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

Software citation & process traceability
Persistent identifiers everywhere

Ocean Observation

Science/expertise

How can we ensure the readers/users can reproduce the results?
Persistent identifiers everywhere

Ocean Observation

Generate Dataset

Science/expertise

How can we ensure the readers/users can reproduce the results?
How can we ensure the readers/users can reproduce the results?
How can we ensure the readers/users can reproduce the results?
How can we ensure the readers/users can reproduce the results?
Persistent identifiers: what about software tools?
Context: who has done what, and how?

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

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
Earth System Science Data

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

RiceAtlas, a spatial database of global rice calendars and production

Alice G. Laborte, Mary Anne Gutierrez, Jane Girly Balanza, Kazuki Saito, Sander J. Zwart, Mirco Boschetti, M.V.R. Murty, Lorena Villano, Jorrel Khalil Aunario, Russell Reinke, Jawoo Koo, Robert J. Hijmans & Andrew Nelson
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.
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
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
Can we go further and have:
"The version used for the DIVA software is the 4.6.9, doi: 10.5281/zenodo.836727"?
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"

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

A. Barth¹, J.-M. Beckers¹, C. Troupin², A. Alvera-Azcárate¹, and L. Vandenbulcke³,⁴

¹GHER, University of Liège, Liège, Belgium
²IMEDEA, Escorlas, Illes Balears, Spain
³seamod.ro/Jailoo srl, Sat Valeni, Com. Salatruclu, Jud. Arges, Romania
⁴CIMAR, University of Porto, Porto, Portugal

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
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

Information infrastructure for Australia’s Integrated Marine Observing System

Authors

Marton G. Hidas, Roger Proctor, Natalia Atkins, Julian Atkinson, Laurent Besnard, Peter Blain, Philip Bohm, Jon Burgess, Kim Finney, Dan Fruehauf, Guillaume Galibert, Xavier Hoenner, Jacqui Hope, Craig Jones, Sebastien Mancini

Open Access | Methodology Article

First Online: 25 May 2016

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, J. Karstensen, P. Testor, P. Alenius, D. Hayes, S. Ruiz, K.J. Heywood, S. Poulouken, L. Mortier, E. Mauri

Get rights and content

https://doi.org/10.1016/j.mio.2016.05.001

doi:10.1016/j.mio.2016.05.001

*discontinued as of 2017*
Definition: online infrastructures whose objective is to persistently store and archive digital artifacts relevant to research:

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

Recent uploads

matplotlib/matplotlib v2.1.0rc1

Michael Droettboom; Thomas A Caswell; John Hunter; Eric Firing; Jens Hedegaard Nielsen; Nelle Varoquaux; Benjamin Root; Elliott Sales de Andrade; Phil Elson; Darren Dale; Jae-Joon Lee; Jouni K. Seppänen; Antony Lee; Ryan May; Damon McDougall; David Stansby; Andrew Straw; Paul Hobson; Tony S Yu; Eric Ma; Christoph Gohlke; Steven Silvester; Charlie Moad; Adrien F. Vincent; Jan Schulz; Peter Würtz; Federico Ariza; Cimarron; Thomas Hisch; Nikita Kniazev

matplotlib: plotting with Python
<table>
<thead>
<tr>
<th>Tool</th>
<th>CKAN</th>
<th>DSpace</th>
<th>Figshare</th>
<th>Zenodo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open Source</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Licence</td>
<td>ckan/ckan</td>
<td>DSpace/DSpace</td>
<td>BSD</td>
<td>zenodo/zenodo</td>
</tr>
<tr>
<td></td>
<td>Affero GNU</td>
<td></td>
<td></td>
<td>GPL-2.0</td>
</tr>
<tr>
<td></td>
<td>GPL v3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1st released</strong></td>
<td>November 2011</td>
<td>November 2002</td>
<td>January 2011</td>
<td>May 2013</td>
</tr>
<tr>
<td>Main technology</td>
<td>Python</td>
<td>Java</td>
<td></td>
<td>Python</td>
</tr>
<tr>
<td>Deployment</td>
<td>Local</td>
<td>Local</td>
<td>Cloud</td>
<td>Cloud</td>
</tr>
<tr>
<td>Integration with GitHub</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Integration with Login</td>
<td>Yes</td>
<td>Not direct</td>
<td>Yes</td>
<td>Login</td>
</tr>
</tbody>
</table>
## Comparison

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

### 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
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>ID account
3. Create new 🤖 (Sign Up)
Upload: you can drop anything

In particular: Software (ok for stable code)
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 Trupin | GHER, University of Liège
Jean-Marie Beckers | GHER, University of Liège
Alexander Barth | GHER, University of Liège
Mohamed Ouerdous | GHER, University of Liège
Digital Object Identifier

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

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

Title

Required.

gher-ulg/DIVA: v4.7.1

Authors

Optional.

Sylvain Watelet

GHER, University of Liège

ORCID (eg. 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 Oubersdous

GHER, University of Liège

ORCID (eg. 0000-0002-1825-0097)
Not necessarily Open!
Upload metadata: funding!

Zenodo is integrated into reporting lines for research funded by the European Commission via OpenAIRE (http://www.openaire.eu). Specify grants which have funded your research, and we will let your funding agency know!

Common strategy for SeaDataCloud, ODIP, ... tools?
Common strategy for SeaDataCloud, ODIP, ... tools?
Zenodo & Github

working hand in hand
Requirements

1. or ID account
2. Source code

Diagram:
- Login
- Upload
- DOI
- ZENODO
- Source Code
Generating doi for software releases

In Zenodo: click on the GitHub tab
In Zenodo: turn on the synchronisation for the selected repositories

<table>
<thead>
<tr>
<th>Repository</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>croupln/python-oceans</td>
<td>OFF</td>
</tr>
<tr>
<td>gher-ulg/Diva_python_tools</td>
<td>ON</td>
</tr>
<tr>
<td>gher-ulg/Documentation</td>
<td>OFF</td>
</tr>
<tr>
<td>gher-ulg/SeaDataCloud</td>
<td>ON</td>
</tr>
<tr>
<td>socib/CMEMS-INSTAC-Dashboard</td>
<td>OFF</td>
</tr>
<tr>
<td>socib/HFRadarReports</td>
<td>OFF</td>
</tr>
</tbody>
</table>
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
Click on the *Release* button
Generating doi for software releases

Fill in the information and ...

Tagging suggestions
It’s common practice to prefix your version names with the letter v. Some good tag names might be v1.8 or v2.3.4.

If the tag isn’t meant for production use, add a pre-release version after the version name. Some good pre-release versions might be v0.2-alpha or v5.9-beta.3.

Semantic versioning
If you’re new to releasing software, we highly recommend reading about semantic versioning.
Generating doi for software releases

...make the release

v1.0

ctroupin released this an hour ago

Release with separated flies for each plot, using the Basemap module.

Downloads

- Source code (zip)
- Source code (tar.gz)
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 1

get DOI for Diva releases
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 🌐
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 😁
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:

- [http://john.albin.net/git/convert-subversion-to-git](http://john.albin.net/git/convert-subversion-to-git)
- [https://www.atlassian.com/git/tutorials/migrating-overview](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

DOI: 10.5281/zenodo.265396
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

Now we can have "The version used for the DIVA software is the 4.6.9, doi: 10.5281/zenodo.400968"

Bonus: "cite as" and social media
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
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
User Experience

1 minute  4 minutes  1 minute  10 minutes  2 minutes

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