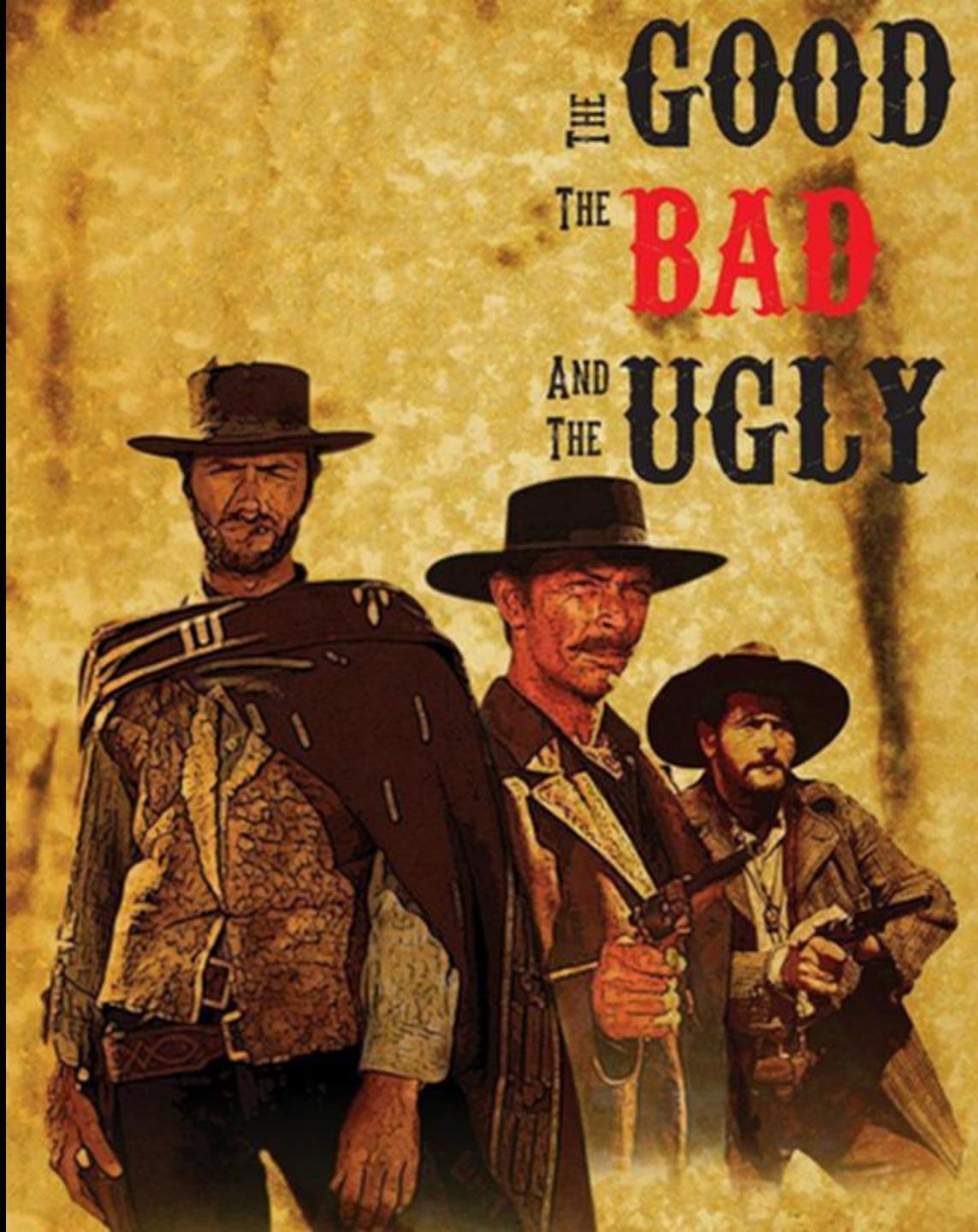


*Where
and
How to
Deposit
your
Data?*

*OSJP
Workshop*



*Ian
Krause*

*Caroline
Salamon*

*ESFL
Library
September*

2017

Program

- Find a suitable repository
- Exercise
- How to deposit your data in Zenodo
- Licenses and citations
- Evaluation

Find a suitable data repository

From active data to a repository: Why & How?

Why?

Preserve

Shared drives and other active data management tools (included GitHub) are not intended nor equipped for long-term storage and preservation whereas data repositories, like Zenodo, are.

Cite

By archiving your active data in a repository you will get a DOI

Discover




Data discoverability enhances the visibility and the impact of your research, and enables a broad dissemination of your research outputs

How?

Exercises

!!! Every (trusted) data repository helps you in making your data safely stored, easily citeable, and discoverable !!!

Comparison: Dryad – Figshare – Zenodo

		
Nonprofit organization	Digital Science (McMillan - NPG)	CERN & OpenAire
multidisciplinary	multidisciplinary	multidisciplinary
multiformat	multiformat	multiformat
CC0	CC0 (data) - CC BY (files)	license to choose
Fee requested (not for «low-income countries»)	Free of charge	Free of charge
Features <ul style="list-style-type: none"> • link data <-> article • view of complete metadata • # of downloads • submission integration • long-term preservation • DOI versioning 	Features <ul style="list-style-type: none"> • link data <-> article • # of views and downloads • negative data • collaborative spaces • DOI versioning 	Features <ul style="list-style-type: none"> • altmetrics • personal collections • login with ORCID and Github • DOI versioning
Cons <p>DPC (data processing charges)</p>	Cons <p>owned by a big commercial company</p>	Cons <p>search tool to be completed</p>

Research data publishing

Data repository

Why?

Long-term accessibility and **preservation**
Increased **discoverability** and **reuse** of data



How to choose?

- Data sharing practices within your community
- Repository's specific features
- Check [Re3data.org](https://re3data.org)
- Ask your librarian 😊

**Comply with
requirements
and policies**



- **Funders**
- **Publishers**
- **Institution**



Disciplinary repositories may be better adapted

EMBL (bioinformatics)

<http://www.ebi.ac.uk/>

- Search by
 - Genomes & metagenomes
 - Nucleotide sequences
 - Protein sequences
 - Gene expression
 - Protein families
- According to www.ebi.ac.uk/about the platform seems well supported, including financially.
- No DOIs, what does guarantee permanence of links?
- How is Long Term Preservation handled?

Crystallography Open Database

<http://www.crystallography.net/cod/>

- Search by formula (drawing)
- Search using specific fields such as Z, volume, Hill notation...
- No information on financing after 2011, will it last?
- No DOIs, what does guarantee permanence of links?
- How is Long Term Preservation handled?



Exercise (in groups)

Find a good repository in your discipline

Form groups by disciplines.

Look for appropriate data repositories for 10 min

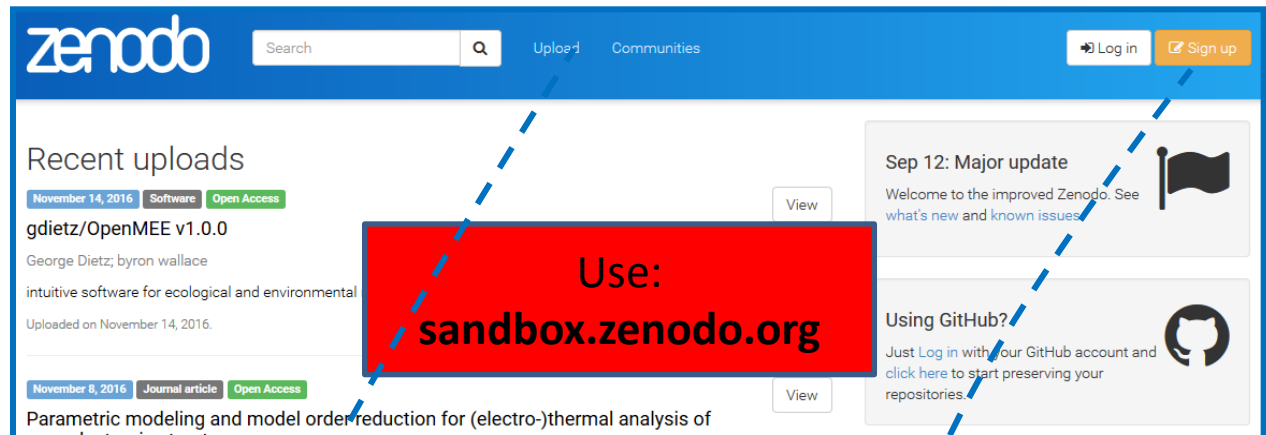
- Is there a good one?
- What are your criteria for your assessment?
(criteria categories: disciplinary, search & access and data permanence)

Each group gives its feedback to all (1-2min/group)

**How to deposit your
Data in Zenodo?**

Deposit in Zenodo

1. Open
Zenodo



Use:
sandbox.zenodo.org

2. Create
account

3. Upload

zenodo

Search

Upload Communities

p01@jankrause.net

Delete Save Publish

New upload

Instructions: (i) Upload maximum one file or file (ii) required fields (marked with a red star) (iii) Press "Save" to save your upload for editing later (iv) When ready, press "Publish" to finalize and make your upload public.

Files

Drag and drop files here

— Or —

Choose files

(minimum 1 file required, max 50 GB per dataset - contact us for larger datasets)

Upload type

required

Publication Poster Presentation Dataset Image Video/Audio Software Lesson

Publication type

Journal article

Basic information

required

Digital Object Identifier

e.g. 10.1234/foo.bar

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

2017-09-14

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

Title

Required

Authors

Family Name, Given names Affiliation

Add another author

Description

Required

Keywords

Add another keyword

Additional notes

Optional

License

required

Access right

Open Access

Embargoed Access

Restricted Access

Closed Access

Required: Open access uploads have considerably higher visibility on Zenodo.

1. Upload files

2. Select datatype

3. Fill in mandatory descriptive metadata:
Title, Authors, Description, publishing date

4. Setup access rights:
Choose license, add optional embargo

5. Hit Upload

Capture and publish your GitHub repository in Zenodo



When you write your software, you can **make the work you share on GitHub citable: archive one of your GitHub repositories in Zenodo and assign it a DOI**

Preliminary steps

zenodo Search Upload Communities Log in Sign up

Recent uploads

November 14, 2016 Software Open Access

gdietz/OpenMEE v1.0.0

George Dietz; byron wallace

intuitive software for ecological and environmental meta-analysis

Unloaded on November 14, 2016

Use:
sandbox.zenodo.org

Sep 12: Major update
Welcome to the improved Zenodo. See [what's new and known issues](#).

Using GitHub?
Just [Log in](#) with your GitHub account and [click here](#) to start preserving your repositories.

Authorize application

Zenodo by @zenodo would like permission to access your account

Review permissions



Personal user data

Email addresses (read-only)



Repository webhooks and services

Admin access



Organizations and teams

Read-only access

Authorize application

Zenodo

Software Preservation Made Simple!

[Visit application's website](#)

[Learn more about OAuth](#)

Capture and publish your GitHub repository in Zenodo

Exercise

zenodo Search Upload Communities lorenza.salvatori@epfl.ch

Home / Account / GitHub

Settings

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

GitHub Repositories (updated 2 minutes ago) 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)

Repositories

If your organization's repositories do not show up in the list, please ensure you have enabled [third-party access](#) to the Zenodo application. Private repositories are not supported.

You have no repositories on GitHub.
Go to [GitHub](#) and create your first or click Sync-button to synchronize latest changes from GitHub.

Capture and publish your GitHub repository in Zenodo

Exercise



Create a new release
Fill in the release notes
Publish your release

Creating a new
release triggers
Zenodo into
archiving your
repository

zenodo

Search

Upload

Communities

Go to Upload

Select your Upload
...and edit metadata

Use the
metadata

Licences and citations

Licences



They cover code, data , text and multimedia:

- CC-BY.
 - Reuse, adapt, publish derivatives.
 - Obligation : cite the creator(s).
- Additional options :
 - non commercial, share alike, no derivatives.
 - resulting in total 6 CC-By licenses.
- CC0 : almost public domain.

Other licenses are specific for code:

- GNU-GPL : Open Software.
- Apache2.0 : smaller codes, libraries.
 - Permissive.
 - No share alike clause.
 - Preservation of copyright notice.
- BSD-3clause - similar.



Apache



Research data citations

DOI = Digital Object Identifier

... is a **unique & persistent** identifier.

It concerns datasets, but not only (also online documents like papers).

Once a DOI is assigned to a dataset, this dataset is **published**.

It is attributed by **repositories** (like Zenodo).

Advantages?

It gives credit to data producers.

Contract with DataCite: the provider (Zenodo) has to maintain access to the dataset.

Alternatives to DOI?

Handle System (system used by Datacite), Persistent URLs, ARKs, etc.

Evaluation



Individual exercise (with evaluation)

Submit your dataset to Zenodo

- Publish...
 - ... a real dataset of yours on zenodo.org.
 - or**
 - ... a fake dataset of your choice on sandbox.zenodo.org. If you don't know what to publish, you can use this one: <http://go.epfl.ch/aEF>.
- Describe the dataset with appropriate metadata and access rights.
- Don't forget the README file to ensure that your data can be correctly interpreted and reanalyzed by others.

To validate your workshop, send us an individual email with the following information by October 27th:

Subject: OSIP2017 Workshop :

Content: The DOI of the dataset you have published

E-mail address: researchdata@epfl.ch

Credits

All the icons are published under CC BY 3.0 license on thenounproject.com:



User attention by TMD



Repository by lastspark



Law by Marie Van den Broeck



Peer Evaluation By Duke Innovation Co-Lab



Friends by Les vieux garçons



Email by Lorena Salagre

Thank you for your attention 😊



Any questions?

researchdata@epfl.ch