

# Складання плану з управління науковими даними

Ірина Кучма, менеджерка програми “Відкритий доступ”

Семінар “Цифрова підтримка наукових досліджень”

9 жовтня 2017 р., м. Київ



Із Зазначенням Авторства 4.0 Міжнародна

# План з управління даними (DMP)

Короткий план, який описує:

- створення/збір даних
- документацію
- доступ до даних для третіх осіб/організацій
- зберігання даних
- бек-ап
- публічний доступ
- архівування

# Планування: починайте з кінця

Який спосіб організації даних зручний для тих, хто використовуватиме ваші дані?



# DMP

1. Опис даних
2. Стандарти / методологія збору і управління даними
3. Етичні питання і питання інтелектуальної власності (обмеження, ембарго, конфіденційність)
4. Доступ до даних (як, коли, хто)
5. Стратегія довготермінового збереження

# Організація даних

## Meaningful file names

Below are tips on meaningful and consistent file names. Read more in '[Choosing a file name](#)'.<sup>(2)</sup>

- ❑ Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD of yymmdd).
- ❑ Do not use strange characters like ?\!@\*%{[<> in the file name.
- ❑ Use traceable file names, such as Project\_Instrument\_locatie\_YYYYMMDD.ext.
- ❑ Make sure to only use each file once in the folder structure. If you store a file in more than one place, several versions of the same file can unwillingly be created.
- ❑ See also [version management](#).

It is good practice to note the file naming and its meaning in a readme.txt.

Even if a researcher is well underway with his project consistent file naming is still an option by using a [bulk file rename utility](#).<sup>(3)</sup> It is important, however, to check if this bulk renamer delivers on its promises.

<http://datasupport.researchdata.nl/en/start-de-cursus/iii-onderzoeksfase/organising-data>



white\_data\_20140708.csv



blue\_data\_20140708.docx

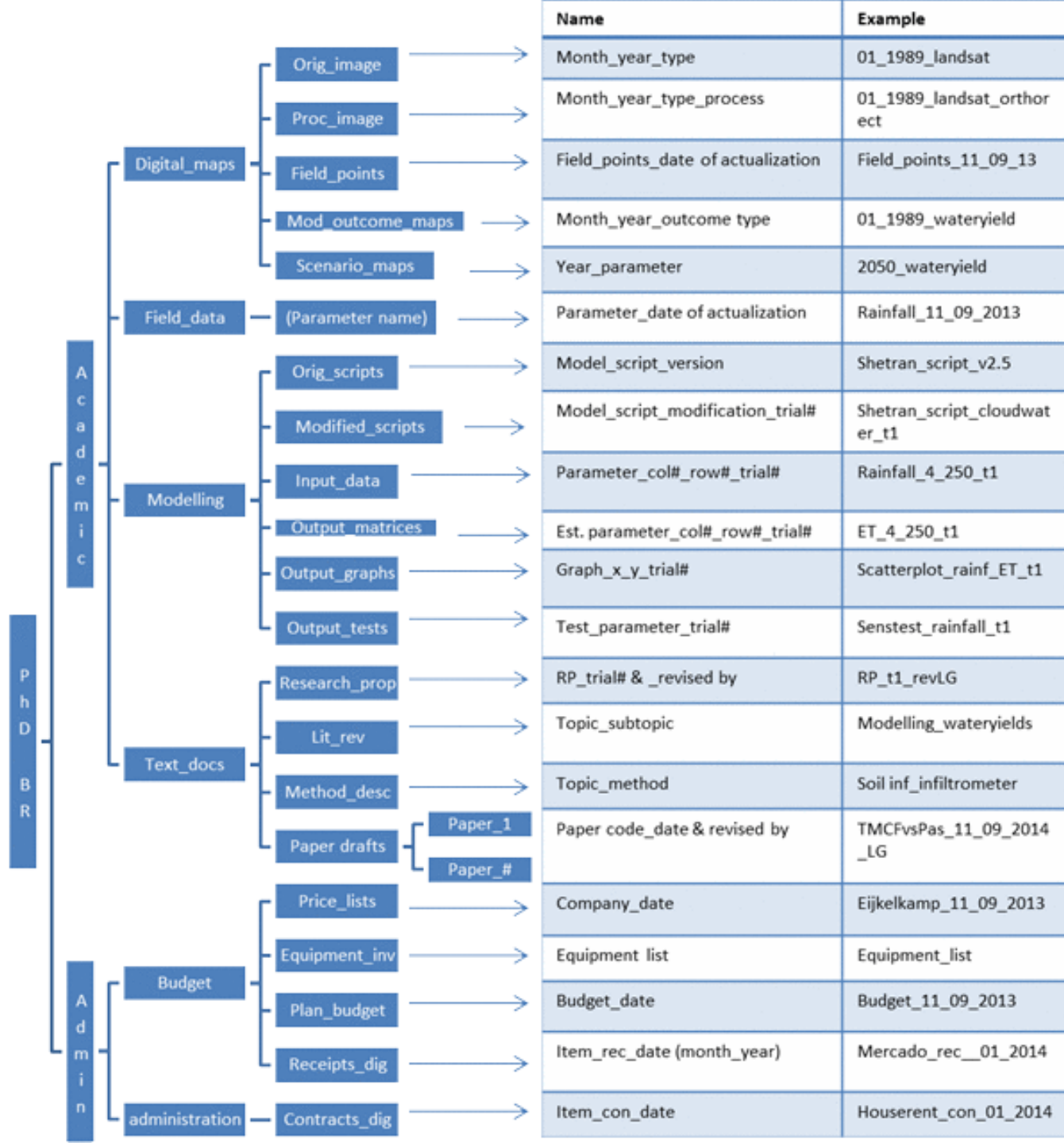


red\_data\_20140708.R



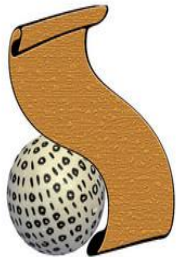
red\_data\_20140708\_v02.R

*File naming and version management*

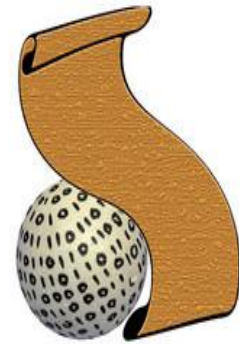
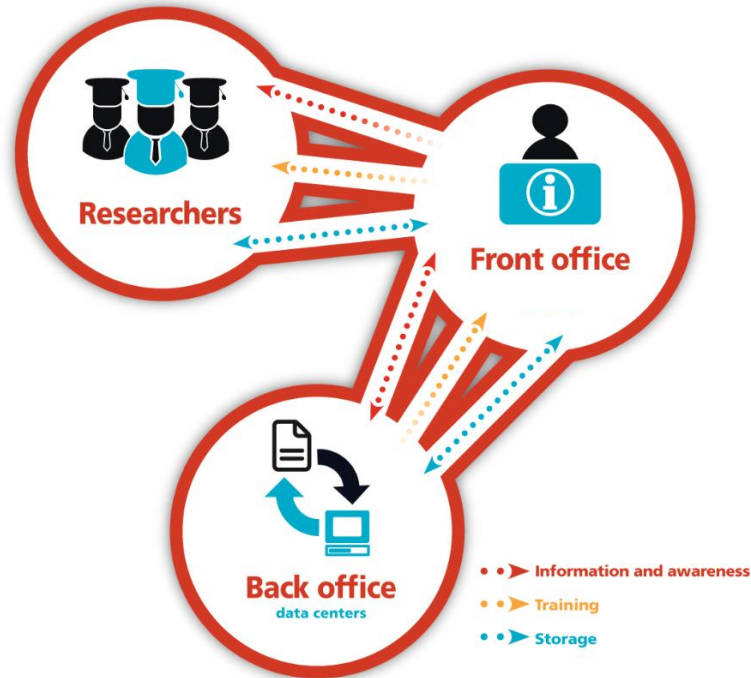




**Комерційні партнери**



**Видавці**  
Data Availability policy



**Організації**  
RDM policy  
Facilities



**Донори**

<https://www.openaire.eu/briefpaper-rdm-infonoads>



Збереження < > Архівування

Архівування < > Знаходження

Знаходження < > Доступність

Доступність < > Зрозумілість

Зрозумілість < > Використання

Флешка ненадійна

Ідентифікатор важливий, але не гарантує  
використання

Дані в пропрієтарних форматах використовують  
менше



# Що зберігати?

Дані

Метадані

Документацію

Програмне

забезпечення

Research Data Alliance (RDA)

<http://rd-alliance.github.io/metadata-directory/standards/>

FAIR Guiding Principles for scientific data management & stewardship

<http://www.nature.com/articles/sdata201618>

How to select and appraise research data:

[www.dcc.ac.uk/resources/how-guides/appraise-select-research-data](http://www.dcc.ac.uk/resources/how-guides/appraise-select-research-data)



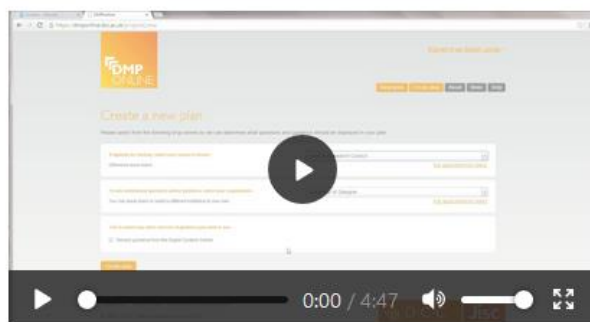
PUBLICATIONS AND DATA

[Home](#)[About](#)[Future plans](#)[Help](#)[Change language](#)

## Welcome.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3).

### Screencast on how to use DMPonline



## Sign in

[Forgot your password?](#)☐ Remember me[Sign in](#)[Or sign in with your institutional credentials](#) (UK users only)

## Create account

New to DMPonline? Create an account today.

<https://dmponline.dcc.ac.uk>

[Contact us](#) | [Terms of use](#)

© 2004 - 2017 Digital Curation Centre (DCC)



# DMPonline

Створити  
DMP

## Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.

If applying for funding, select your research funder.

European Commission (Horizon 2020)

Otherwise leave blank.

[Not applicable/not listed.](#)

To see institutional questions and/or guidance, select your organisation.

University of Glasgow

You may leave blank or select a different organisation to your own.

[Not applicable/not listed.](#)

Tick to select any other sources of guidance you wish to see.

- ☐ DCC guidance
- ☐ EUDAT
- ☐ School of Humanities
- ☐ Computing

Create plan

Обрати  
шаблон  
донора

Обрати  
університет

Обрати додаткові  
поради

# Опис даних: приклади

The final dataset will include self-reported demographic and behavioural data from interviews with the subjects and laboratory data from urine specimens provided. From [NIH data sharing statements](#)

Every two days, we will subsample *E. affinis* populations growing under our treatment conditions. We will use a microscope to identify the life stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook and then copy the data into an Excel spreadsheet. The Excel spreadsheet will be saved as a comma separated value (.csv) file.

From DataOne – [E. affinis DMP example](#)

# Метадані: приклади

Metadata will be tagged in XML using the **Data Documentation Initiative (DDI) format**. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and **all information necessary for a secondary analyst** to use the data accurately and effectively.

From [ICPSR Framework for Creating a DMP](#)

We will first document our metadata by taking careful notes in the laboratory notebook that refer to specific data files and **describe all columns, units, abbreviations, and missing value identifiers**. These notes will be transcribed into a **.txt document that will be stored with the data file**. After all of the data are collected, we will then use EML (Ecological Metadata Language) to digitize our metadata. **EML is one of the accepted formats used in ecology**, and works well for the types of data we will be producing. We will create these metadata using Morpho software, available through KNB. The metadata will fully describe the data files and the context of the measurements.

From DataOne – [E. affinis DMP example](#)



REUSABLE DATA



## In this section

[Briefing Papers](#)[How-to Guides & Checklists](#)[Developing RDM Services](#)[Curation Lifecycle Model](#)[Curation Reference Manual](#)[Policy and legal](#)[Data Management Plans](#)[Tools](#)[Case studies](#)[Repository audit and assessment](#)

### Standards

[Disciplinary Metadata](#)[DIFFUSE](#)[Publications and presentations](#)[Roles](#)[Curation journals](#)[Informatics research](#)[External resources](#)[Online Store](#)

## Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Please note that a [community-maintained version of this directory](#) has been set up under the auspices of the Research Data Alliance.

### Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

## Metadata

RDA | Metadata Directory

[View the standards](#)

[View the extensions](#)

[View the tools](#)

[View the use cases](#)

Browse by subject areas

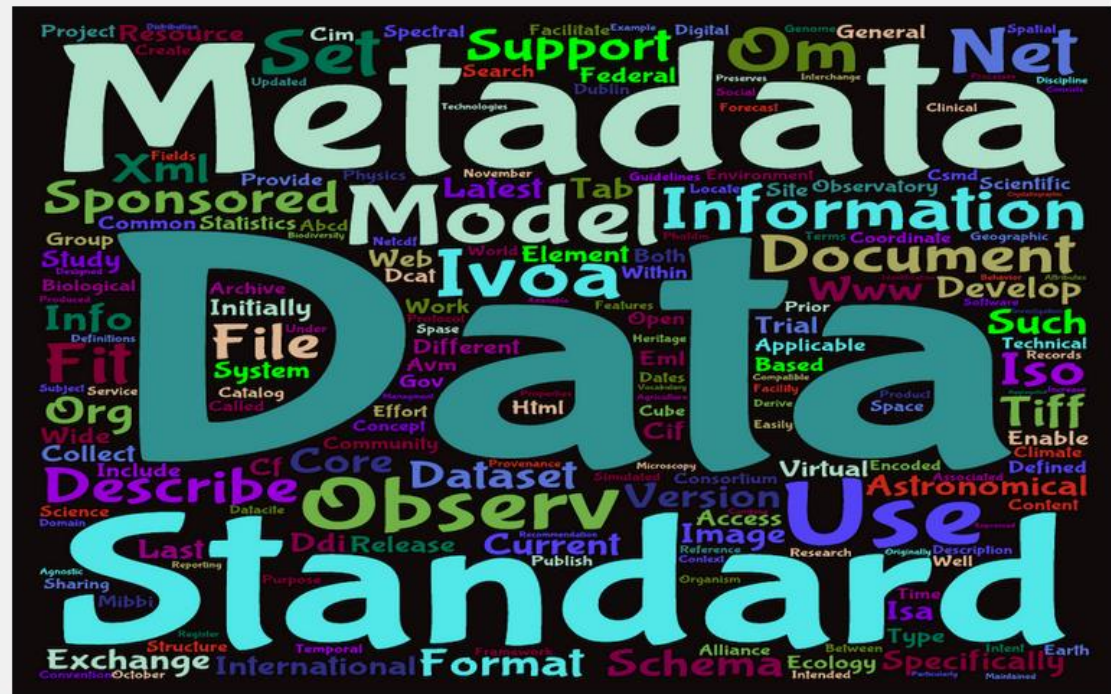
Contribute

Add standards

Add extensions

Add tools

Add use cases



## Metadata Standards Directory Working Group

The RDA Metadata Standards Directory Working Group is supported by individuals and organizations involved in the development, implementation, and use of metadata for scientific data. The overriding goal is to develop a collaborative, open directory of metadata standards applicable to scientific data can help address infrastructure challenges.

The RDA Metadata Standards Directory is maintained by [Sean Chen](#), [Kate Anne Alderete](#), and [Alex Ball](#).

The theme is maintained by [Dustin Allen](#).

This page was generated by [GitHub Pages](#).

<http://rd-alliance.github.io/metadata-directory>



# Доступ до даних: приклади

The videos will be made available **via the [bristol.ac.uk](http://bristol.ac.uk) website** (both as streaming media and downloads) HD and SD versions will be provided to accommodate those with lower bandwidth. Videos will also be made available **via Vimeo**, a platform that is already well used by research students at Bristol. **Appropriate metadata will also be provided** to the existing Vimeo standard.

All video will also be available **for download and re-editing by third parties**. To facilitate this **Creative Commons** licenses will be assigned to each item. In order to ensure this usage is possible, the **required permissions will be gathered** from participants (using a suitable release form) before recording commences.

From [University of Bristol Kitchen Cosmology DMP](#)

We will make the data and associated documentation available to users under a **data-sharing agreement** that provides for: (1) a commitment to using the data **only for research purposes** and not to identify any individual participant; (2) a commitment to **securing the data** using appropriate computer technology; and (3) a commitment to **destroying or returning the data after analyses** are completed.

From [NIH data sharing statements](#)

## Обмеження : приклади

Because the STDs being studied are reportable diseases, we will be **collecting identifying information**. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there **remains the possibility of deductive disclosure of subjects** with unusual characteristics. Thus, we will make the data and associated documentation available to users **only under a data-sharing agreement**.

From [NIH data sharing statements](#)

## Обмеження: приклади (2)

1. Share data **privately within 1 year.**

*Data will be held in Private Repository, but metadata will be public*

2. Release data to **public within 2 years.**

*Encouraged after one year to release data for public access.*

3. **Request, in writing, data privacy up to 4 years.**

*Extensions beyond 3 years will only be granted for compelling cases.*

4. Consult with creators of private CZO datasets prior to use.

*Is required to **seek consent before using private data** they can access*

From [Boulder Creek Critical Zone Observatory DMP](#)

# Архівування: приклади

The investigators will **work with staff at the UKDA** to determine **what to archive and how long** the deposited data should be retained. Future long-term use of the data will be ensured by **placing a copy of the data into the repository**.

From [ICPSR Framework for Creating a DMP](#)

Data will be provided in **file formats considered appropriate for long-term access**, as recommended by the UK Data Service. For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Appropriate **documentation necessary** to understand the data will also be provided. Anonymised data will be held for **a minimum of 10 years** following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule. Biological samples (output 3) will be **deposited with the UK BioBank** for future use.

From [Writing a Wellcome Trust Data Management and Sharing Plan](#)

# Доступ до даних

Учасники проекту



«робочі» дані

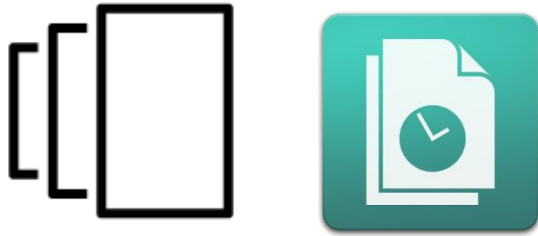
(Відкритий) доступ



Стабільність,  
пошук, цитування,  
ліцензування

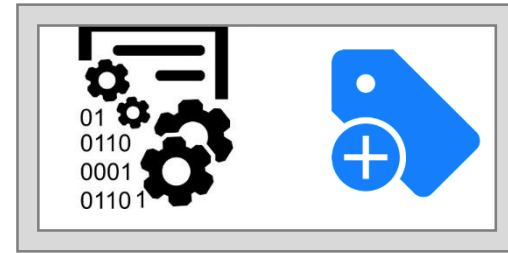
# Збереження даних

Збереження і бек-ап  
активного проекту



Мережеве  
файлосховище, диск

Довготермінове  
архівування і збереження



Цифровий репозитарій


<https://www.openaire.eu/opendatapilot-repository>


PARTICIPATE

SEARCH

MONITOR

SUPPORT

OPEN ACCESS

# How to select a repository?

Updated on 15 November 2016



## INFORMATION ABOUT SELECTING A DATA REPOSITORY

### OVERVIEW

In deciding where to store your data, you may have a number of choices about who will look after it. The choice may be straightforward if you have an established data management facility in your domain or institution, or even within your research group or department. When data preservation standards or norms exist in your discipline, these should be followed. Your research funder may recommend a data centre or self-deposit archive. In order of preference:

1. Use an external data archive or repository already established for your research domain to preserve the data according to recognised standards in your discipline.
2. If available, use an institutional research data repository, or your research group's established data management facilities.
3. Use a cost-free data repository such as [Zenodo](https://zenodo.org/).
4. Search for other data repositories here: [re3data.org](https://re3data.org/). On top of specific research disciplines you can filter on access categories, data usage licenses, trustworthy data repositories (with a certificate or explicitly adhering to archival standards) and whether a repository gives the data a persistent identifier.



## Open Access

### BACKGROUND

Overview

Policies and Mandates

Open Access in FP7

Open Access in H2020

### IN PRACTICE

EU Member States

### PROJECTS

OpenAIRE-Connect

### PILOTS

FP7 Post-Grant OA Pilot

Open Research Data Pilot

Create a Data Management Plan

Select a Data Repository



# Як обрати репозитарій даних?

## Сертифікат Trustworthy Digital Repository



Data Seal of Approval: <http://datasealofapproval.org/en>

nestor seal: [http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel\\_node.html](http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html)

ISO 16363: <http://www.iso16363.org>



## Recent uploads

<https://zenodo.org>

August 30, 2017 (v2) Working paper Open Access

View

## Introducing Parsl: A Python Parallel Scripting Library

Babuji, Yadu; Brizius, Alison; Chard, Kyle; Foster, Ian; Katz, Daniel S.; Wilde, Michael; Wozniak, Justin

Researchers frequently rely on large-scale and domain-specific workflows to conduct their science. These workflows may integrate a variety of independent software functions and external applications. However, developing and executing such workflows can be difficult, requiring complex...

Uploaded on September 15, 2017

1 more version(s) exist for this record

September 6, 2017 (v4) Software Open Access

View

## geodynamics/pylith: PyLith v2.2.1

Brad Aagaard; Charles Williams; Matthew Knepley

Added new examples. examples/3d/subduction: New suite of examples for a 3-D subduction zone. This intermediate level suite of examples illustrates a wide range of PyLith features for quasi-static simulations. examples/2d/subduction: Added quasi-static spontaneous rupture earthquake cycle...

Uploaded on September 6, 2017

3 more version(s) exist for this record

August 31, 2017 (v1) Figure Open Access

View

## Fig. 1 in Vanmanenia oricampus I a new species of loach from the Plain of Jars I Laos (Teleostei: Gastromyzontidae)

Maurice Kottelat

Fig. 1. Vanmanenia oricampus I MHNG 2767.094 I holotype I 47.5 mm SL; Laos: Nam Ngum on Plain of Jars.

Uploaded on September 6, 2017

December 31, 2015 (v1) Journal article Open Access

View

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 European Commission via OpenAIRE.
- **Flexible licensing** — because not everything is under Creative Commons.
- **Safe** — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

Read more about Zenodo and its [features](#).

# Introducing Zenodo!

(All) Research.  
Shared.

— your one stop research shop!

all research outputs from across all fields of research are welcome! Zenodo accepts any file format as well as both positive and negative results. We choose to promote peer-reviewed openly accessible research, and we curate the uploads posted on the front-page.

Citeable.  
Discoverable.

— be found!

Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. Zenodo further supports harvesting of all content via the OAI-PMH protocol.

Communities

— create your own repository

Zenodo allows you to create your own collection and accept or reject uploads submitted to it. Creating a space for your next workshop or project has never been easier. Plus, everything is citeable and discoverable!

Safe

— more than just a drop box!

Your research output is stored safely for the future in same cloud infrastructure as research data from CERN's [Large Hadron Collider](#) and using CERN's battle-tested repository software [INVENIO](#), which is used by some of the world's largest repositories such as [INSPIRE HEP](#) and [CERN Document Server](#).

Reporting

— tell your funding agency!

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Just upload your research to Zenodo, and we will take care of the reporting for you. We plan to expand this feature with further funding agencies in the future, so stay tuned!

Flexible Licensing

— not everything is under Creative Commons

Zenodo encourages you to share your research as openly as possible to maximize use and re-use of your research results. However, we also acknowledge that one size does not fit all. Therefore, we allow for uploading under a variety of different licenses and access levels. You are responsible for respecting applicable copyright and license conditions for the files you upload.

# Get started!

Make your first upload - all research outputs from across all fields of research are welcome.

[New Upload](#)

Delete

Save

Publish

## New upload

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

Files ▼

Choose files

Start upload

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 ▼



# Застереження

Застереження	Вирішення
Неадекватне використання через нерозуміння мети і параметрів дослідження	
Безпека і конфіденційність	
Відсутність посилання на автора	
Втрата переваг при суперництві за гранти	

## Застереження (2)

Застереження	Вирішення
Неадекватне використання через нерозуміння мети і параметрів дослідження	 метадані
Безпека і конфіденційність	 метадані
Відсутність посилання на автора	 метадані
Втрата переваг при суперництві за гранти	 метадані

## Застереження (3)

Застереження	Вирішення
Неадекватне використання через нерозуміння мети і параметрів дослідження	Надайте короткий опис мети, застереження щодо використання, додаткову інформацію
Безпека і конфіденційність	Вкажіть, хто може отримати доступ до даних і застереження щодо використання
Відсутність посилання на автора	Вкажіть, обов'язкове цитування і застереження щодо використання
Втрата переваг при суперництві за гранти	Створіть публічну версію з узагальненим описом обробки даних

# Ліцензування наукових даних

## CREATIVE COMMONS



Некомеційне

Що таке комерційне?



Без похідних



Значно обмежує  
використання

A Digital Curation Centre and JISC Legal  
'working level' guide

**How to License  
Research Data**

Alex Ball (DCC)

Horizon 2020:

Digital Curation Centre, 2012.  
Licensed under Creative Commons Attribution 2.5 Scotland:  
<http://creativecommons.org/licenses/by/2.5/scotland/>



# Селектор ліцензій EUDAT

Do you own copyright and similar rights in your dataset and all its constitutive parts?

Yes

No

Do you allow others to make commercial use of you data?

Yes

No

## Creative Commons Attribution (CC-BY)

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

## Public Domain Dedication (CC Zero)

CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

<http://ufal.github.io/public-license-selector>

# Що зберігати?

Критерії:

- Те, що потрібно, наприклад дані, які використовувалися у публікаціях
- Те, що неможливо відворити, наприклад, характеристики середовища
- Те, що може потенційно знадобитися іншим
- Те, що має наукову, культурну, історичну цінність
- Те, що слід знищити за умовами угоди

How to select and appraise research data: [www.dcc.ac.uk/resources/how-guides/appraise-select-research-data](http://www.dcc.ac.uk/resources/how-guides/appraise-select-research-data)

# Корисні посилання

How to develop a DMP <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>

RDM brochure and template  
[https://dans.knaw.nl/en/about/organisation-and-policy/information-material?set\\_language=en](https://dans.knaw.nl/en/about/organisation-and-policy/information-material?set_language=en)

OpenAIRE guidelines [www.openaire.eu/opendatapilot-dmp](http://www.openaire.eu/opendatapilot-dmp)

ICPSR framework for a DMP  
[www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html](http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html)

# Корисні посилання

Where to keep research data <http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data>

Five steps to decide what data to keep <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>

Re3data <http://www.re3data.org/>

Figshare <https://figshare.com/>

Genbank <https://www.ncbi.nlm.nih.gov/genbank/>

How to write a lay summary <http://www.dcc.ac.uk/resources/how-guides/write-lay-summary>

Lay summaries <https://www.bhf.org.uk/research/information-for-researchers/how-to-apply/lay-summaries>

З подякою

Marjan Grootveld:

[marjan.grootveld@dans.knaw.nl](mailto:marjan.grootveld@dans.knaw.nl)

Sarah Jones: [sarah.jones@glasgow.ac.uk](mailto:sarah.jones@glasgow.ac.uk)

**Acknowledgements:**

Thanks to DANS and DCC for reuse of slide



# Postgraduate Data Management Plan

---

## 1. Overview

<b>1.1 Postgraduate Researcher:</b>
<b>1.2 Project title:</b>
<b>1.3 Project start and end dates:</b>
<b>1.4 Project context:</b>

## 2. Defining your data

<b>Describe your data (e.g. type, format, volume)</b>
---

# DMP

Опишіть ваші дані (тип, формат, обсяг)

Управління даними (збереження і бек-ап, структура, версії, документація, тощо)

Архівування даних

Які дані будуть доступними і як (наприклад, в якому репозитарії, під якою ліцензією)

**Дякую!**  
**Запитання?**

[iryna.kuchma@eifl.net](mailto:iryna.kuchma@eifl.net)  
Twitter: @irynakuchma

**eifl** **KNOWLEDGE  
WITHOUT  
BOUNDARIES**

[www.eifl.net](http://www.eifl.net)