

OpenAIRE  
Belgium

# Open Research Data

Open research data and data management  
for Horizon 2020 projects



OpenAIRE

Emilie Hermans  
Project Assistant OpenAIRE, UGent

can be reused under the [CC BY license](#)

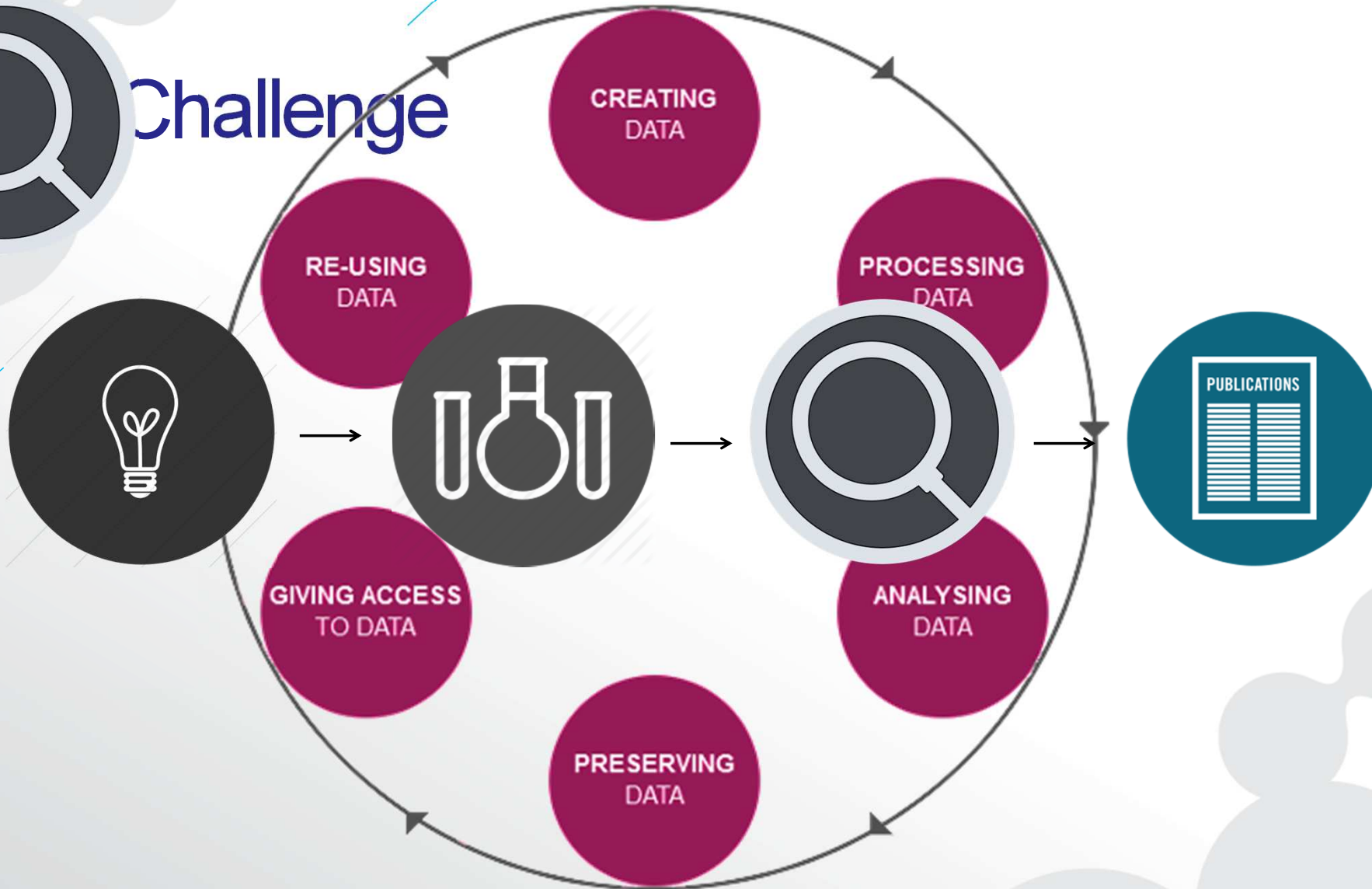


# Challenge





# Challenge



# DATA MANAGEMENT AND OPEN DATA

Prevents data loss



Maximize usefulness



Fosters creativity



Citizens science



Promote integrity



Increases transparency



Credit & longer shelf life <sup>1</sup>

1. e.g. Piwowar HA, Vision TJ. (2013) Data reuse and the open data citation advantage. PeerJ 1:e175 <https://doi.org/10.7717/peerj.175>, Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. [doi:10.1371/journal.pone.0000308](https://doi.org/10.1371/journal.pone.0000308)

# EC: The Open Research Data Pilot

Flexible ORD pilot:  
From limited to default in 2017

- Foster Open Science
- Avoid duplication of research and loss of resources

Data Management  
Planning

Open Access  
to research data  
(or partially opt-out)

# Which areas are participating?

Projects started in 2014-2016

## Limited ORD Pilot

- Limited ORD pilot: some areas: Check [Article 29.3](#)
- DMP/Dataset
- Possibility to opt-in or opt-out

From 2017

## Extended ORD Pilot

- Participating is **default** option for **all** projects
- 1 DMP/Project
- Possibility to opt-out

**Costs eligible** (Article 6.2.D.3 of the Model Grant Agreement)

# (PARTIALLY) OPTING-OUT

**As open as possible as closed as necessary**

## Reasons e.g.

- Exploitation of results
  - Confidentiality
  - Protection of personal data
  - Would jeopardize the main aim of the action
  - No data generated
  - Any other legitimate reason
- Projects can opt out at any stage:
    - Complete opt-out via project amendment
    - Complete or partially opt-out:  
describe issues in project DMP

# Requirements of the Data Pilot



1. Data Management Plan (DMP)



2. Deposit data in data repository



3. Provide information to validate results



4. Open up data



# What is a DMP?



Handling of data during and after project



Living document: update



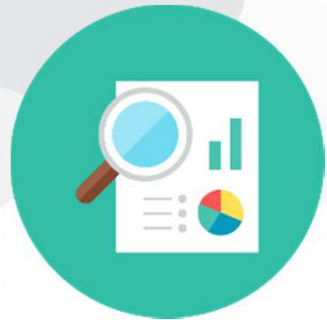
Reflects on curation,  
preservation, sustainability and  
security



What parts will be  
open and how?

# Content of H2020 DMP

Template: EC guidelines on FAIR Data Management



Data  
summary



FAIR  
Data  
principles



Resources

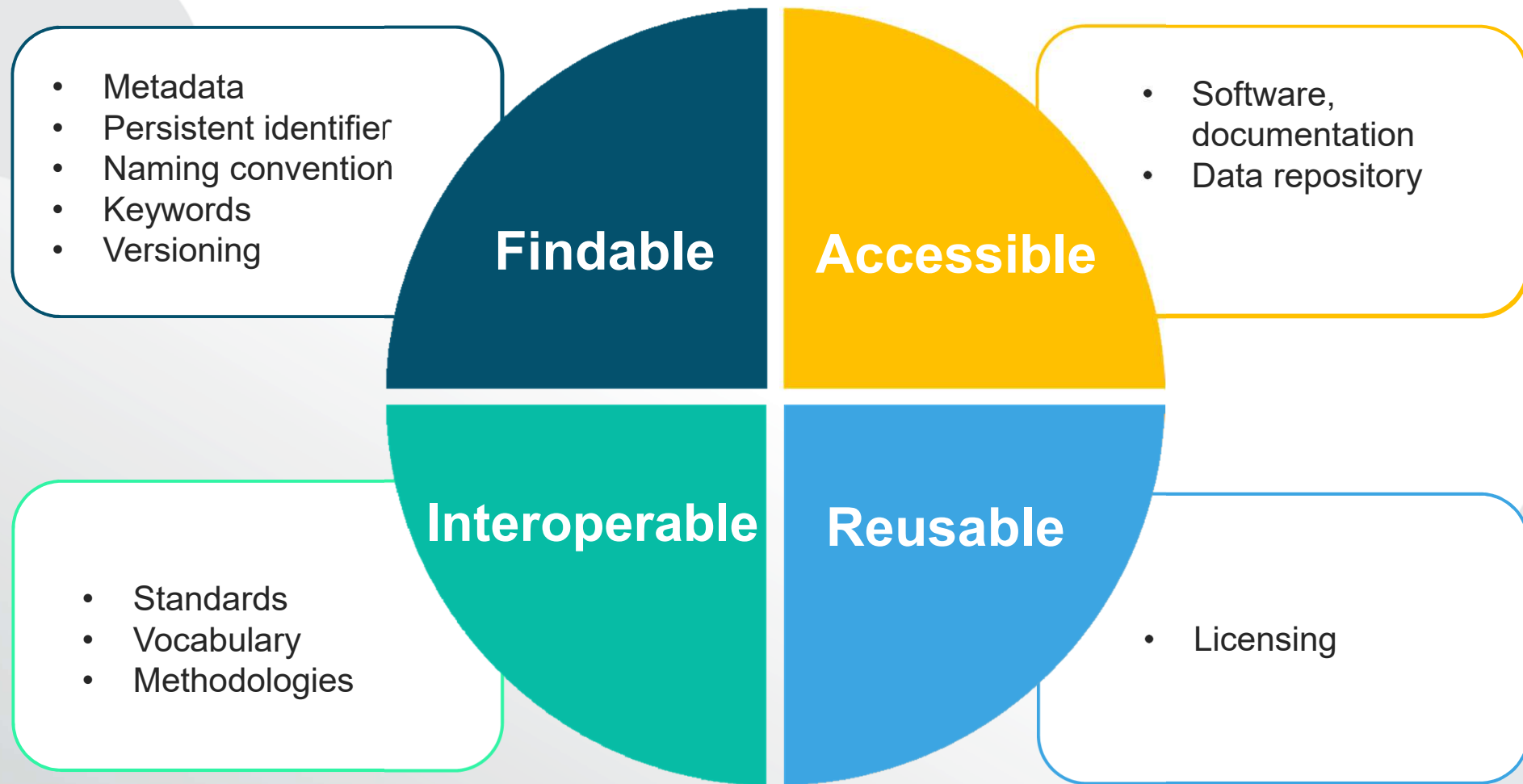


Data  
security



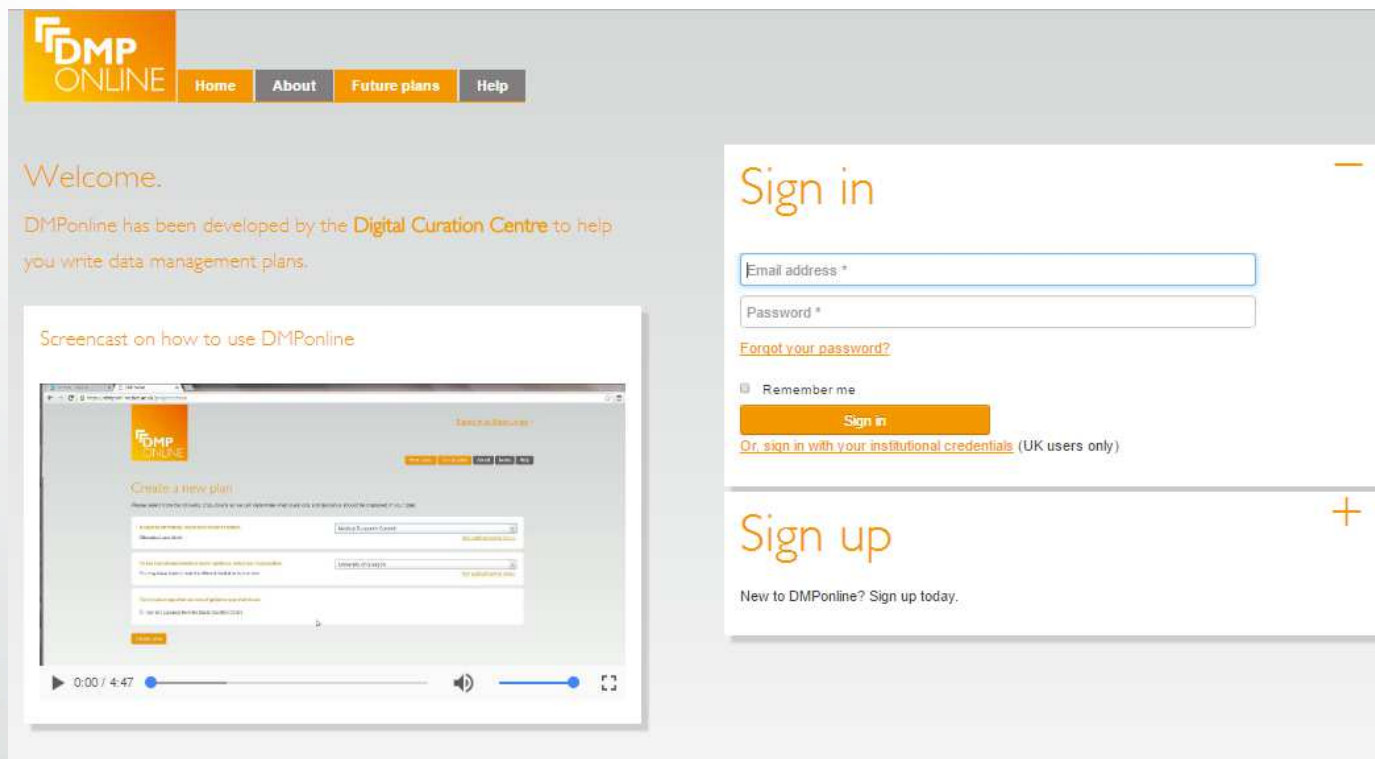
Ethical  
aspects

# FAIR data principles



# How to write a DMP

[dmponline.dcc.ac.uk](http://dmponline.dcc.ac.uk)



The screenshot displays the DMPonline website interface. At the top left is the 'DMP ONLINE' logo. A navigation menu contains 'Home', 'About', 'Future plans', and 'Help'. The main content area features a 'Welcome.' message, a brief description of the service, and a video player titled 'Screencast on how to use DMPonline'. On the right side, there is a 'Sign in' section with input fields for 'Email address \*' and 'Password \*', a 'Remember me' checkbox, a 'Sign in' button, and a link for 'Forgot your password?'. Below this is a 'Sign up' section with the text 'New to DMPonline? Sign up today.' and a '+' icon.

# Create and confirm plan

Signed in as **Emile Hermans**

**DMP ONLINE** View plans Create plan About Future plans Help

## 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.**  
Otherwise leave blank.

**To see institutional questions and/or guidance, select your organisation.**  
You may leave blank or select a different organisation to your own.

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

- DCC guidance

Create plan

European Commission (Horizon 2020)

Funder

- Arts & Humanities Research Council
- Biotechnology and Biological Sciences Research Council
- Cancer Research UK
- Economic and Social Research Council
- Engineering and Physical Sciences Research Council
- European Commission (Horizon 2020)
- Medical Research Council

# Plan Details

My plan (Horizon 2020 DMP)

Plan details | Initial DMP | Detailed DMP | Final review DMP | Share | Export

Please fill in the basic project details below and click 'Update' to save.

Plan name: My plan (Horizon 2020 DMP)

ID:

Grant number:

Principal Investigator/Researcher: Emilie Hemans

Principal Investigator/Researcher ID:

Plan data contact:

Description:

Save Cancel

This plan is based on:

Funder | European Commission (Horizon 2020)

## Versions

## Share with partners

Plan details **Initial DMP** Detailed DMP Final review DMP Share Export

### 1. Data summary (1 question, 0 answered)

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

**B** *I* [List] [List] [Link] [Table]

Save

Not answered yet

**2. FAIR data (4 questions, 0 answered)** +

**3. Allocation of resources (1 question, 0 answered)** +

**4. Data security (1 question, 0 answered)** +

**5. Ethical aspects (1 question, 0 answered)** +

**6. Other (1 question, 0 answered)** +

Guidance Share note

DCC guidance on Existing Data +

DCC guidance on Data Volumes +

DCC guidance on Data Format +

# Guidance

**2. FAIR data** (4 questions, 0 answered)

In general terms, your research data should be 'FAIR' (that is findable, accessible, interoperable and re-usable). These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

**B** *I* [List icons] [Link icon] [Grid icon]

**Save**

**Not answered yet**

**Guidance** **Share note**

**EC Guidance**

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

**DCC guidance on Metadata**

**Questions to consider:**

- How will you capture / create the metadata?
- Can any of this information be created automatically?
- What metadata standards will you use and why?

**Guidance:**

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a 'readme' text file, in file headers etc.

Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a [catalogue of disciplinary metadata standards](#).

**DCC guidance on Documentation**

Guidance based on guidelines EC

Guidance and links from DCC



## Export to various formats

### My plan (Horizon 2020 DMP)

Plan details Initial DMP Detailed DMP Final review DMP Share **Export**

From here you can download your plan in various formats. This may be useful if you need to submit your plan as part of a grant application. Select what format you wish to use and click to 'Export'.

#### Initial DMP —

Format

pdf	Export
csv	
html	
json	
pdf	
text	
xml	
docx	

#### Detailed DMP +

#### Final review DMP +

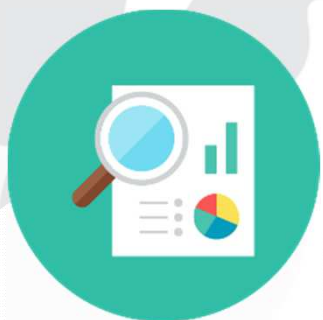
# RECOMMENDATIONS



## Data summary

# Data collection

- Origin: generated, collected, reused
  - Does similar data exist? What about reintegration or reuse?
  - Re-use? Provide the source and check IPR
- **Types** e.g. digital/non-digital data, qualitative/quantitative, audio files, surveys, databases, field notes.....



## Data summary

# Data files formats

Use data formats that are:

- Open standard possible?
- In an easily re-usable format
- Commonly used by research community

Use consequent naming convention

Structured organizing of files

### Examples of preferred format choices:

Text	.odt, .txt, .xml, .html, .rtf
Tabular Data	.csv (comma separated values), .xml, .rdf, .SPSS portable
Images	.tif, .jpeg2000, .png, .svg,
Structured data	.xml, .rdf

**Any standard used in your field**

Accessible

# Documenting data

- Make your data understandable: project level (context) and data level (e.g. codebooks, protocol)
- Be clear what methods you use
  - E.g. lab notebook, end-to-end code/scripts for statistics
  - Software can help: R, MatLab, Python...

**Findable**

# Create searchable data

## Using metadata

- Data about data
- Consists of set of attributes
- Machine readable
- Helps prevent inappropriate use
- Use metadata standards of your domain

Accessible

# Where to deposit data?

## Research data repository

- Matches data needs
- Disciplinary/Institutional data repository
- Directory of data repositories:  
[www.Re3data.org](http://www.Re3data.org)
- [Zenodo](https://zenodo.org/)





# Re3data.org

Search for Repositories (1314 reviewed repositories)

fish ✕ 🔍 Search

▼

**Subject** **Content Type** **Country** (of the responsible institutions)

Add subjects  Add content types  Add countries

Certificates  Open Access  Persistent Identifier

Reset filter

42 results ( 1 – 25 ) « 1 2 » Sort by Weight ▼

**e-Atlas**  
Australia's Tropical Land and Seas

📄 🔒 © 🌐 🇺🇸 🇸🇩

Subjects: Geography Geosciences (including Geography) Human Geography Natural Sciences

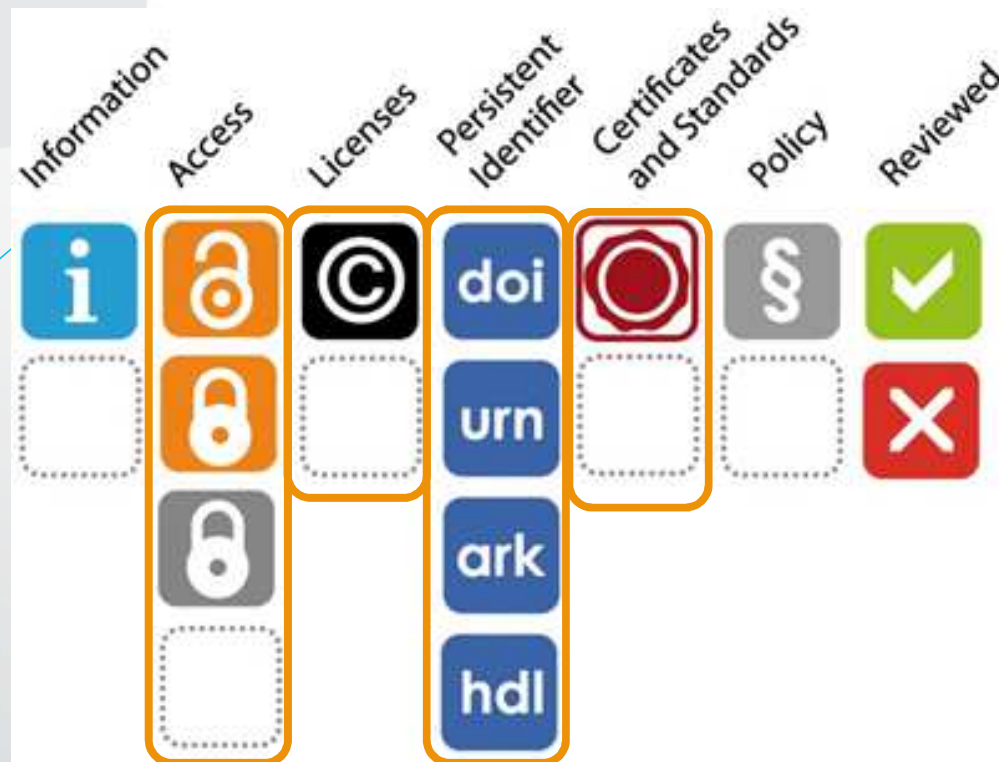
Content types: Archived data Databases Images Plain text Raw data Scientific and statistical data formats Software applications  
Standard office documents Structured graphics Structured text other

Countries: Australia

The eAtlas is a website, mapping system and set of data visualisation tools for presenting research data in an accessible form that promotes greater use of this information. The eAtlas will serve as the primary data and knowledge repository for all NERP Tropical Ecosystems Hub projects, which focus on the on the Great Barrier Reef, Wet Tropi... more »



# Re3data



- Access
- Licenses
- Persistent identifier



→ Trustworthy Digital repository

# What to deposit?

Everything needed to validate results presented in scientific publications



## DATA

- Validate results
- Selection



## METADATA



## DOCUMENTATION

- Tools, software....
- Read\_me file?

OPEN DATA?

## Open data

- Keep it simple
- Apply an open license:

- e.g. creative commons:



- Data repositories can provide licenses
    - [Re3data.org](https://re3data.org)



The terms of use and licenses of the data are provided by the research data repository.

# Example

Understandable  
for humans

Machine readable  
metadata

Tools

Open Data

Open license

## DOI

DOI: 10.5281/zenodo.46266

## Keyword(s):

Integrative Modeling Platform (IMP) Chemical crosslinks Electron microscopy class average MODELLER PMI

## Published in:

Mol Cell Proteomics: 13 (2014) no. 11, pp. 2927-43

## Related publications and datasets:

Supplement to:

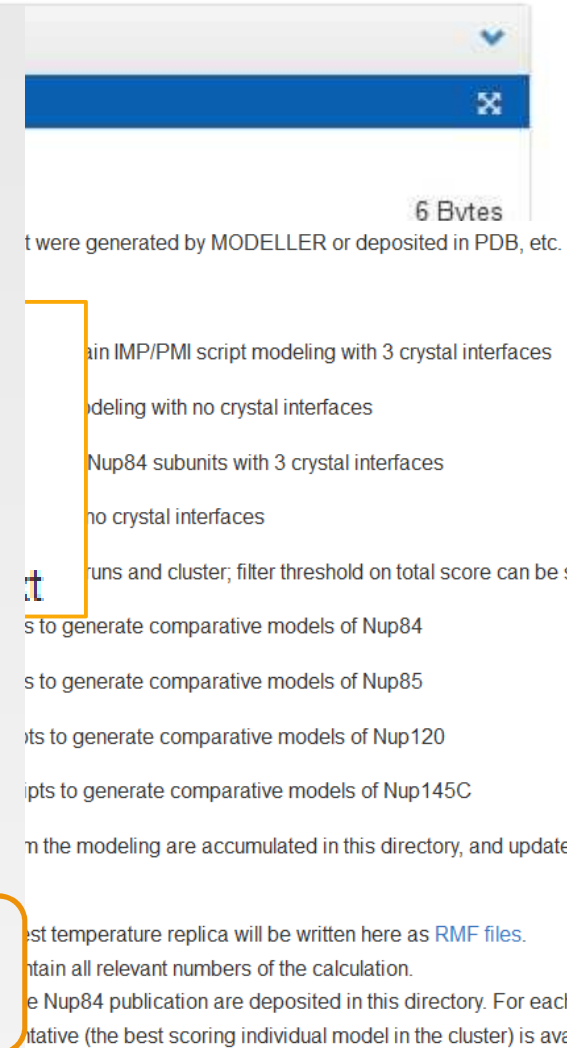
<https://github.com/integrativemodeling/nup84/tree/v1.0>, 10.1074/mcp.M114.041673

## Collections:

Communities > Integrative Modeling  
Communities > Sali Lab at UCSF  
Datasets  
Open Access

## License (for files):

GNU Library or "Lesser" General Public License (LGPL)



6 Bytes

t were generated by MODELLER or deposited in PDB, etc.

ain IMP/PMI script modeling with 3 crystal interfaces

odeling with no crystal interfaces

Nup84 subunits with 3 crystal interfaces

no crystal interfaces

t runs and cluster; filter threshold on total score can be s

s to generate comparative models of Nup84

s to generate comparative models of Nup85

ts to generate comparative models of Nup120

pts to generate comparative models of Nup145C

n the modeling are accumulated in this directory, and updated

est temperature replica will be written here as RMF files.

tain all relevant numbers of the calculation.

e Nup84 publication are deposited in this directory. For each

ative (the best scoring individual model in the cluster) is ava

RMF format, together with the top five best scoring models in PDB format. An accompanying stat file contains us statistics on the simulation, such as whether each of the crosslinks was satisfied.



# Zenodo



For all content types

Type

- Publication (37312) +
- Software (11356)
- Image (3412) +
- Dataset (3163)
- Presentation (1359)
- Poster (741)
- Video (209)
- Lesson (160)

Create communities

Search bar with 'Upload' and 'Communities' buttons. 'Communities' is highlighted with an orange box. 'Log in' and 'Sign up' buttons are also visible.

and secure storage architecture  
 han; Niazi, Khan; Hakimzadeh, Kamal; Parák, Boris  
 pts that must be built into BBMRI-ERIC IT services by design, as trust and  
 medical research infrastructure. This document focuses on providing a  
 ementing IT services of ...

View

ndanao Island, southern Philippines, II: the  
 ao and adjacent islands  
 meron D.; Diesmos, Arvin C.; Alcalá, Angel C.; Brown, Rafe M.  
 reptile species distribution data from the northeast Mindanao faunal region,  
 s subcenter of endemic vertebrate biodiversity. Together with all publicly  
 rsity ...

View

October 12, 2016 Journal article Open Access

View

Open Science in H2020  
 Complete genome sequence of *Methanoculleus bourgensis* strain MAB1, the syntrophic  
 partner of mesophilic acetate-oxidising bacteria (SAOB)

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

With GitHub Integration

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!





# OpenAIRE

www.openaire.eu/search

## HERMIONE<sup>SC39</sup>

**Title** Hotspot Ecosystem Research and Man's Impact on European seas

**Funding** EC | FP7 | SP1 | ENV

**Call** FP7-ENV-2008-1

**Contract (GA) number** 226354

**Start Date** 2009/04/01

**End Date** 2012/09/30

**Open Access mandate** yes

**Special Clause 39** yes

**Organizations** UCC, UPMC, WCMC, UGent, AWI, UniHB, CSIC, CNRS-LN2, UNIVERSITY OF THESSALY - UTH, UB, UNEP, MPG, IFREMER, UAzores, SOUTHAMPTON, UGOT, JacobsUni, MEDIAN, HELLENIC CENTRE FOR MARINE RESEARCH, HWU, SIO, HAVFORSKNINGSINSTITUTTET, THE SCOTTISH ASSOCIATION FOR MARINE SCIENCE, ArchimediX, NIOZ, IH, National Marine Aqua, UAVR, FAU, CARDIFF UNIVERSITY, CNR, Acquario di Genova, GEOMAR, NERC, SENCKENBERG GESELLSCHAFT FUR NATURFORSCHUNG, NUI GALWAY, UIT, CONISMA, KNAW, UNIABDN, LIV

**More information** [Detailed project information \(CORDIS\)](#)

### SHARE - BOOKMARK



### APP BOX

- [Publication details](#)
- [Dynamically incorporate publications in your site \(HTML\)](#)
- [View EC progress report \(HTML\)](#)
- [Download EC progress report \(CSV\)](#)

[LINK RESEARCH RESULTS](#)

[DEPOSIT PUBLICATIONS](#)

Link your data to publications or project

## Mats of psychrophilic thiotrophic bacteria associated with cold seeps of the Barents Sea

Grünke, S.; A. Lichtschlag; de Beer, D.; Felden, J.; Salman, V.; A. Ramette; H. N. Schulz-Vogt; A. Boetius (2012)

**Publisher:** Copernicus GmbH

**Journal:** Biogeosciences

**Types:** Article

**Subjects:** Biology (General), Q, DOAJ:Earth Sciences, DOAJ:Biological Sciences, DOAJ:Earth and Environmental Sciences, Geology, QE1-996.5, DOAJ:Biological and Life Sciences, QH301-705.5, Science

**Identifiers:** [doi:10.5194/bgd-9-3917-2012](#), [doi:10.5194/bg-9-2947-2012](#)

This study investigated the bacterial diversity associated with microbial mats of polar deep-sea cold seeps. The mats were associated with high upward fluxes of sulfide produced by anaerobic oxidation of methane, and grew at temperatures close to the freezing point of seawater. They ranged from small patches of 0.2–5 m in diameter (gray mats) to extensive fields covering up to 850 m<sup>2</sup> of seafloor (white mats) and were formed by diverse sulfide-oxidizing bacteria differing in color and size. Overall, both the dominant mat-forming thiotrophs as well as the bacterial communities inhabiting the mats differed in composition as determined by microscopy, 16S rRNA gene sequencing, and ribosomal intergenic spacer analysis. While the smaller mats were composed of only 1–2 types of gliding *Beggiatoa* filaments, molecular analyses showed that most of the dominant mat-forming sulfide oxidizers were phylogenetically different from, but still closely related to, thiotrophs known from warmer ocean realms. The psychrophilic nature of the polar mat-forming thiotrophs was tested by visual observation of active mats at in situ temperature compared to their warming to >4 °C. The temperature range of mat habitats and the variation of sulfide and oxygen fluxes appear to be the main factors supporting the diversity of mat-forming thiotrophs in cold seeps at continental margins.

[LINK TO PROJECT](#)

[LINK TO RESEARCH DATA](#)

Publications (360)

Research Data (478)

Statistics



# OpenAIRE

Training and support material

[www.openaire.eu/opendatapilot](http://www.openaire.eu/opendatapilot)

Information on:

- Open research data pilot
- Creating a data management plan
- Selecting a data repository
- Dealing with personal data

Support material:

Briefing papers, factsheets, webinars, workshops, FAQs, helpdesk

The collage features several documents from the OpenAIRE Horizon 2020 project. At the top left is the 'OpenAIRE Horizon 2020 FactSheets' document, which includes a section titled 'Open Research Data Pilot in Horizon 2020: How can OpenAIRE help?'. Below this is a 'Briefing paper for Researchers, Project coordinators, and Research Managers participating in the EC Open Research Data Pilot in Horizon 2020'. Another document is titled 'Personal data and the Open Research Data Pilot: How can OpenAIRE help?'. A 'Do you have a grant?' section is also visible, listing various research areas like 'Research Infrastructures', 'Future and Emerging Technologies', and 'Societal Challenge: Secure, Clean and Efficient Energy'. At the bottom right, there is a 'Briefing paper for researchers, research admin in institutions and project coordinators' dated February 2016. The footer of the collage includes the 'Horizon-EMEREA-00011 Topic e-INFRAstructure for Open Access Research & Innovation action Grant Agreement 64910' logo.

## STEP 1

WRITE A DMP  
dmponline.dcc.ac.uk



Update at

- 6 months
- Periodic evaluation
- Final review

## STEP 2

FIND REPOSITORY  
Matches data needs



Data Repositories

- discipline/institutional
- [www.re3data.org](http://www.re3data.org)
- Zenodo

## STEP 3

DEPOSIT DATA  
(Open) Data  
Metadata  
Other tools



- Standard File Formats
- Standards metadata schema
- (Open) Licences

## SUPPORT

Supporting  
infrastructure and  
information



- EC guidelines
- [OpenAIRE.eu](http://OpenAIRE.eu)
- [dcc.ac.uk](http://dcc.ac.uk)



# Questions!



[www.openaire.eu](http://www.openaire.eu)



[@openaire\\_eu](https://twitter.com/openaire_eu)



[Facebook.com/groups/openaire](https://www.facebook.com/groups/openaire)



<https://www.linkedin.com/groups/OpenAIRE3893548>



[Emilie.Hermans@UGent.be](mailto:Emilie.Hermans@UGent.be)  
[info@openaccess.be](mailto:info@openaccess.be)