

Research Data Management at CERN

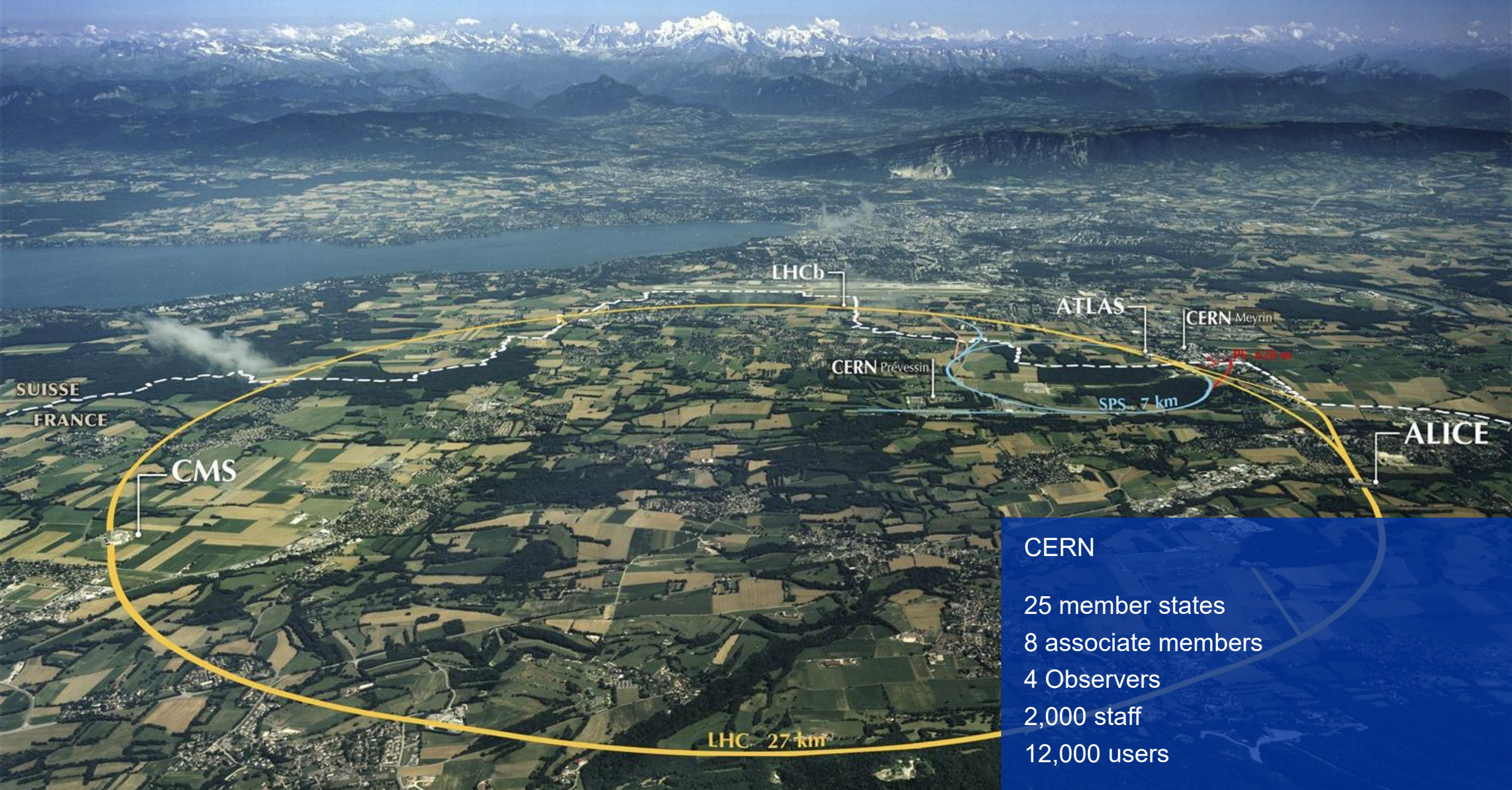
Pablo Saiz

10 September 2025



**“FAIR data?
I do not have time
for that.
And it is not worth
the effort”**





CERN

- 25 member states
- 8 associate members
- 4 Observers
- 2,000 staff
- 12,000 users



CERN Open Science



Open Science

Open Science AREAS

Open Access

Open Data

Open Code

Open Educational Resources

Open Methodology

Open Peer Review

Open Economics Guide (<https://openeconomics.zbw.eu/en/>)

Creative Commons Licences

BY: Attribution of author

NC: Non-commercial
(Prevents commercial use of the work)

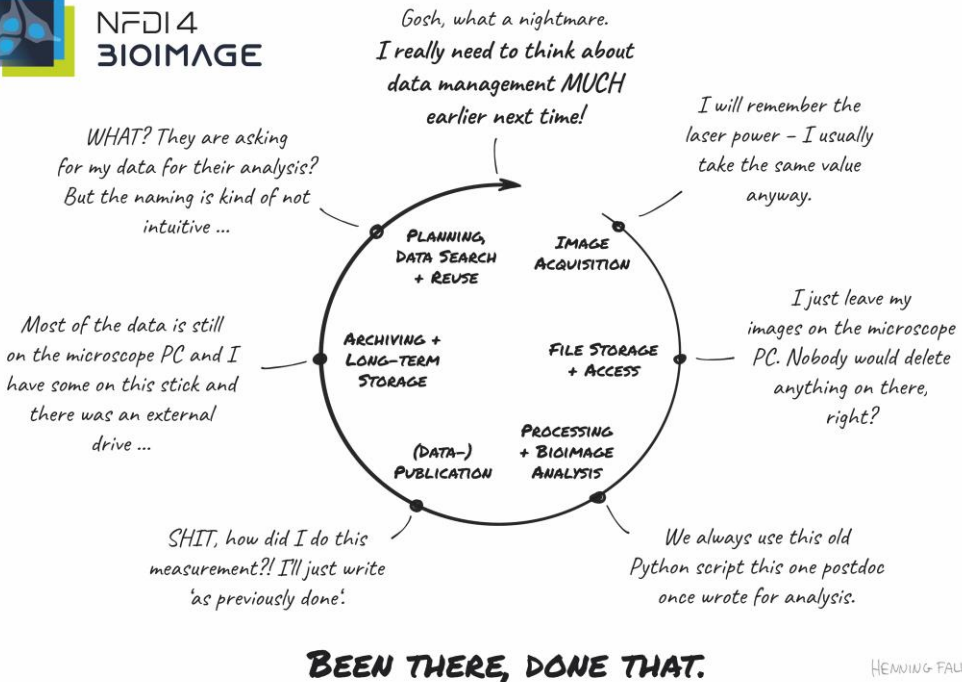
ND: No Derivatives
(No editing of the work allowed)

SA: Share-alike
(The work may be modified, but must be republished under the same licence.)

My article shall ...
... be free to use
... nobody should earn money with it and
... changes must be republished under the same licence.

Open Economics Guide (<https://openeconomics.zbw.eu/en/>)

Research Data Management Lifecycle



NFDI4BIOIMAGE Consortium (2024): NFDI4BIOIMAGE data management illustrations by Henning Falk, Zenodo, <https://doi.org/10.5281/zenodo.14186100>, is used under a CC BY 4.0

Data Management Plan

Data Management Plan THE ELEMENTS

What data are produced and used?

What is the data type?

What data should or must be stored and why?

Which supplementary information is necessary to understand the data?

When does the data selection take place?

How long should the data be stored?

When and where are the data transferred?

Who is allowed to use the data?

Open Economics Guide (<https://openeconomics.zbw.eu/en>)

Data Management Plans

Data Management Plans have become a requirement for projects, grants, stipends etc around the world, i.e. in CERN's member states. With that many in the CERN community are facing the challenge to regularly write or update a Data Management Plan (DMP). In a data driven environment like CERN, such documentation can also be considered essential to ensure the longevity of project and research results.

CERN provides templates for DMPs as well as support and guidance for anyone with questions or concerns regarding their DMPs.

You can find the CERN Data Management Plan template [here](#).

If you seek support please contact dmp-support@cern.ch.

FAIR principles



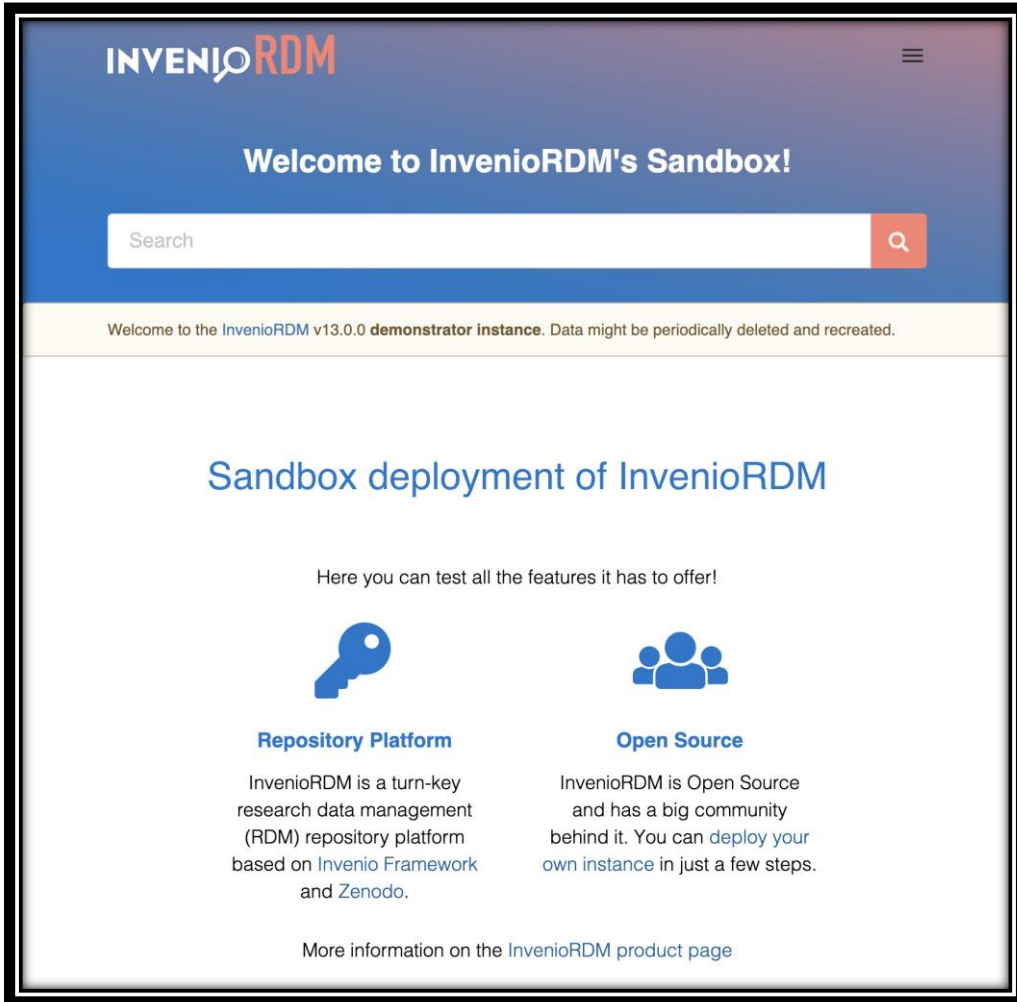
The Turing Way Community, & Scriberia. (2020). Illustrations from the Turing Way book dashes. Zenodo. <https://doi.org/10.5281/zenodo.3695300>

FAIR: Findable

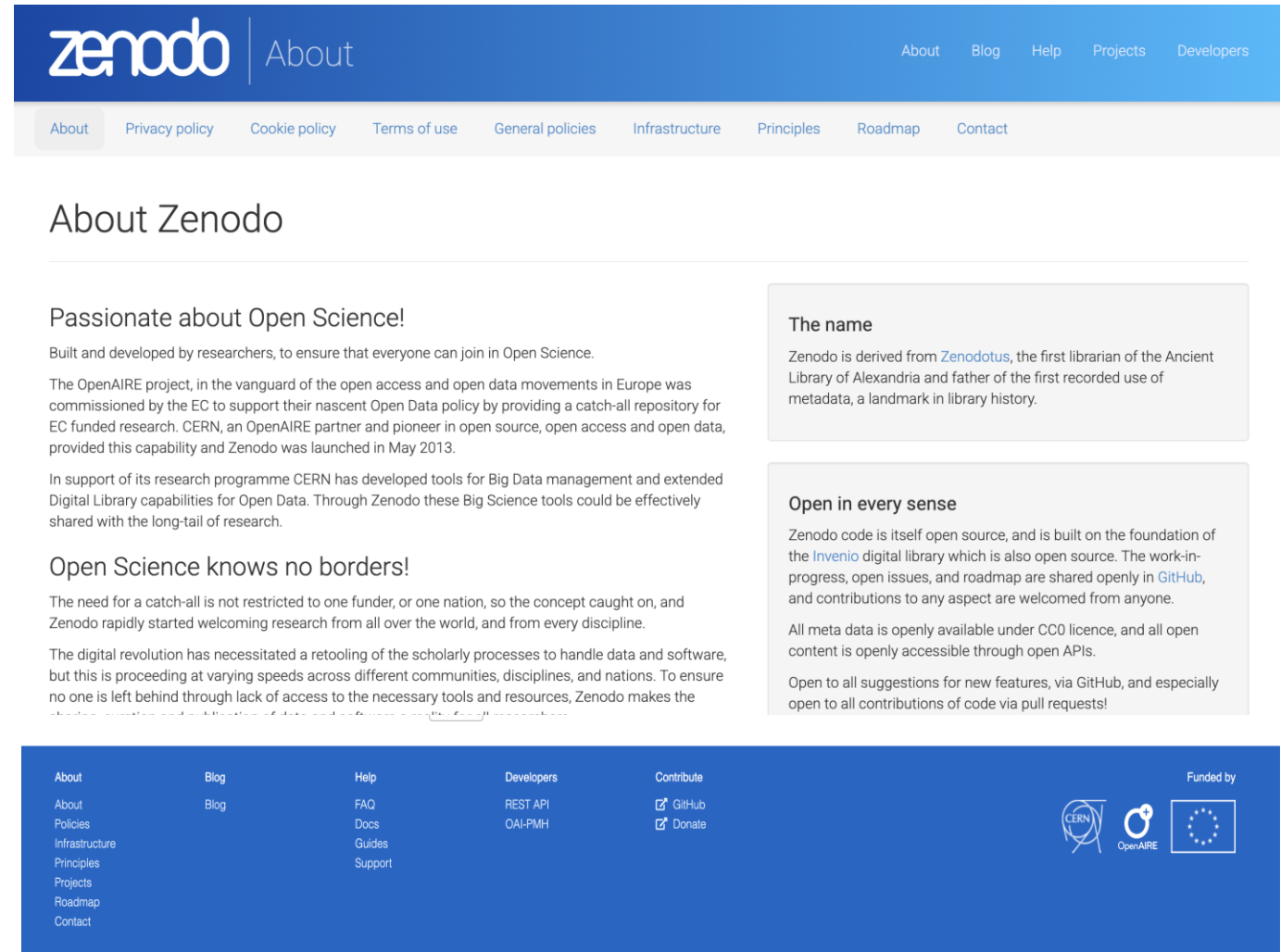


**Digital
Object
Identifier**

Digital repositories: Invenio RDM /ZENODO



The screenshot shows the InvenioRDM website interface. At the top left is the 'INVENIO RDM' logo. Below it is a navigation menu with 'About', 'Blog', 'Help', 'Projects', and 'Developers'. A search bar is present with the text 'Search' and a magnifying glass icon. A yellow banner reads: 'Welcome to the InvenioRDM v13.0.0 demonstrator instance. Data might be periodically deleted and recreated.' The main heading is 'Sandbox deployment of InvenioRDM'. Below this, a text block says: 'Here you can test all the features it has to offer!'. There are two columns of content. The first column has a key icon, the heading 'Repository Platform', and text: 'InvenioRDM is a turn-key research data management (RDM) repository platform based on Invenio Framework and Zenodo.' Below this is a link: 'More information on the InvenioRDM product page'. The second column has a group of people icon, the heading 'Open Source', and text: 'InvenioRDM is Open Source and has a big community behind it. You can [deploy your own instance](#) in just a few steps.'



The screenshot shows the Zenodo website. At the top left is the 'zenodo' logo and an 'About' link. To the right is a navigation menu with 'About', 'Blog', 'Help', 'Projects', and 'Developers'. Below this is a secondary navigation menu with 'About', 'Privacy policy', 'Cookie policy', 'Terms of use', 'General policies', 'Infrastructure', 'Principles', 'Roadmap', and 'Contact'. The main heading is 'About Zenodo'. Below this is a section titled 'Passionate about Open Science!'. The text reads: 'Built and developed by researchers, to ensure that everyone can join in Open Science. The OpenAIRE project, in the vanguard of the open access and open data movements in Europe was commissioned by the EC to support their nascent Open Data policy by providing a catch-all repository for EC funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability and Zenodo was launched in May 2013. In support of its research programme CERN has developed tools for Big Data management and extended Digital Library capabilities for Open Data. Through Zenodo these Big Science tools could be effectively shared with the long-tail of research.' Below this is a section titled 'Open Science knows no borders!'. The text reads: 'The need for a catch-all is not restricted to one funder, or one nation, so the concept caught on, and Zenodo rapidly started welcoming research from all over the world, and from every discipline. The digital revolution has necessitated a retooling of the scholarly processes to handle data and software, but this is proceeding at varying speeds across different communities, disciplines, and nations. To ensure no one is left behind through lack of access to the necessary tools and resources, Zenodo makes the [structure, content and publication of data and software](#) open to all.' To the right of the main text are two boxes. The first is titled 'The name' and contains the text: 'Zenodo is derived from [Zenodotus](#), the first librarian of the Ancient Library of Alexandria and father of the first recorded use of metadata, a landmark in library history.' The second is titled 'Open in every sense' and contains the text: 'Zenodo code is itself open source, and is built on the foundation of the [Invenio](#) digital library which is also open source. The work-in-progress, open issues, and roadmap are shared openly in [GitHub](#), and contributions to any aspect are welcomed from anyone. All meta data is openly available under CC0 licence, and all open content is openly accessible through open APIs. Open to all suggestions for new features, via [GitHub](#), and especially open to all contributions of code via pull requests!'. At the bottom of the page is a footer with a grid of links: 'About', 'Blog', 'Help', 'Developers', 'Contribute'. Under 'About' are 'About', 'Policies', 'Infrastructure', 'Principles', 'Projects', 'Roadmap', 'Contact'. Under 'Blog' are 'Blog'. Under 'Help' are 'FAQ', 'Docs', 'Guides', 'Support'. Under 'Developers' are 'REST API', 'OAI-PMH'. Under 'Contribute' are 'GitHub', 'Donate'. On the right side of the footer are logos for 'CERN', 'OpenAIRE', and 'Funded by' with the European Union flag.

ZENODO Communities

The screenshot shows the Zenodo Help page. At the top, there is a navigation bar with the Zenodo logo and 'Help' text. Below this, there are links for 'About', 'Blog', 'Help', 'Projects', and 'Developers'. A secondary navigation bar contains 'Help', 'FAQ', 'Docs', and 'Guides'. The main content area is titled 'Frequently Asked Questions' and includes a search bar. Below the search bar, there are four columns of links: 'General', 'Account', 'Upload/Deposit', and 'Collaborate and share'. A 'Tip' section follows, providing information about OpenAIRE. Below the tip, there are four columns of links: 'Get started', 'Deposit', 'Collaborate and share', and 'Communities'. The 'Communities' link is circled in red. Below this, there are three columns of links: 'Profile', 'GitHub and Software', and 'About'.

Docs / Communities / About communities

About communities

Communities are shared areas on Zenodo where projects, institutions, domains, and conferences can curate and manage their research outputs.

About communities

Your project, institution, department, domain or conference can collaborate on Zenodo using our communities feature, which serves as an shared area curated and managed by your team.

Each person always login on Zenodo using their personal account. A person can create a new community. Once the community is created, you can invite an unlimited number of people to your community. Each person is given a role below which grants them different levels of access to the community and its records. A member of the community can hold one of the following roles:

Roles	Reader	Curator	Manager	Owner
Can view files in restricted records	X	X	X	X
Can accept/decline submissions	-	X	X	X
Can edit records metadata	-	X	X	X
Can invite members	-	-	X	X
Can manage members	-	-	X	X
Can modify community settings	-	-	-	X
Can delete community	-	-	-	X

Any user on Zenodo can submit records for inclusion in a community. The community curators are responsible for curating the records and finally accept/decline the record into the community.

ZENODO for EURAD-2

zenodo Search records... Communities My dashboard

eurad2 European Partnership on Radioactive Waste Management

European Partnership on Radioactive Waste Management - 2

Part of EU Open Research Repository <https://www.ejp-urad.eu/> ANDRA and 27 more organizations

New upload

Records Members

1 results found Sort by Newest

Versions

View all versions

Access status

Open

Resource types

September 9, 2025 (v1) Project deliverable Open

EURAD-2 Data management plan

Miron, George Dan ; DEBAYLE, Christophe ; Kudriashova, Yevheniia

No description

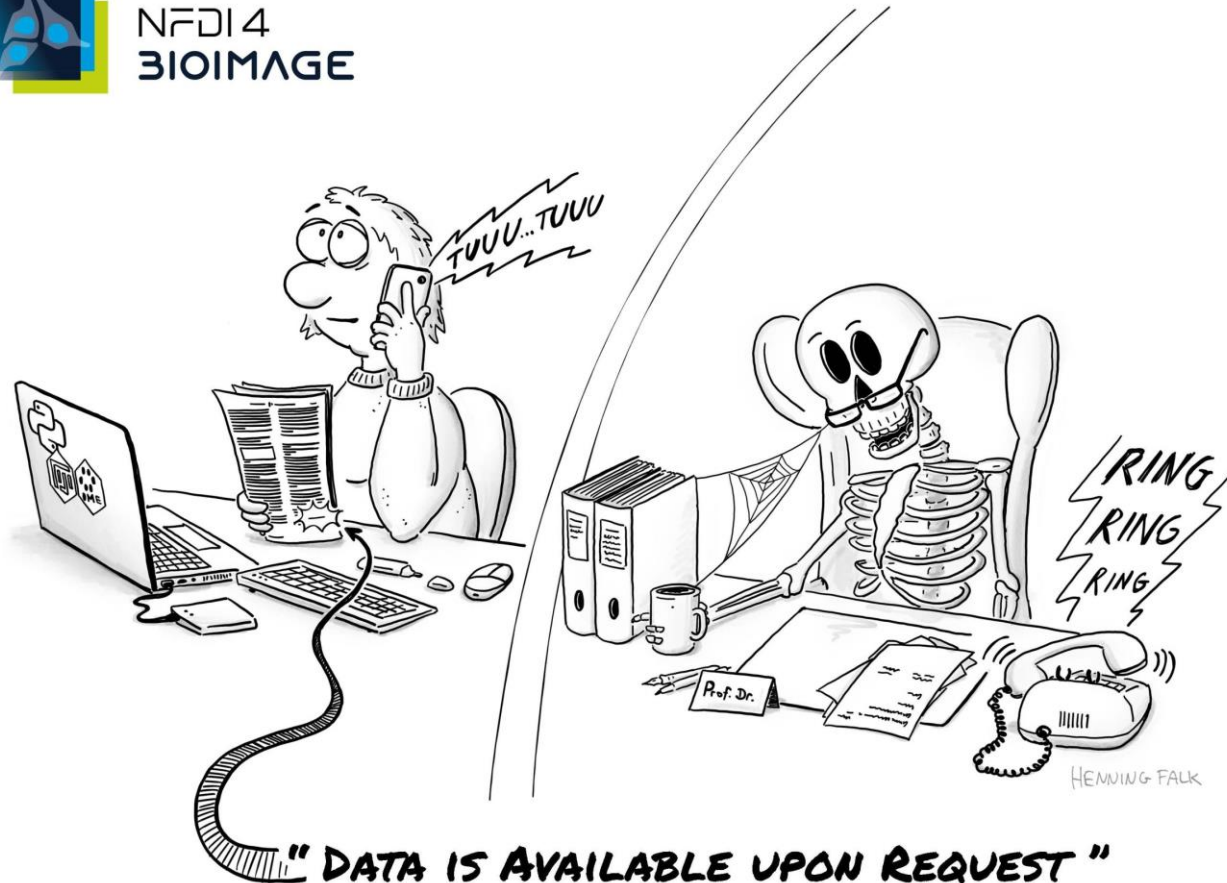
Part of EU Open Research Repository , European Partnership on Radioactive Waste Management - 2

Uploaded on September 9, 2025

23 3

10 results per page

FAIR: Accesible



" DATA IS AVAILABLE UPON REQUEST "

This cartoon is part of a series of post cards raising awareness for the work of the NFDI4BIOIMAGE consortium. Published under [a CC BY 4.0 licence](#).

Cold Storage for CERN Open Data

opendata
CERN

Help

Explore more than **five petabytes** of open data from particle physics!

Search

search examples: [collision datasets](#), [keywords:education](#), [energy:7TeV](#)

Simulated dataset ZZTo4L_TuneCP5_13TeV_powheg_pythia8 in MINIAODSIM format for 2016 collision data

ZZTo4L_TuneCP5_13TeV_powheg_pythia8RunIISummer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM_CMS_Collaboration

Cite as: CMS Collaboration (2024). Simulated dataset ZZTo4L_TuneCP5_13TeV_powheg_pythia8 in MINIAODSIM format for 2016 collision data. CERN Open Data Portal. DOI:10.7434/OPENDATA.CMS.MNIAODSIM

Description

Simulated dataset ZZTo4L_TuneCP5_13TeV_powheg_pythia8 in MINIAODSIM format for 2016 collision data. See the description of the simulated dataset names in: [About CMS simulated dataset names](#). These simulated datasets correspond to the collision data collected by the CMS experiment in 2016.

Cross section

For pp collisions at 13TeV, this sample has a cross section of $(1.325e+00 \pm 1.220e-03)$ pb, calculated using the method described in [CMS guide for cross sections](#). This cross section takes into account a matching efficiency of 1.0 and a filtering efficiency of 1.000e+00, based on generator settings and/or filters. If this sample was generated at NLO, it has 5.057e-03% events with negative weights. This is a leading-order value, and cross sections evaluated at next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) should be used.

Related datasets

The corresponding MINIAODSIM dataset: [ZZTo4L_TuneCP5_13TeV_powheg_pythia8RunIISummer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM](#)

Dataset characteristics

52137000 events, 604 files, 1.7 TiB in total.

System details

Recommended [global tag](#) for analysis: 100X_mcRun2_asymptotic_v17
Recommended release for analysis: CMS16_10_8_30
Recommended container images for analysis is available in the following locations ([see guide](#)):

- `docker://cernopendata/cosw_18_6_38-slc7_and64_gcc780:latest`
- `gitlab-registry.cern.ch/cos-cs-cs/cosw-docker-opendata/cosw_18_6_38-slc7_and64_gcc780:latest`

How were these data generated?

These data were generated in several steps (see also CMS Monte Carlo production overview):

Step LHE GEN

Release: CMS16_10_8_22
Global tag: 100X_mcRun2_asymptotic_v13
Generators: powheg ZZ

Files and indexes

Availability: PARTIAL [Request all files](#)

Please note that only a subset of files are available for this dataset. If you are interested in accessing all of them, please request them. Note that to online storage may take several weeks or months in case of a large amount of data.

Index description	Index size	Availability
CMS_mc_RunII_Summer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM_100X_mcRun2_asymptotic_v17-v1_2430000_file_index.json	7.5 GB	Partial [] List files
CMS_mc_RunII_Summer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM_100X_mcRun2_asymptotic_v17-v1_2520000_file_index.json	262.8 GB	On demand [] List files
CMS_mc_RunII_Summer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM_100X_mcRun2_asymptotic_v17-v1_2530000_file_index.json	304.5 GB	On demand [] List files
CMS_mc_RunII_Summer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM_100X_mcRun2_asymptotic_v17-v1_40000_file_index.json	275.9 GB	On demand [] List files Download files
CMS_mc_RunII_Summer20UL16MiniAODv2_ZZTo4L_TuneCP5_13TeV_powheg_pythia8_MNIAODSIM_100X_mcRun2_asymptotic_v17-v1_50000_file_index.json	879.9 GB	On demand [] List files Download files

Disclaimer

These open data are released under the Creative Commons Zero v1.0 Universal license.

PUBLIC DOMAIN

Neither the experiment(s) | CMS | nor CERN endorse any works, scientific or otherwise, produced using these data. This release has a unique DOI that you are requested to cite in any applications or publications.

Request to make data available

Please confirm you want to request all files of the record.

This action takes time. The more data requested, the longer it will take.

I confirm that I want to request 604 files (1.7 TiB of data)

If you want to be notified, enter your email

Enter your email

Cancel OK

Long-term preservation

Experiment

- ALICE (16)
- ATLAS (130)
- CMS (51,955)
- DELPHI (25,458)
- JADE (1)
- LHCb (103)
- OPERA (904)
- PHENIX (1)
- TOTEM (1)

Availability

- online (16,346)

Year

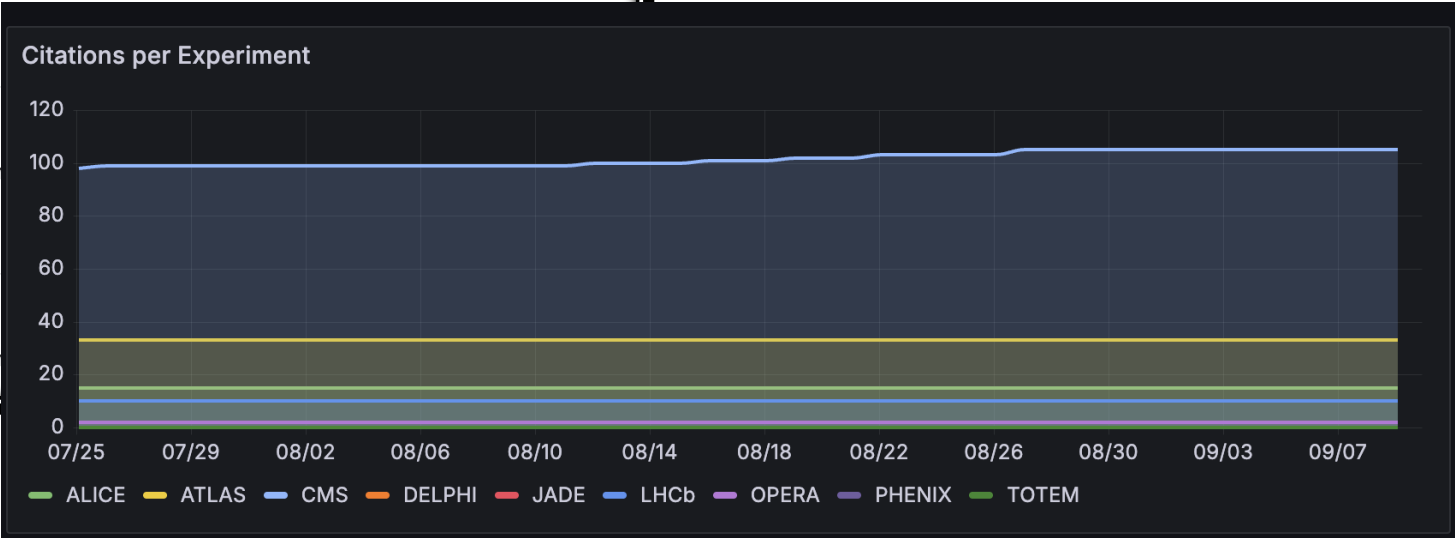
DELPHI simulation data xs_hzha03pyth6156hcee_e206.5_m47.5_ka0_1l_e1
Extended Short DST simulation a0_e1 done at ecms=206.5 , Karlsruhe

DELPHI simulation data xs_hzha03pyth6156hcee_e206.5_m47.5_ka0_1l_e1
Extended Short DST simulation a0_e1 done at ecms=206.5 , Karlsruhe

DELPHI simulation data xs_hzha03pyth6156hcqq_e206.5_m102.5_la0_1l_u1
Extended Short DST simulation a0_u1 done at ecms=206.5 , Lyon

DELPHI simulation data xs_hzha03pyth6156hcqq_e206.5_m102.5_la0_1l_u1
Extended Short DST simulation a0_u1 done at ecms=206.5 , Lyon

DELPHI simulation data xs_hzha03pyth6156hcqq_e206.5_m102.5_la0_1l_u1
Extended Short DST simulation a0_u1 done at ecms=206.5 , Lyon



FAIR: Interoperable



Online illustrations by Storyset



FAIR: Reusable



News Feature | Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

[Monya Baker](#)

[Nature](#) 533, 452–454 (2016) | [Cite this article](#)

266k Accesses | 3290 Citations | 5216 Altmetric | [Metrics](#)

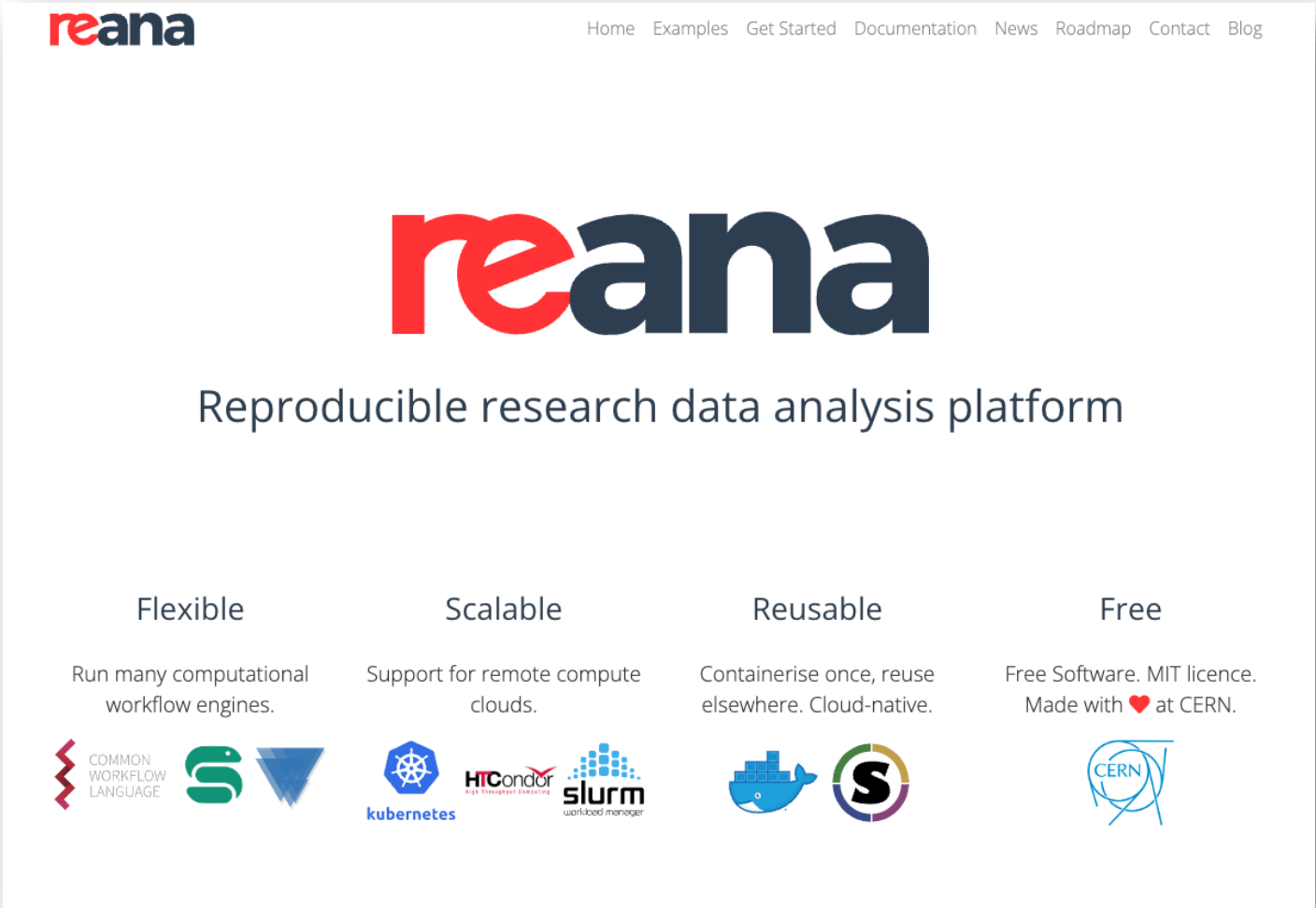
[i](#) This article has been [updated](#)

Survey sheds light on the 'crisis' rocking research.

More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature's* survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

The data reveal sometimes-contradictory attitudes towards reproducibility. Although 52% of those surveyed agree that there is a significant 'crisis' of reproducibility, less than 31% think that failure to reproduce published results means that the result is probably wrong, and most say that they still trust the published literature.

Reproducible research data analysis












The screenshot shows the reana website homepage. At the top left is the reana logo. To its right is a navigation menu with links: Home, Examples, Get Started, Documentation, News, Roadmap, Contact, and Blog. The main heading is 'reana' in a large, bold font, with 're' in red and 'ana' in dark blue. Below this is the subtitle 'Reproducible research data analysis platform'. The page is divided into four columns, each representing a key feature: Flexible, Scalable, Reusable, and Free. Each column contains a brief description and logos of associated technologies or partners.

reana

Home Examples Get Started Documentation News Roadmap Contact Blog

reana

Reproducible research data analysis platform

Flexible	Scalable	Reusable	Free
Run many computational workflow engines.	Support for remote compute clouds.	Containerise once, reuse elsewhere. Cloud-native.	Free Software. MIT licence. Made with ❤️ at CERN.
 COMMON WORKFLOW LANGUAGE  	  	 	

More sources

Open Economics

- 01 Meet the Open Science policies of your institution and your research funding agencies.
- 02 Create an ORCID iD for yourself.
- 03 Make sure that as much of your research as possible is assigned a DOI.
- 04 Get an overview of Open Access journals in your specialist subject.
- 05 Consider which parts of your research process you can open up.

Open Economics Guide (<https://openeconomics.zbw.eu/en>)

English Dansk

Availability	Discipline	FAIR dimension	Phases in research life cycle	Service Name	Description
International	Generic (Tabular data)	__I	Process/Analyze	OpenRefine	OpenRefine is a standalone open source desktop application for data cleanup and transformation to other formats (i.e. data wrangling)
International	Generic	FA(I)(R)	Publish/Disseminate, Archive, Discover & Re-use, Release, Preserve	Zenodo	Zenodo is a general-purpose open access research data repository, hosted by CERN (Switzerland) that provides a place for researchers to deposit datasets. Researchers in any subject area to are able to upload files up to 50 GB. It has an integration with GitHub to make code hosted in GitHub citable. Support: zenodo.org/support
International	Generic - Highly recognized in Social Sciences	FA(I)(R)	Publish/Disseminate, Archive, Discover & Re-use, Release, Preserve	Harvard Dataverse	Dataverse is a data repository that is widely used within the Social Sciences. Researchers can login with their institutional credentials via WAYF. Data can be made findable by applying discipline-specific metadata schemes and digital object identifiers (DOIs). Data is made reusable by specifying relevant re-use licenses.

FAIR for beginners

“FAIR data?
I do not have time
for that.
And it is not worth
the effort”

Let's talk about FAIR data...

“Making
your data FAIR
is a gradual process
with small steps one at a time.
In the end, it will become
best practice in your
research field”



www.vidensportal.deic.dk/FAIR

FAIR data are: Findable, Accessible, Interoperable and Reusable.

Making your data FAIR means maximizing the project's output, increasing your impact and enhancing your recognition as a researcher.

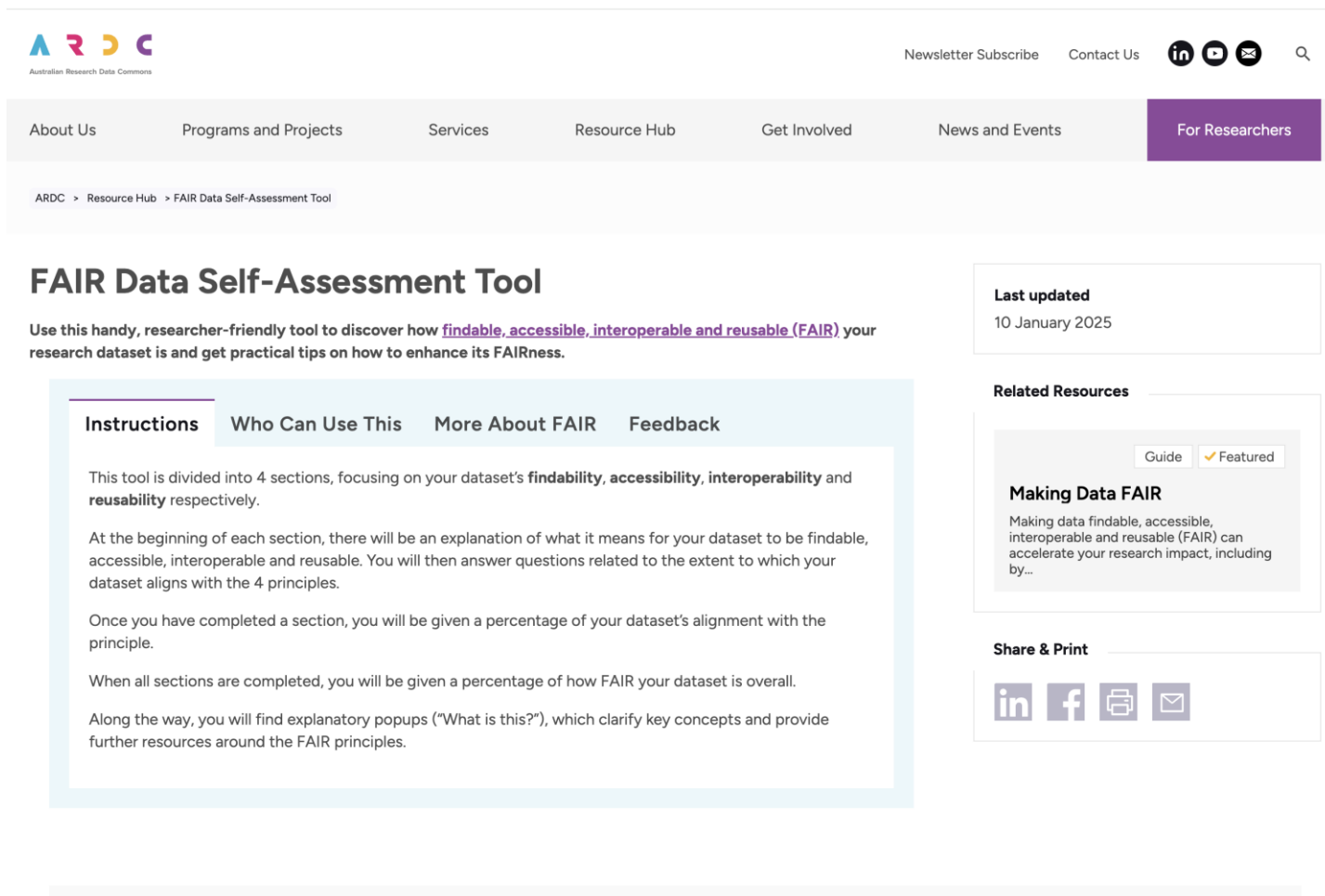



©2018 by DeiC/Deff/DTU/CBS/AAU/KU/Rigsarkivet/ Det Kgl Bibliotek
Policies is made available under a Creative Commons 4.0 licence (international)







home.cern

FAIR data self assessment tool



 Australian Research Data Commons

Newsletter Subscribe Contact Us    

About Us Programs and Projects Services Resource Hub Get Involved News and Events **For Researchers**

ARDC > Resource Hub > FAIR Data Self-Assessment Tool

FAIR Data Self-Assessment Tool

Use this handy, researcher-friendly tool to discover how [findable, accessible, interoperable and reusable \(FAIR\)](#) your research dataset is and get practical tips on how to enhance its FAIRness.

Instructions Who Can Use This More About FAIR Feedback

This tool is divided into 4 sections, focusing on your dataset's **findability, accessibility, interoperability and reusability** respectively.

At the beginning of each section, there will be an explanation of what it means for your dataset to be findable, accessible, interoperable and reusable. You will then answer questions related to the extent to which your dataset aligns with the 4 principles.

Once you have completed a section, you will be given a percentage of your dataset's alignment with the principle.

When all sections are completed, you will be given a percentage of how FAIR your dataset is overall.

Along the way, you will find explanatory popups ("What is this?"), which clarify key concepts and provide further resources around the FAIR principles.

Last updated
10 January 2025

Related Resources

Guide Featured

Making Data FAIR
Making data findable, accessible, interoperable and reusable (FAIR) can accelerate your research impact, including by...

Share & Print

