Marjan Grootveld
Data Archiving and Networked Services (DANS)

Horizon 2020 Open Research Data Pilot and tools from OpenAIRE and EUDAT

OpenAIRE: un'infrastruttura per la Scienza Aperta in Europa
Rome, May 30 2016

European Commission

Agenda

- What is DANS?
- The Open Research Data Pilot
- OPENAIRE support kit for the Data Pilot
- Planning data management
- EUDAT data services
- The data management plan – Part 1

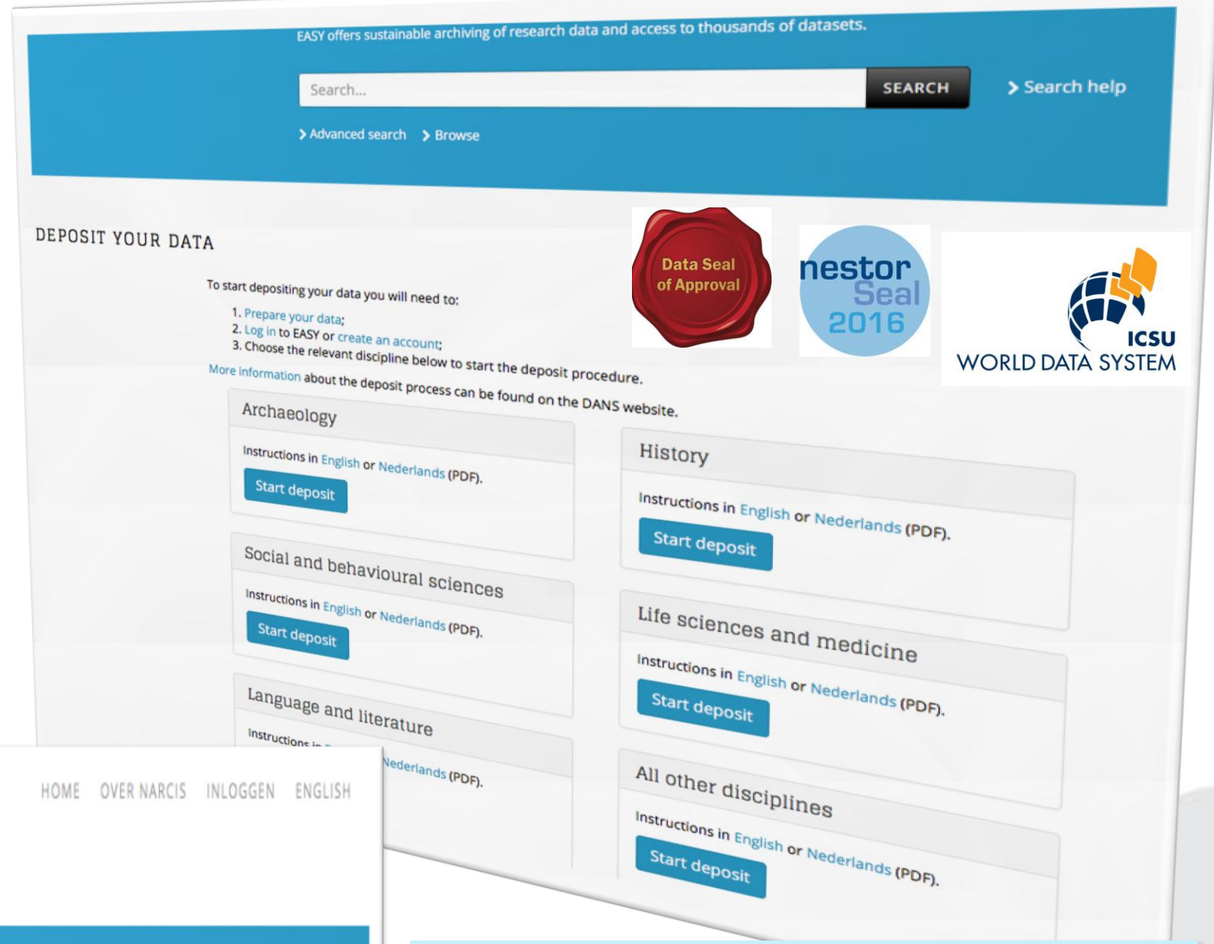
About DANS

Institute of Dutch Academy and Research Funding Organisation (KNAW & NWO) since 2005

Mission: promote and provide permanent access to digital research information

First predecessor dates back to 1964 (Steinmetz Foundation), Historical Data Archive 1989

DataverseNL for short- and mid-term storage



EASY: certified long-term Electronic Archiving System for self-deposit

NARCIS: Gateway to scholarly information in the Netherlands

Data Archiving and Networked Services

DANS

NARCIS

HOME ABOUT NARCIS LOGIN NEDERLANDS

The gateway to scholarly information in the Netherlands

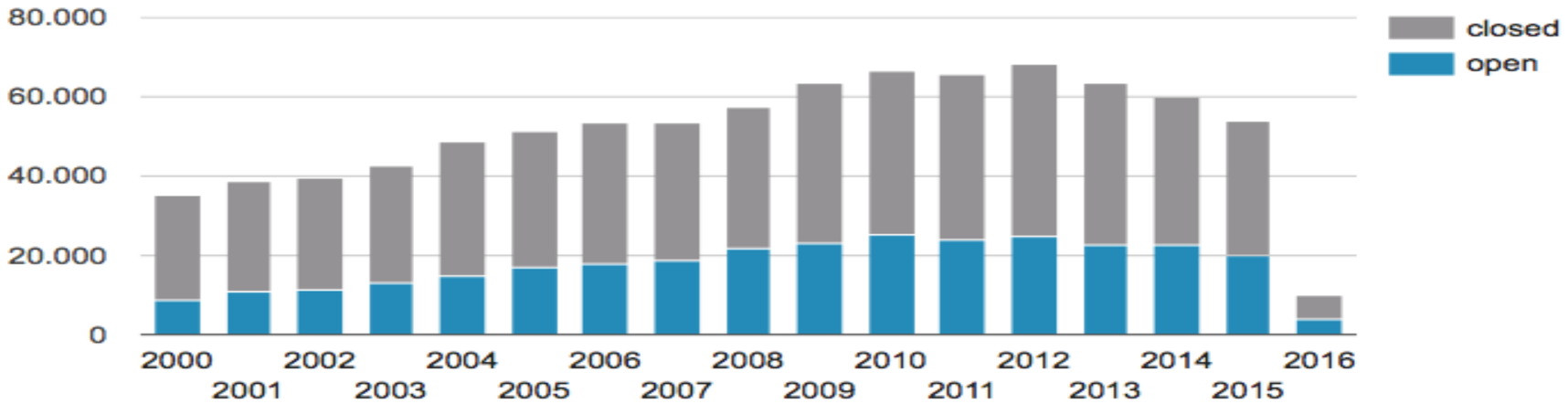
> Submit Content to NARCIS

455.000 Open Access

Search... SEARCH

1,249,951	150,965	63,766	52,859	2,952	1,710 beta release
PUBLICATIONS	DATA SETS	RESEARCH	PEOPLE	ORGANISATIONS	ENHANCED PUBLICATIONS

OPEN AND CLOSED ACCESS SCHOLARLY PUBLICATIONS IN NARCIS PER YEAR OF PUBLICATION



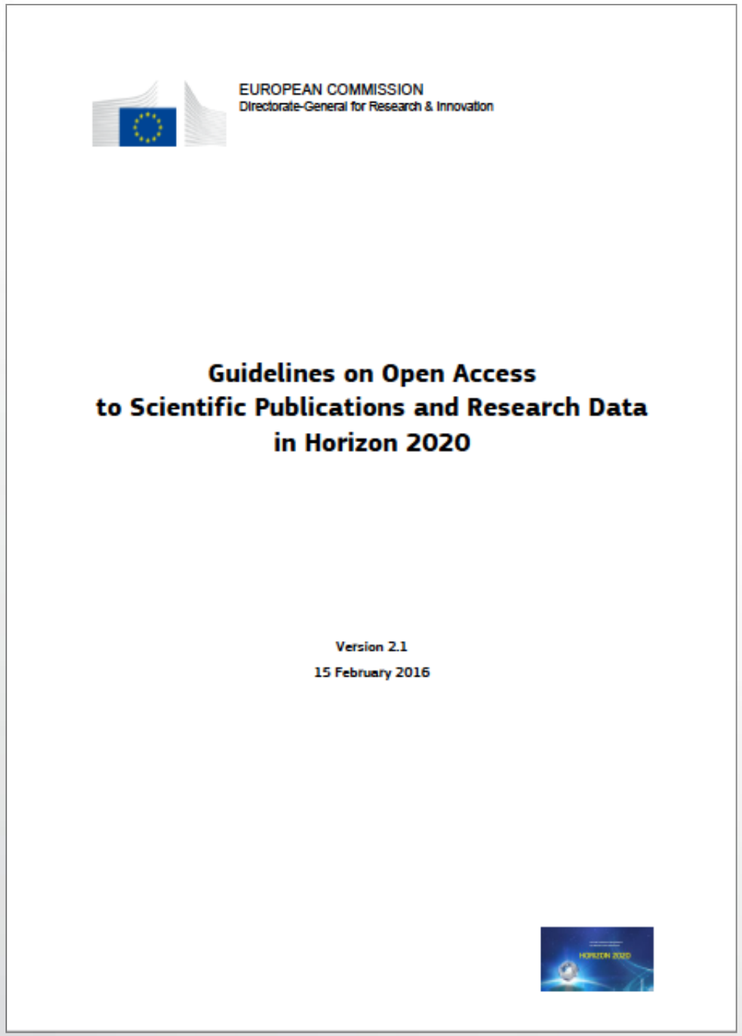
This chart shows the actual number of open and closed access publications (articles, doctoral theses, books, reports et cetera) in NARCIS, since 2000.

> More statistics

http://www.narcis.nl/

OPEN RESEARCH DATA PILOT

Official Guidelines 2016 - 1

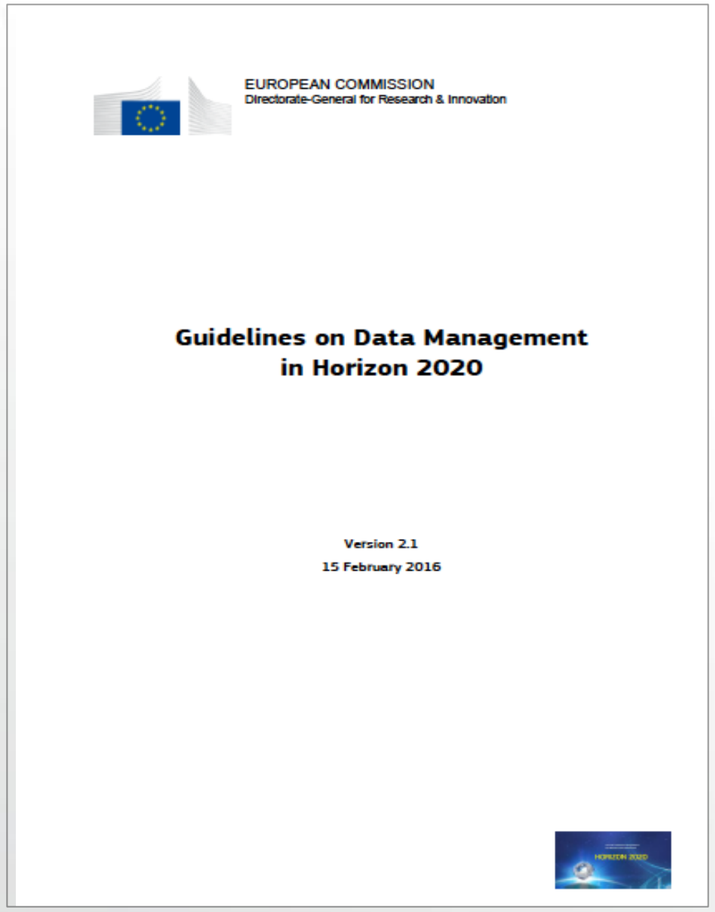


Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

- **These 2016 Guidelines state the official requirements of the Open Research Data Pilot.**
- **In these Guidelines also links to the AGA - Annotated Model Grant Agreement**

Official Guidelines 2016 - 2


EUROPEAN COMMISSION
Directorate-General for Research & Innovation

Guidelines on Data Management in Horizon 2020

Version 2.1
15 February 2016

Guidelines on Data Management in Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

This document is intended to help applicants and beneficiaries of projects under Horizon 2020 meet their responsibilities as regards research data quality, sharing and security.

Requirements Open Data Pilot

1. **Data Management Plan required within six months after project grant (regular project deliverable)**
2. **Deposit the research data in a research data repository**
3. **Make the research data openly accessible**

Participation in the pilot is not part of the project evaluation.

To whom does the Data Pilot concern?

Current situation:

- **Researchers funded by Horizon 2020 within 9 specified call areas.**
- ***Opt out* and *opt in* are possible and are being used (35%, resp. 13% - Nov '15).**
- **Call areas: https://www.openaire.eu/opendatapilot**

As of 2017:

- ***European Cloud Initiative* to give Europe a global lead in the data-driven economy.**
- **Open data will become the default option. The pilot will be extended to cover all call areas. Opting out remains possible.**
- **Press release: http://europa.eu/rapid/press-release_IP-16-1408_en.htm**

Daniel Spichtinger (EC) at OpenCon 14-11-15: 3,699 Horizon 2020 signed grant agreements – 149/431 projects in core areas opted out - 409/3268 projects in other areas opted in

Reasons for opting out:

- **participation is incompatible with the Horizon 2020 obligation to protect results that can reasonably be expected to be commercially or industrially exploited;**
- **participation is incompatible with the need for confidentiality in connection with security issues;**
- **participation is incompatible with rules on protecting personal data;**
- **the project will not generate / collect any research data; or**
- **there are other legitimate reasons not to take part in the Pilot.**

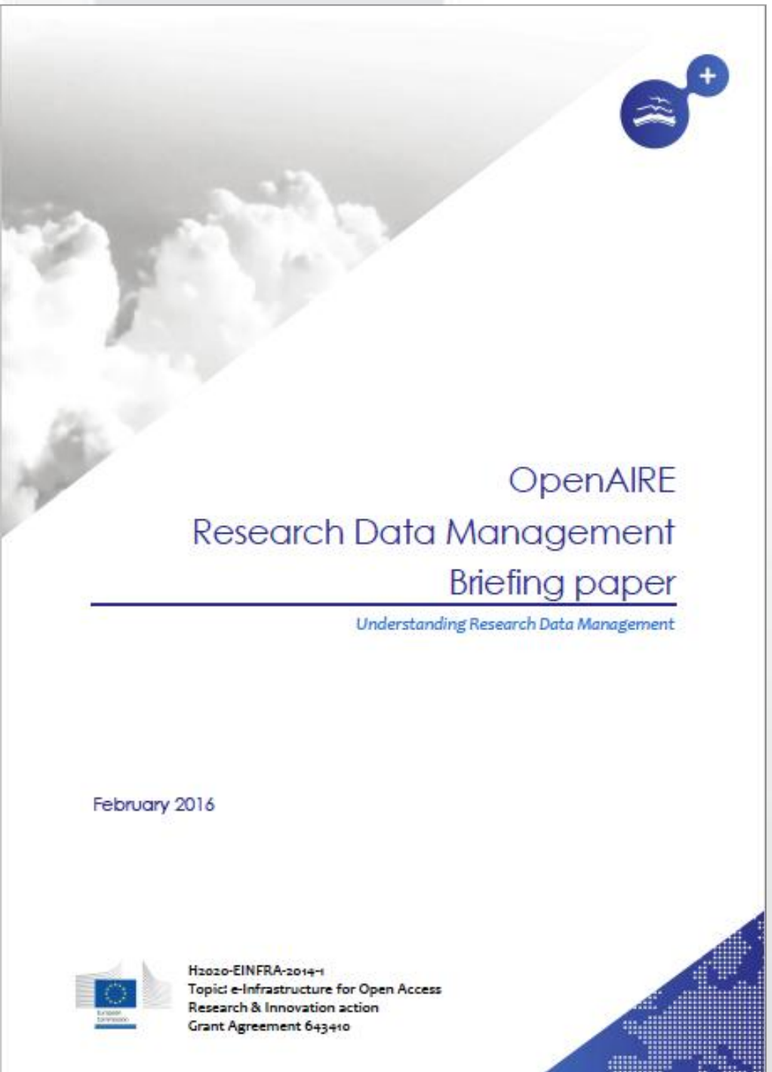
Note that** partial **opt out is possible** – **and preferable to full opt

OPENAIRE SUPPORT KIT FOR THE DATA PILOT

For NOADs:

https://www.openaire.eu/openaire-noad-info-dissemination-kit

Briefing Paper RDM



OpenAIRE Research Data Management Briefing Paper

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

- **This extensive briefing paper gives an overview of Research Data Management with practical sections about data management planning, and archiving the research data for reuse.**

Information at the OpenAIRE website

- **Open Research Data Pilot**

https://www.openaire.eu/opendatapilot

- What is the pilot? Which H2020 strands are required to participate? What practical steps should the researcher take?

- **Create a Data Management Plan**

https://www.openaire.eu/opendatapilot-dmp

- Information about how to create a Data Management Plan. First steps; When to write and revise your Data Management Plan

- **Select a Data Repository**

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

- Information about how to select a repository

- **Frequently Asked Questions about the Open Research Data Pilot**

https://www.openaire.eu/support/faq

Webinars

Webinar page https://www.openaire.eu/webinars/

- **E.g. about Pilot, Zenodo and data management, but also about OA publications and how to make your repository OpenAIRE compatible.**
- **Various presenters; slides included.**
- **Upcoming joint OpenAIRE-EUDAT webinar on July 7 ‘How to write a DMP?’**

OpenAIRE factsheets

OpenAIRE Horizon2020 FactSheets

Visit the OpenAIRE Portal www.openaire.eu

Open Research Data Pilot in Horizon 2020
How can OpenAIRE help?

Briefing paper for Researchers, Project coordinators, and Research Managers participating in the EC Open Research Data Pilot in Horizon 2020.

What is the Open Research Data Pilot?

Open data is data that is free to use, reuse, and redistribute. The Open Research Data Pilot aims to make the research data generated by selected Horizon 2020 projects open. It will be carefully monitored and used to inform future EC policy.

If your Horizon 2020 project is part of the pilot, you must:

- Develop (and keep up-to-date) a Data Management Plan (DMP).
- Deposit your data in a research data repository.
- Make sure third parties can freely access, mine, exploit, reproduce and disseminate it.
- Make clear what tools will be needed to use the raw data to validate research results (or provide the tools themselves).

The pilot applies to (1) the data (and metadata) needed to validate results in scientific publications, and (2) other curated and/or raw data (and metadata) that you specify in the DMP.

What's in it for you?

- Be part of the new era of Open Science, integrating transparency, effectiveness and timeliness into all areas of scientific methods and processes.
- Reach more people, have greater impact.
- Avoid duplication of effort and help preserve data for future researchers.
- Simplify final reporting thanks to an up-to-date DMP.

OpenAIRE

Do you have a Horizon 2020 project grant?

Check your grant agreement to see if you are part of the Pilot.

Are you part of the Pilot?

If your project stems from one of these Horizon 2020 areas, you are automatically part of the pilot:

- Future and Emerging Technologies
- Research infrastructures – part e-Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – with the exception of raw materials topics
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society

Can you opt in? Yes, projects in other areas can participate on a voluntary basis.

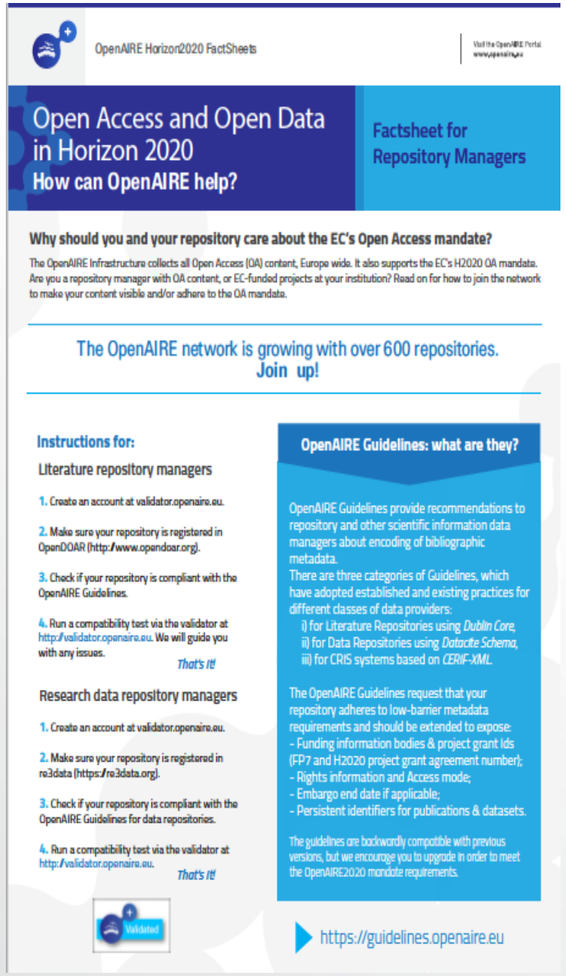
Can you opt out? We hope you don't, but projects may at any stage opt out of the Pilot. See the EC Guide on OA (link overleaf) for eligible reasons.

How can OpenAIRE help?

OpenAIRE provides a range of resources, FAQs, webinars and support pages. OpenAIRE has local representatives in all EU countries: the National Open Access Desks or NOADs. Contact them via our helpdesk system at www.openaire.eu

If you have no other suitable place to deposit your data, you can use OpenAIRE's catch-all repository, Zenodo

zenodo



OpenAIRE Horizon2020 FactSheets

Visit the OpenAIRE Portal www.openaire.eu

Open Access and Open Data in Horizon 2020
How can OpenAIRE help?

Factsheet for Repository Managers

Why should you and your repository care about the EC's Open Access mandate?

The OpenAIRE Infrastructure collects all Open Access (OA) content, Europe wide. It also supports the EC's H2020 OA mandate. Are you a repository manager with OA content, or EC-funded projects at your institution? Read on for how to join the network to make your content visible and/or adhere to the OA mandate.

The OpenAIRE network is growing with over 600 repositories. Join up!

Instructions for:

Literature repository managers

1. Create an account at validator.openaire.eu.
2. Make sure your repository is registered in OpenDOAR (http://www.opendoar.org).
3. Check if your repository is compliant with the OpenAIRE Guidelines.
4. Run a compatibility test via the validator at http://validator.openaire.eu. We will guide you with any issues.

That's it!

Research data repository managers

1. Create an account at validator.openaire.eu.
2. Make sure your repository is registered in re3data (https://re3data.org).
3. Check if your repository is compliant with the OpenAIRE Guidelines for data repositories.
4. Run a compatibility test via the validator at http://validator.openaire.eu.

That's it!

Validated

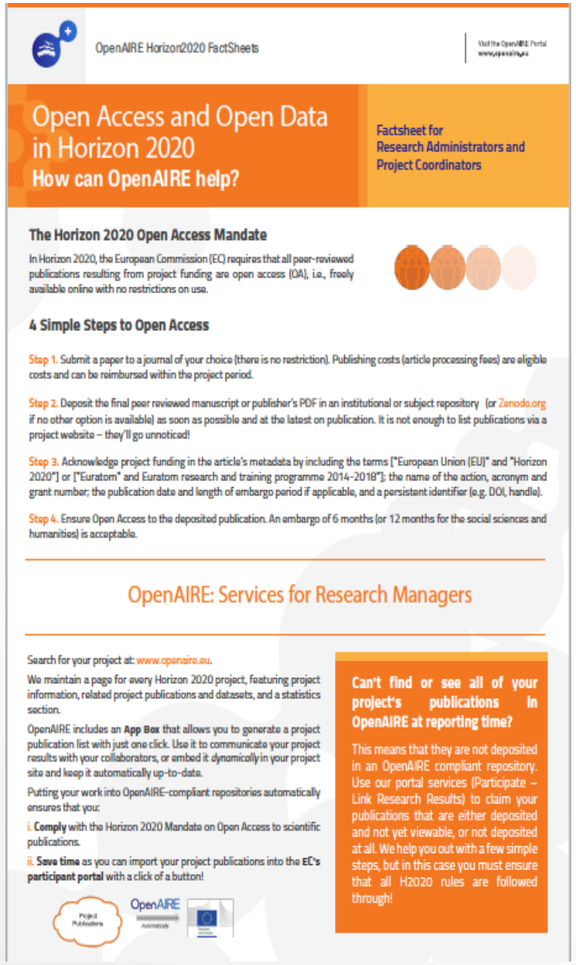
OpenAIRE Guidelines: what are they?

OpenAIRE Guidelines provide recommendations to repository and other scientific information data managers about encoding of bibliographic metadata.
There are three categories of Guidelines, which have adopted established and existing practices for different classes of data providers:
i) for Literature Repositories using *Dublin Core*,
ii) for Data Repositories using *Datacite Schema*,
iii) for CRIS systems based on *CERIF-XML*.

The OpenAIRE Guidelines request that your repository adheres to low-barrier metadata requirements and should be extended to expose:
- Funding information bodies & project grant Ids (FP7 and H2020 project grant agreement number);
- Rights information and Access mode;
- Embargo end date if applicable;
- Persistent identifiers for publications & datasets.

The guidelines are backwardly compatible with previous versions, but we encourage you to upgrade in order to meet the OpenAIRE2020 mandate requirements.

https://guidelines.openaire.eu



OpenAIRE Horizon2020 FactSheets

Visit the OpenAIRE Portal www.openaire.eu

Open Access and Open Data in Horizon 2020
How can OpenAIRE help?

Factsheet for Research Administrators and Project Coordinators

The Horizon 2020 Open Access Mandate

In Horizon 2020, the European Commission (EC) requires that all peer-reviewed publications resulting from project funding are open access (OA), i.e., freely available online with no restrictions on use.

4 Simple Steps to Open Access

Step 1. Submit a paper to a journal of your choice (there is no restriction). Publishing costs (article processing fees) are eligible costs and can be reimbursed within the project period.

Step 2. Deposit the final peer reviewed manuscript or publisher's PDF in an institutional or subject repository (or Zenodo.org if no other option is available) as soon as possible and at the latest on publication. It is not enough to list publications via a project website – they'll go unnoticed!

Step 3. Acknowledge project funding in the article's metadata by including the terms ["European Union (EU)" and "Horizon 2020"] or ["Euratom" and Euratom research and training programme 2014-2018"]; the name of the action, acronym and grant number; the publication date and length of embargo period if applicable, and a persistent identifier (e.g. DOI, handle).

Step 4. Ensure Open Access to the deposited publication. An embargo of 6 months (or 12 months for the social sciences and humanities) is acceptable.

OpenAIRE: Services for Research Managers

Search for your project at: www.openaire.eu.

We maintain a page for every Horizon 2020 project, featuring project information, related project publications and datasets, and a statistics section.

OpenAIRE includes an **App Box** that allows you to generate a project publication list with just one click. Use it to communicate your project results with your collaborators, or embed it *dynamically* in your project site and keep it automatically up-to-date.

Putting your work into OpenAIRE-compliant repositories automatically ensures that you:

i. **Comply** with the Horizon 2020 Mandate on Open Access to scientific publications.

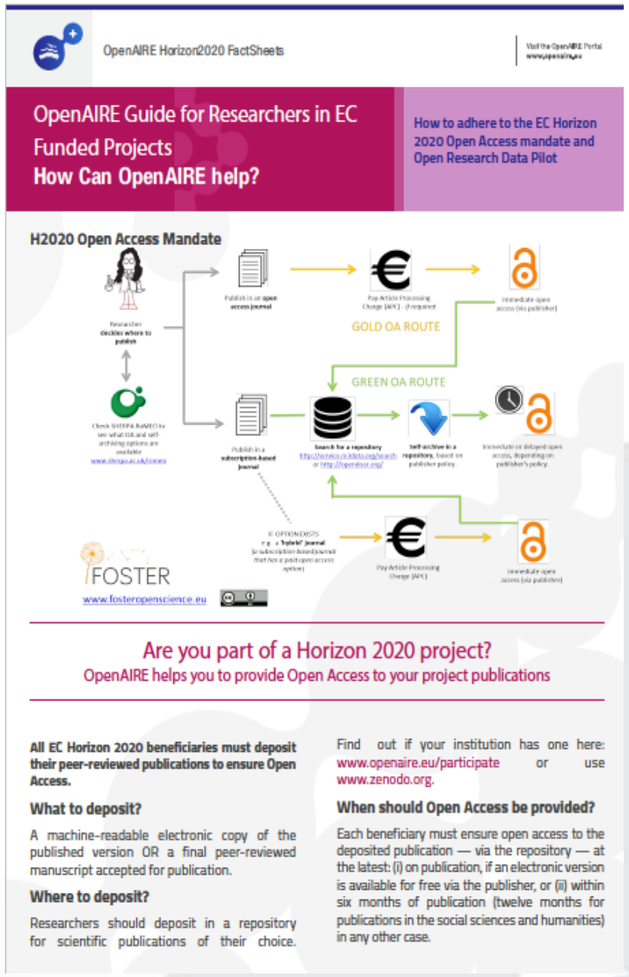
ii. **Save time** as you can import your project publications into the **EC's participant portal** with a click of a button!

Project Publications Automatically

OpenAIRE

Can't find or see all of your project's publications in OpenAIRE at reporting time?

This means that they are not deposited in an OpenAIRE compliant repository. Use our portal services (Participate – Link Research Results) to claim your publications that are either deposited and not yet viewable, or not deposited at all. We help you out with a few simple steps, but in this case you must ensure that all H2020 rules are followed through!



OpenAIRE Horizon2020 FactSheets

Visit the OpenAIRE Portal www.openaire.eu

OpenAIRE Guide for Researchers in EC Funded Projects
How Can OpenAIRE help?

How to adhere to the EC Horizon 2020 Open Access mandate and Open Research Data Pilot

H2020 Open Access Mandate

FOSTER

www.fosteropenscience.eu

Are you part of a Horizon 2020 project?

OpenAIRE helps you to provide Open Access to your project publications

All EC Horizon 2020 beneficiaries must deposit their peer-reviewed publications to ensure Open Access.

What to deposit?

A machine-readable electronic copy of the published version OR a final peer-reviewed manuscript accepted for publication.

Where to deposit?

Researchers should deposit in a repository for scientific publications of their choice.

Find out if your institution has one here: www.openaire.eu/participate or use www.zenodo.org.

When should Open Access be provided?

Each beneficiary must ensure open access to the deposited publication — via the repository — at the latest: (i) on publication, if an electronic version is available for free via the publisher, or (ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.

European Commission

Postcard

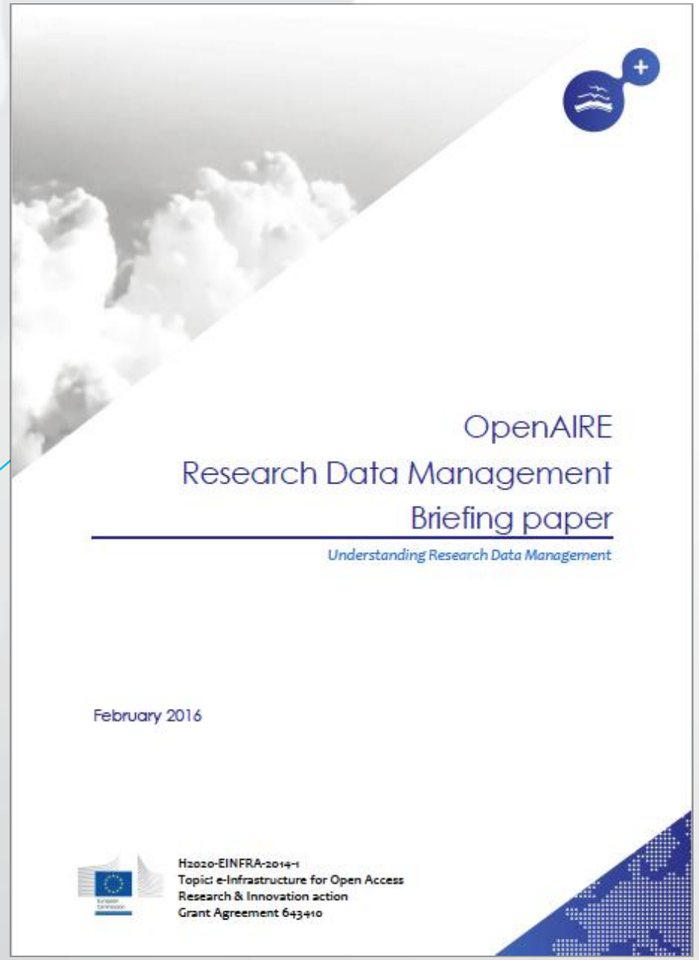
Postcard: Are you part of EC's Open Research Data Pilot?

- **https://box.openaire.eu/index.php/apps/files/?dir=/OpenAIRE2020/Dissemination_Materials/Postcards#pdfviewer**

- **This postcard can be used as dissemination material.**

PLANNING
DATA MANAGEMENT

Open access to research data



Researchers should specify which data will be shared openly in the Data Management Plan. What does "open" mean here? Again, there are several definitions, but the pilot Guidelines on Data Management state **"Open access to research data** refers to the right to access and re-use digital research data. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user."

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Which data does the pilot apply to?

- **Data, including associated metadata, needed to validate the results in scientific publications.**
- **Other curated and/or raw data, including associated metadata, as specified in the Data Management Plan**

Research data means data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or another research output is based. Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and may be held in any format or media.

What should be deposited?

- The **data** needed to validate results in scientific publications (minimally!).
- The associated **metadata**: the dataset's creator, title, year of publication, repository, identifier etc.
  - Follow a metadata standard in your line of work, or a generic standard, e.g. Dublin Core or DataCite., and be FAIR.
  - The repository will assign a persistent ID to the dataset: important for discovering and citing the data.
- **Documentation** like code books, lab journals, informed consent forms – domain-dependent, and important for understanding the data and combining them with other data sources.
- **Software**, hardware, tools, syntax queries, machine configurations – domain-dependent, and important for using the data. (Alternative: information about the software etc.)

Basically, everything that is needed to replicate a study should be available for others.

Research Data Alliance (RDA) http://rd-alliance.github.io/metadata-directory/standards/

FAIR Guiding Principles for scientific data management http://www.nature.com/articles/sdata201618

Archive the data openly, unless…

Grant Agreement, Art. 29.3, *Open Access to research data*:

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

- **Confidentiality and security issues can be good reasons *not* to publish or share – all – data. Note in the DMP the reasons for not giving access, and deposit that part of the data under a Restricted Access regime.**
- **When regenerating data would be cheaper than archiving, don't archive. Spend time on selecting what data you'll need and want to retain. Motivate your criteria in the DMP.**

See http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
For selection criteria see https://www.openaire.eu/opendatapilot

Where to find a repository?

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. More information for **selecting a data repository.**
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.

   zenodo
   Research. Shared.
4. Search for other research data repositories in http://re3data.org/

   re3data.org
   REGISTRY OF RESEARCH DATA REPOSITORIES

More information: https://www.openaire.eu/opendatapilot-repository
Zenodo: http://www.zenodo.org/
Re3data.org: http://www.re3data.org/

How to select a repository? 1/2

Main criteria for choosing a data repository:

- **Certification as a ‘Trustworthy Digital Repository’, with an explicit ambition to keep the data available in the long term.**
  - **Network of trustworthy digital repositories for long-term preservation of the data after the research is finished.**
  - **Three common certification standards for TDRs:**

Data Seal of Approval: http://datasealofapproval.org/en/
nestor seal for DIN 31644: http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html
ISO 16363: http://www.iso16363.org/

How to select a repository? 2/2

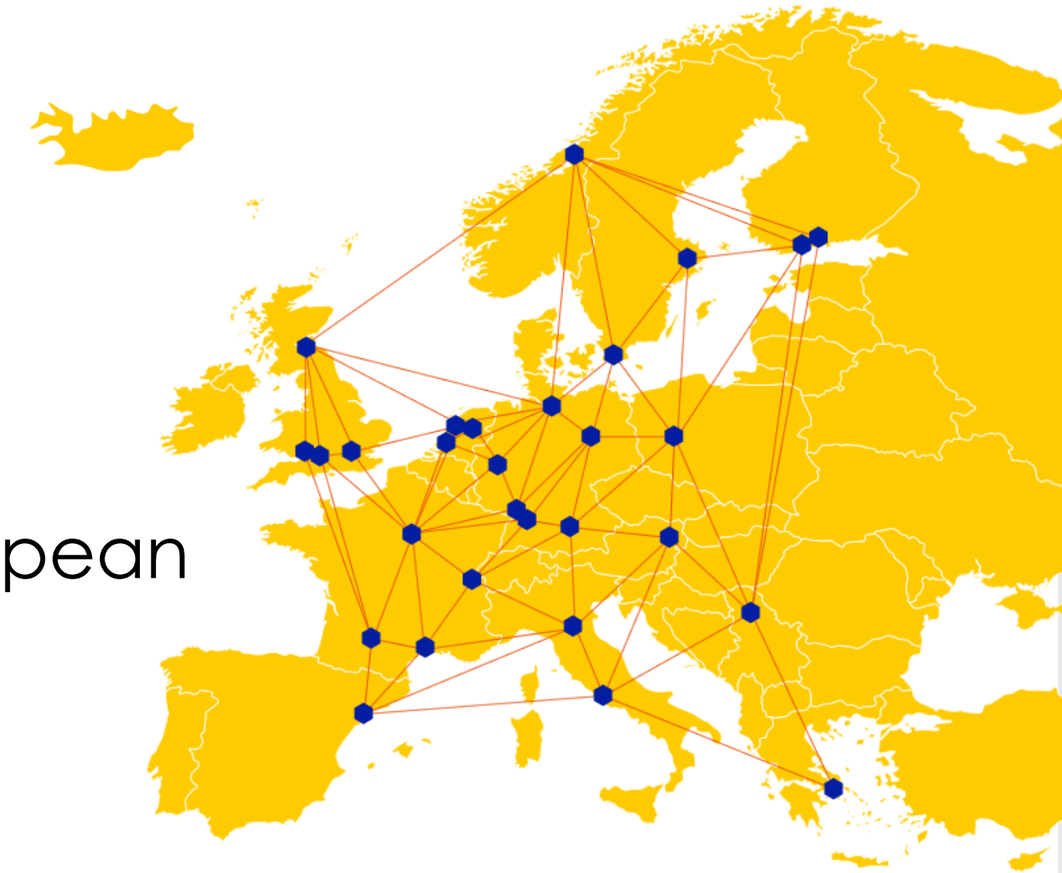
Main criteria for choosing a data repository:

- **Certification as a ‘Trustworthy Digital Repository’, with an explicit ambition to keep the data available in the long term.**
- **Matches your particular data needs: e.g. certain file formats; mixture of Open and Restricted Access. So contact the repository of your choice when writing the first version of your DMP, or earlier.**
- **Provides guidance on metadata and on how to cite the data that has been deposited.**
- **Gives your submitted dataset a persistent and globally unique identifier: for sustainable citations – both for data and publications – and to link back to particular researchers and grants.**

EUDAT DATA SERVICES

EUDAT project

EUDAT offers common data services to both research communities and individuals through a network of 35 European organisations.

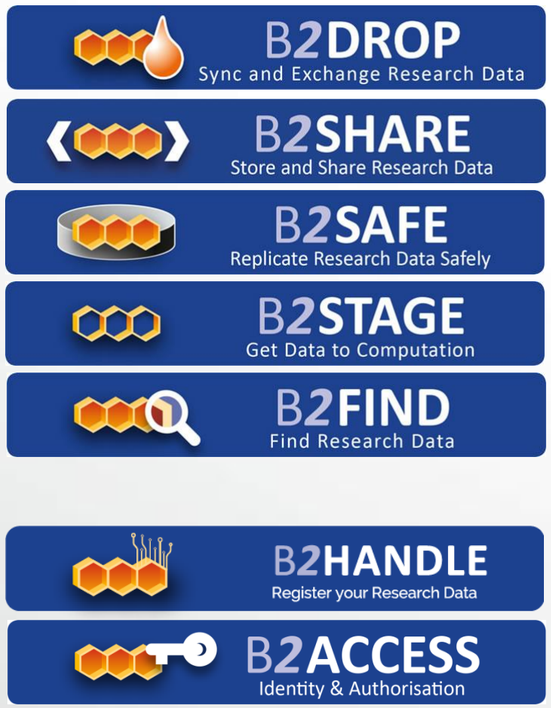


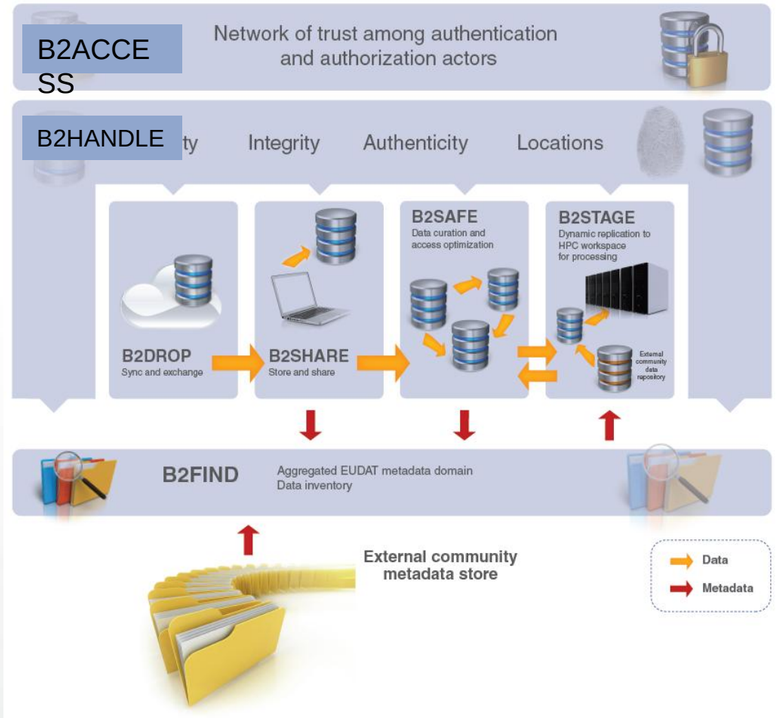
EUDAT offers data services

EUDAT services are designed, built and implemented based on user community requirements.

EUDAT services

B2 SERVICE SUITE



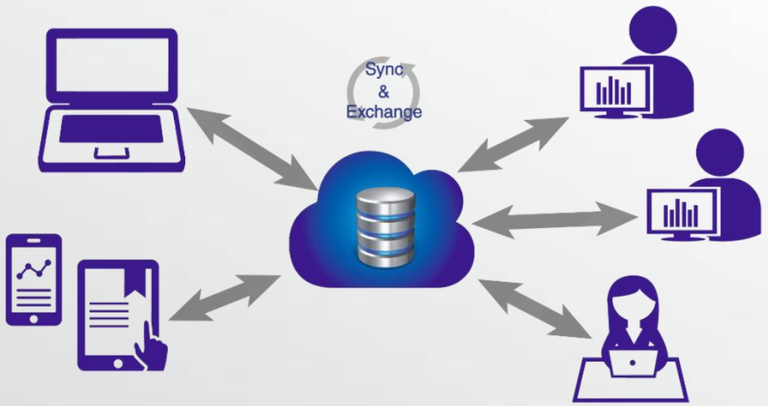


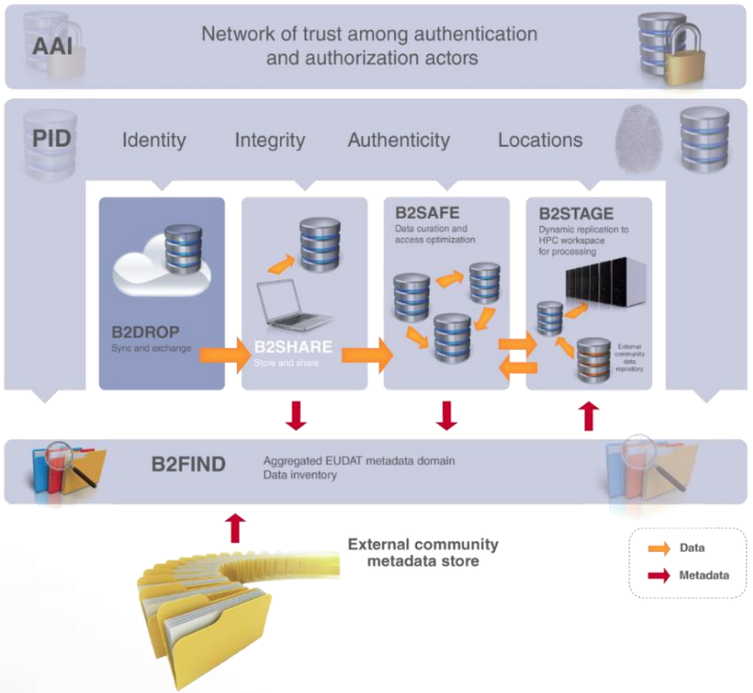
https://eudat.eu/services

e.g. B2DROP – a solution for researchers and scientists to:

b2drop.eudat.eu

- Store and exchange data with colleagues and team members, including research data not finalized for publishing
- share data with fine-grained access controls
- synchronize multiple versions of data across different devices





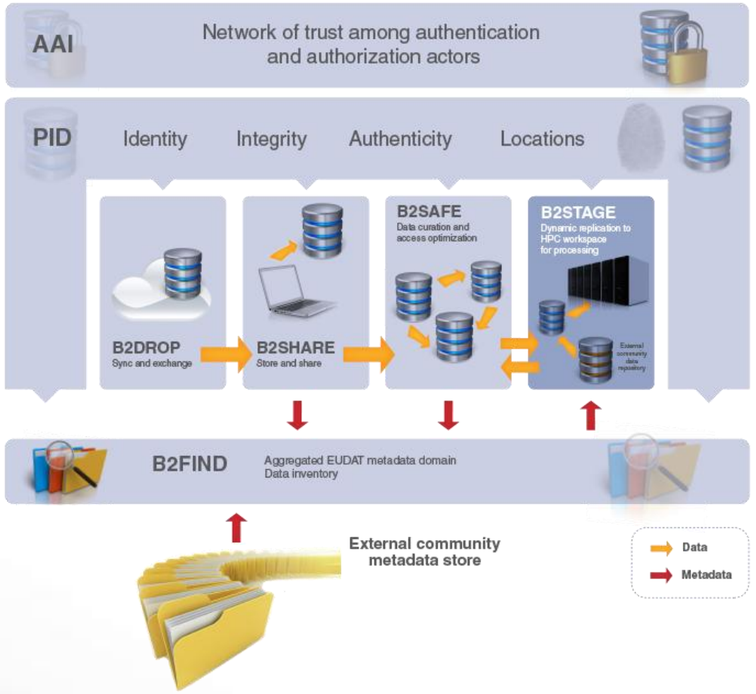
Features:

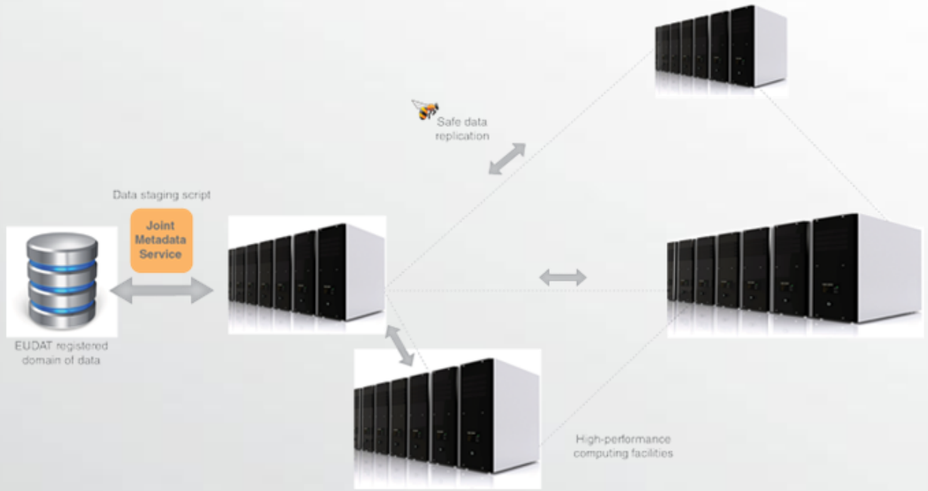
- 20GB storage per user
- Living objects, so no PIDs
- Versioning and offline use
- Desktop synchronisation
- B2DROP is hosted at the Jülich Supercomputing Centre
- Daily backups of all files in B2DROP are taken and kept on tape

e.g. B2STAGE - Facilitating communities to:

eudat.eu/b2stage

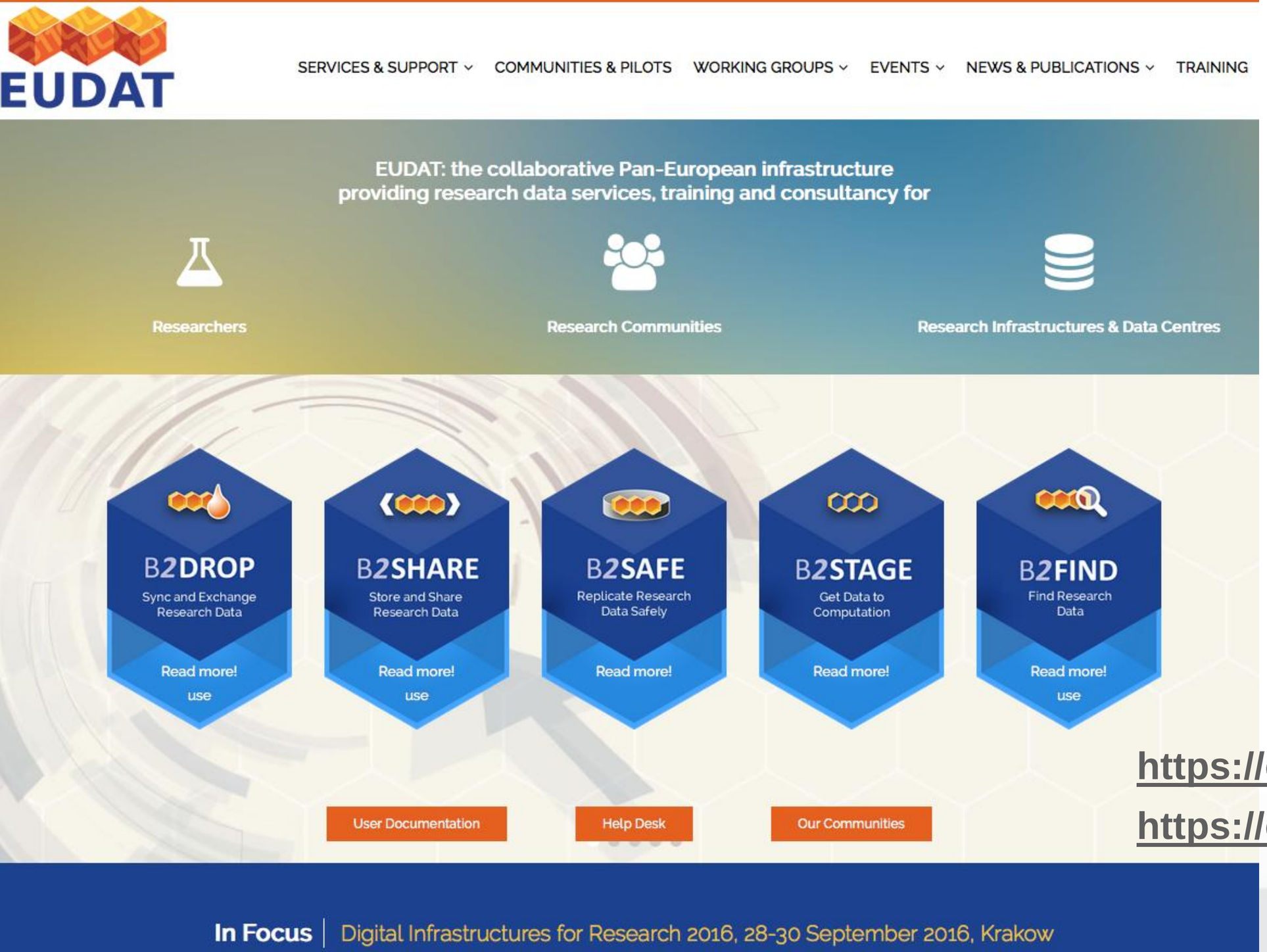
- move large amounts of data between data stores and high-performance compute resources
- re-ingest computational results back into EUDAT
- deposit large data sets into EUDAT resources for long-term preservation





Features:

- high-speed transfer
- reliable and light-weight
- manages permanent PIDs



https://eudat.eu/training

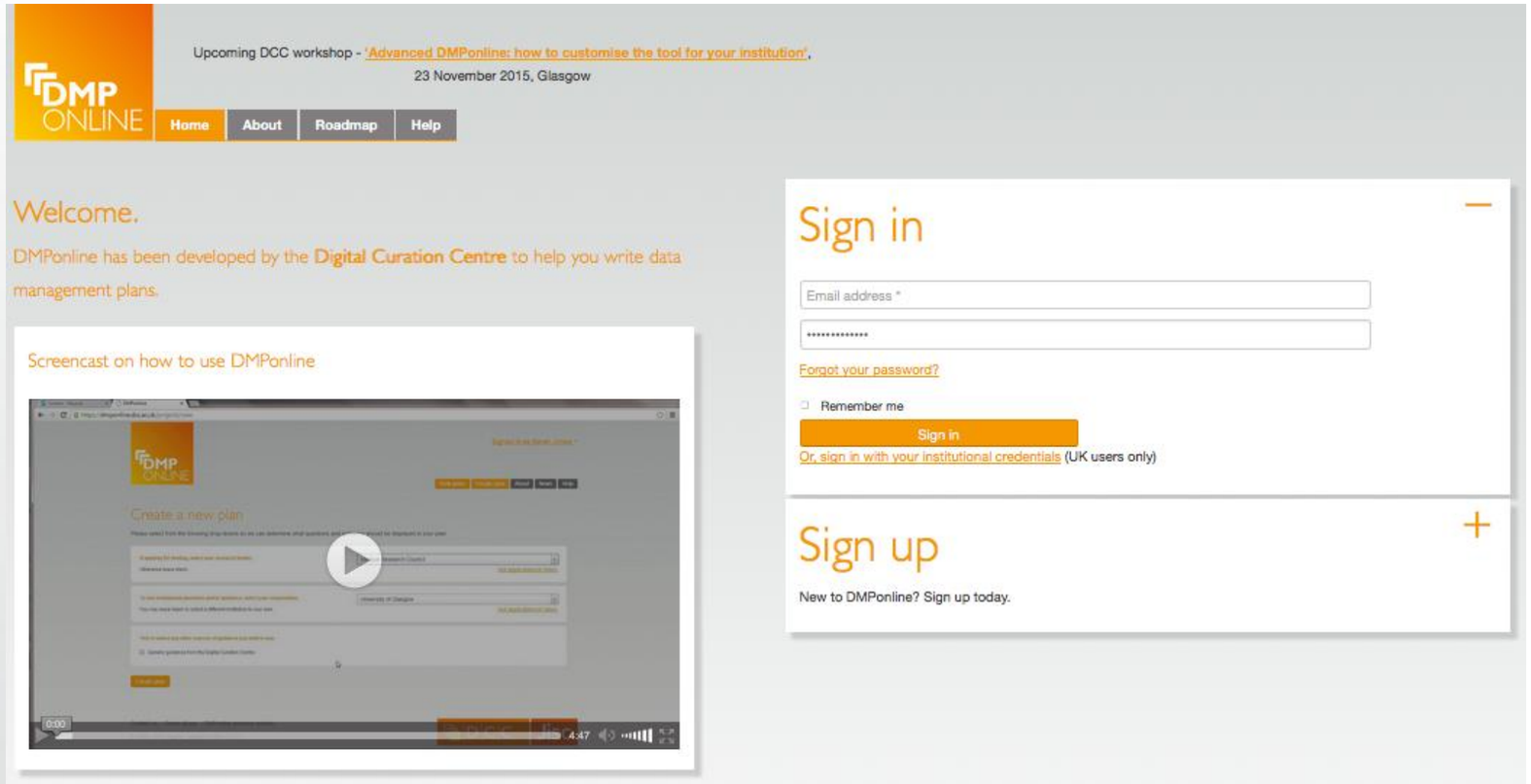
https://eudat.eu/events/webina

DATA MANAGEMENT PLAN (PART 1)

How to write a DMP

1

- **Template available from https://dmponline.dcc.ac.uk/**



- **And from a few national DMPonline sites, e.g. in Spain and Belgium**

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 commo

If applying for funding, select your research funder.

Otherwise leave blank.

2

European Commissio

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

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

3

ELIXIR

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

- [x] DCC guidance

4

Create plan

Initial DMP

An initial DMP should be completed within 6 months of starting the project.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets project.

The DMP is not a fixed document, but evolves during the lifespan of the project.

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

Sections	Questions
For each data set specify the following:	- Data set reference and name - Data set description - Standards and metadata - Data sharing - Archiving and preservation (including storage and backup)

Mid-term Review DMP

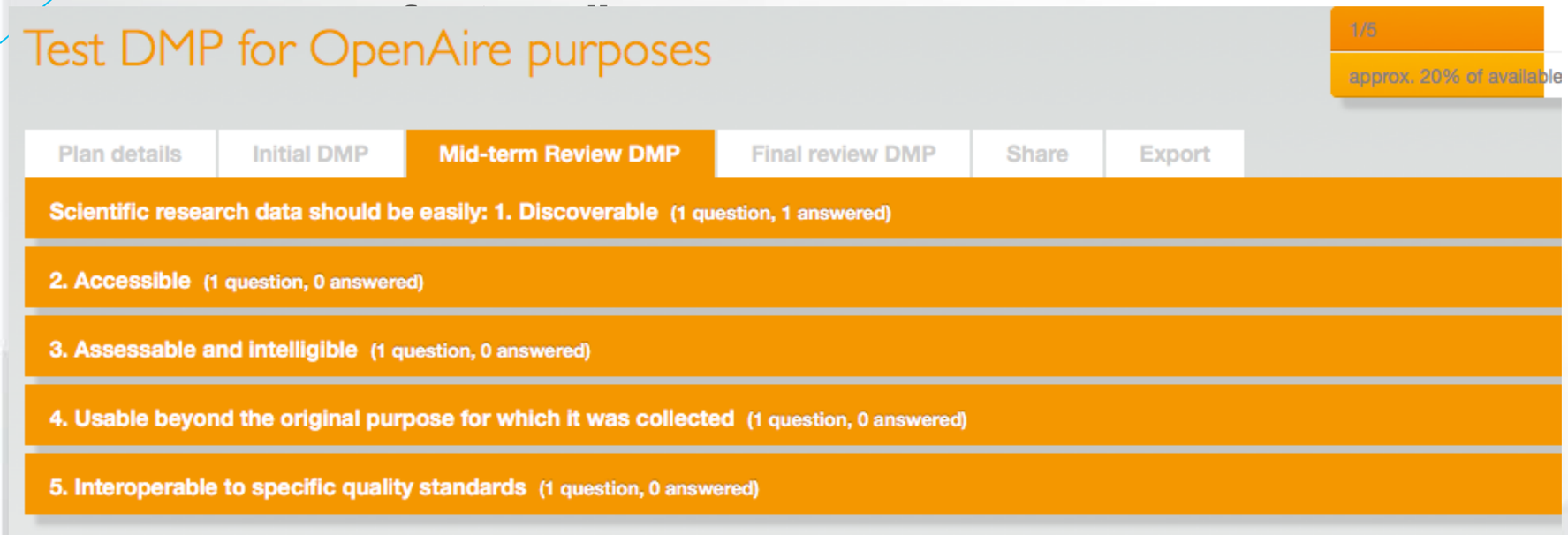
Final review DMP

Briefly specify

- how data will be captured/created
- how it will be documented
- according to what standards
- who will be able to access it
- where it will be stored
- how it will be backed up, and
- where and how it will be shared and preserved long-term

Deliver the DMP

EC: “Since DMPs are expected to mature during the project, more developed versions of the plan can be included as additional deliverables at later stages. (…) New versions of the DMP should be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external



Software Sustainability Institute

About | Blog | Community | Policy | Software | Training | R

Writing and using a software management plan

A software management plan should minimally include:

- information on what software outputs (including documentation and other related material) are expected to be produced [Note: it's not as important to define the functionality in detail, as this may change in a research software project];
- who is responsible for releasing the software [Note: this is often the PI, project manager or lead developer but may a specific role];
- the revision control process to be used [Note: it is important to choose a revision control / configuration management system that all members of the team will use];
- what license will be used for each output [Note: this might require input from your funder or institution].

A software management plan could also:

- identify the software development model to be used;
- identify the external software that will be brought into the project, and the associated licences;
- what method will be used to accept each output (e.g. a specific user will review it);
- dependencies between outputs and with external dependencies;
- major risks that might impact on the delivery of the outputs.

See http://www.software.ac.uk/resources/guides/software-management-plans

A DMP is about ‘keeping’ data

- Storing data < > archiving data
- Archived data < > findable data
- Findable < > accessible
- Accessible < > understandable
- Understandable < > usable

- a USB stick is not safe
- Figshare is not a Trustworthy Digital Repository
- a persistent identifier is essential but no guarantee for usability
- Data in a proprietary format are not sustainable

… and about sharing data

Questions?

www.openaire.eu

@openaire_eu

facebook.com/groups/openaire

linkedin.com/groups/OpenAIRE-3893548

marjan.grootveld@dans.knaw.nl