

# EXPANDS

**European Open Science Cloud Photon  
and Neutron Data Services**

## **Publishing the (PaN) Experiment**

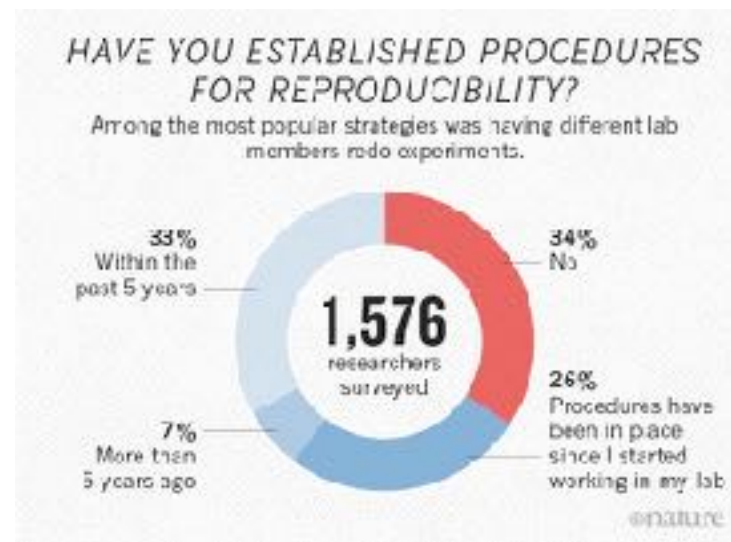
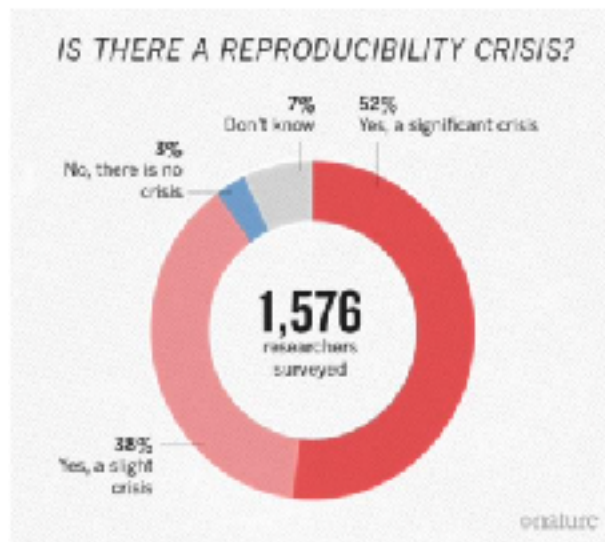
**Symposium for Librarians and Data Managers**

Oliver Knodel (HZDR)



# Why Publishing the Experiment?

- Publishing the experiment ideally leads to **comprehensible science** through documentation,
- Allowing other scientists to contribute far beyond the original findings,
- Increases visibility and recognition through **reproducibility**,
- Supports other scientists in improving research and methods,
- And is wished (and sometimes required) by funders.



Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). <https://doi.org/10.1038/533452a>



# Open Science and FAIR Research

## The F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub> Principles

Our research should follow the FAIR principles definition which is referenced from:

### Note

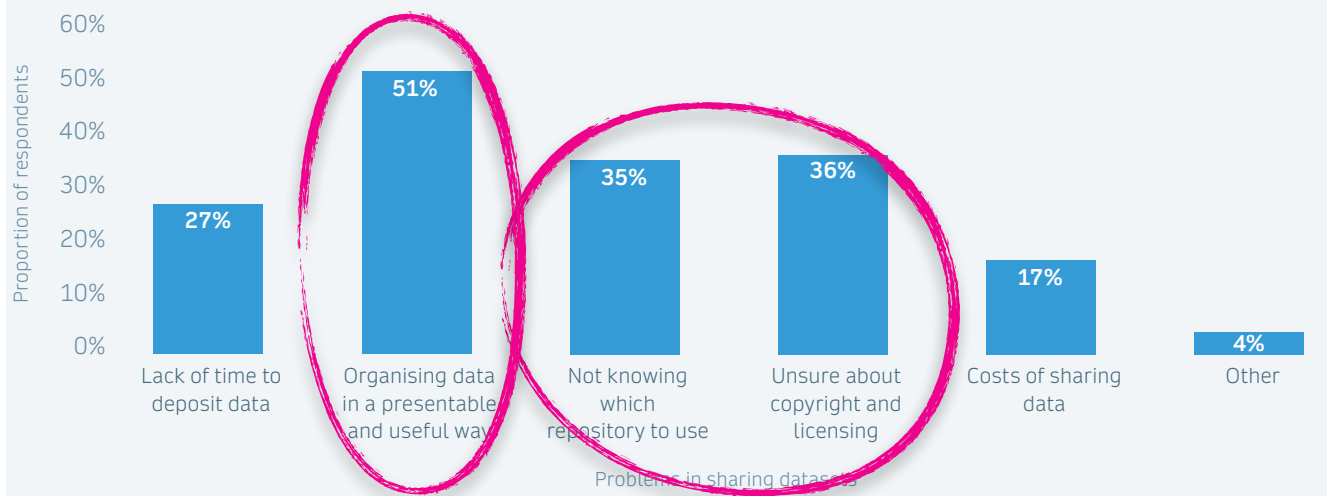
Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) (2016).



# The Main Challenges in Publishing Experiments

- A survey gives insights into the main challenge to data sharing.
- ExPaNDS wants to support scientists in:
  - Choosing a suitable repository and a license,
  - Describing relationships between methods and data in a (training) workflow.
- Our overall goal is to support the **F**indable **A**ccessible **I**nteroperable **R**eusable principles!

Figure 12: Problems in sharing data in the physical sciences (n=487)

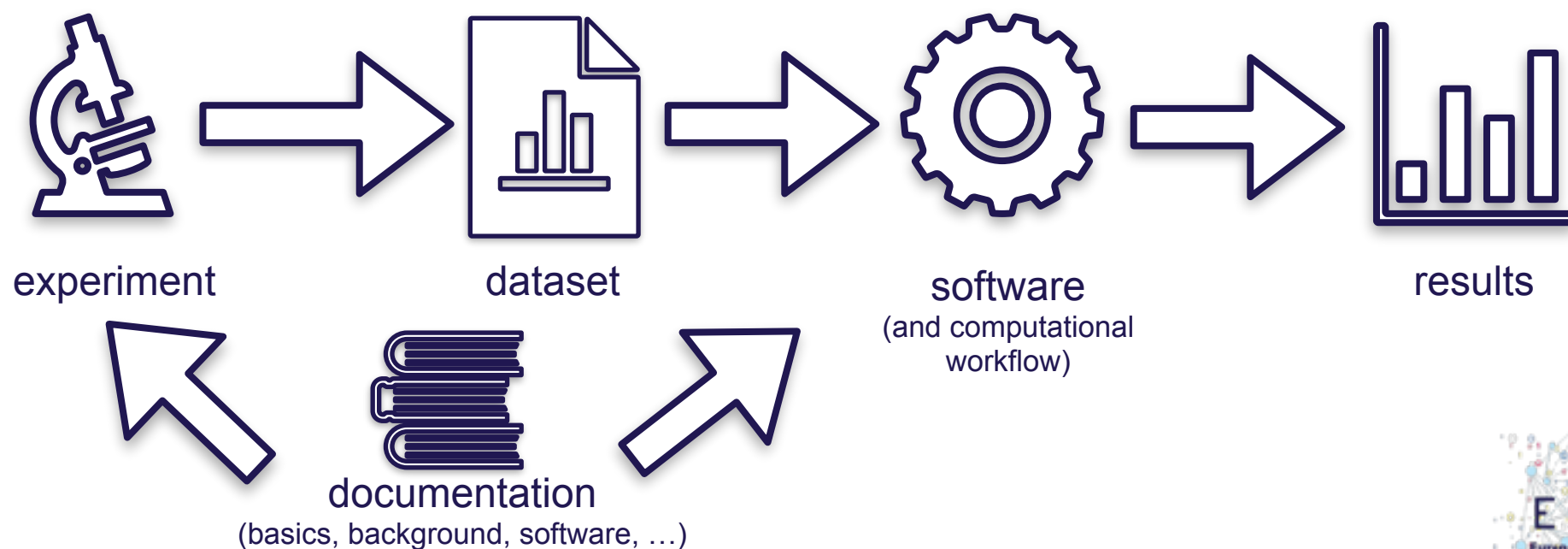


Springe Nature Whitepaper (2018): “Practical challenges for researchers in data sharing”, David Stuart, Grace Baynes, Iain Hrynaszkiwicz, Katie Allin, Dan Penny, Mithu Lucraft, Mathias; <https://doi.org/10.6084/m9.figshare.5975011.v1>



# Components of a Comprehensible Experiment

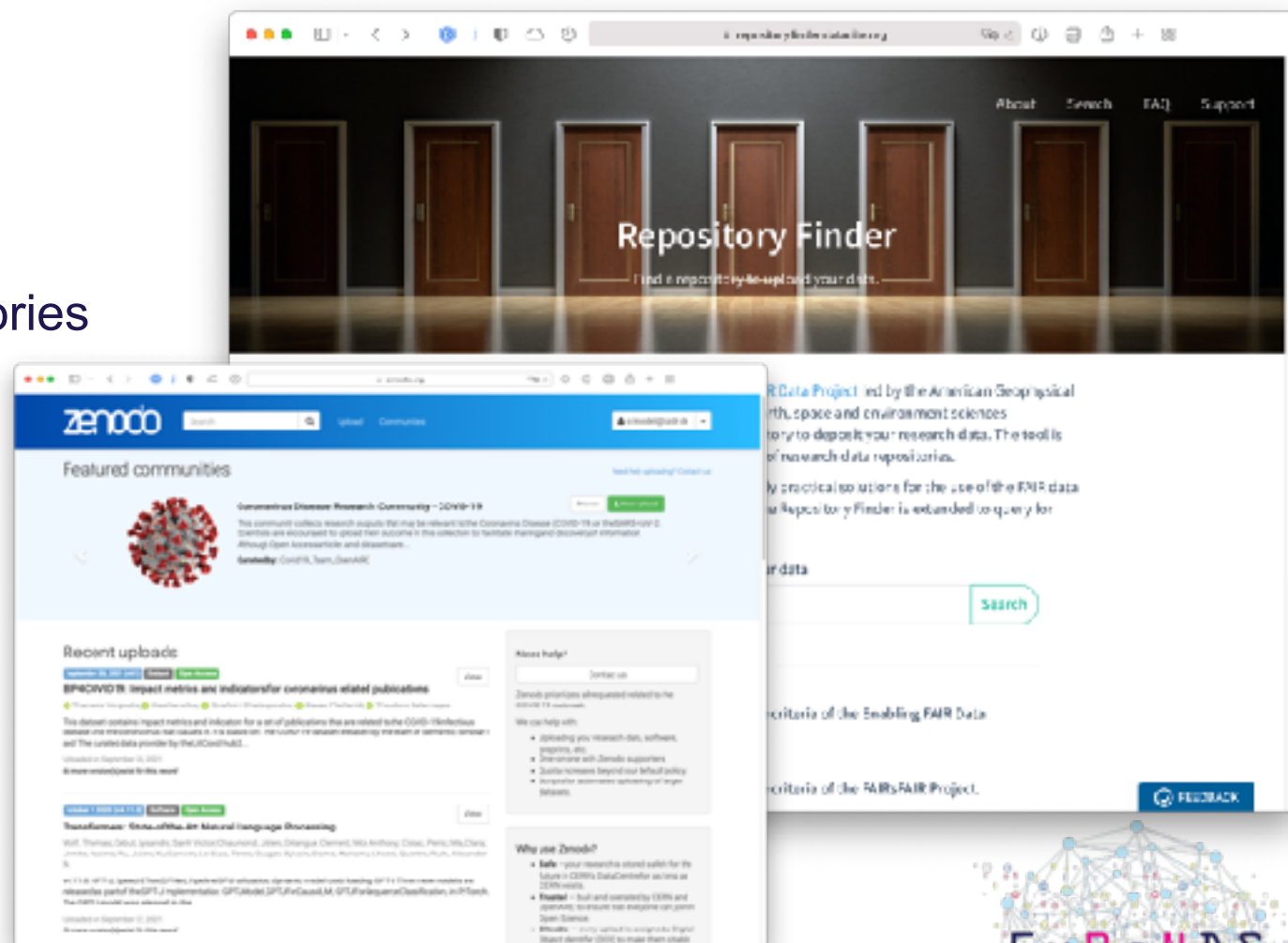
- Dataset(s) with raw and post-processed data,
- Documentation (Electronic Logbook, ...),
- Software (source code) and
- Overall workflow describing how all belongs together.



# Repositories for Data Publication

<https://repositoryfinder.datacite.org>

- Institutional repository,
- General-purpose open-access repositories or
- Domain specific or curated repositories can also be found with:



This project receives funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641

# Choosing Formats and Dataset Structures

## Guidelines

- Non-proprietary,
- Unencrypted,
- Uncompressed,
- In common usage by the research community,
- Adherent to an open, documented standard.



Name	Size	Download
Catalog of Genets.pdf	271.2 MB	Preview Download
ms2-0f7d0c9a8f81e277b6e8b8b1		
6-3-5-7-MMM-Data.txt	185.7 MB	Download
ms2-0f7d0c9a8f81e277b6e8b8b1		
GCES Eye & Dist. Detector, Jr.	17.9 kB	Download
ms2-0f7d0c9a8f81e277b6e8b8b1		
GCES Event-day Distribution, Jr.	17.9 kB	Download

## Further hints

- The dataset should have a self-explanatory and intuitive structure.
- A (short) description of the data (not the experiment) should be available in the datasets metadata.



# Licences for Research Data



- Every digital object needs a license.
- Without license nobody has the right to download the data objects...
- For research data we recommend:



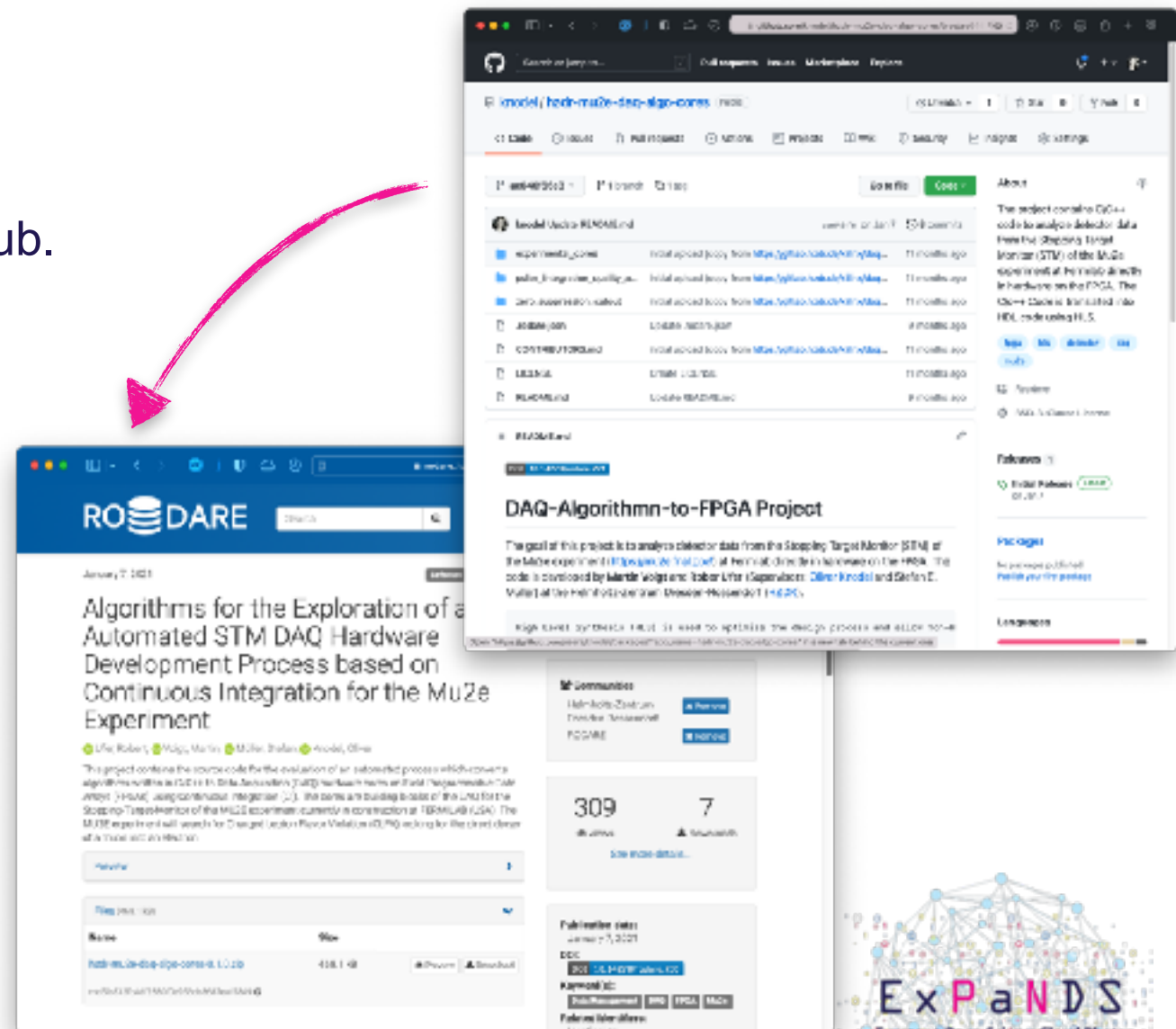
- With **Attribution 4.0 International (CC-BY-4.0)** you are free to:
  - Share: copy and redistribute the material in any medium or format.
  - Adapt: remix, transform, and build upon the material for any purpose, (even commercially).
  - Conditions: give appropriate credit, provide a link to the license.
- **Attribution-NonCommercial 4.0 International (CC-BY-NC-4.0)** restricts
  - NonCommercial: You may not use the material for commercial purposes.
- For software we recommend to use: <https://choosealicense.com>





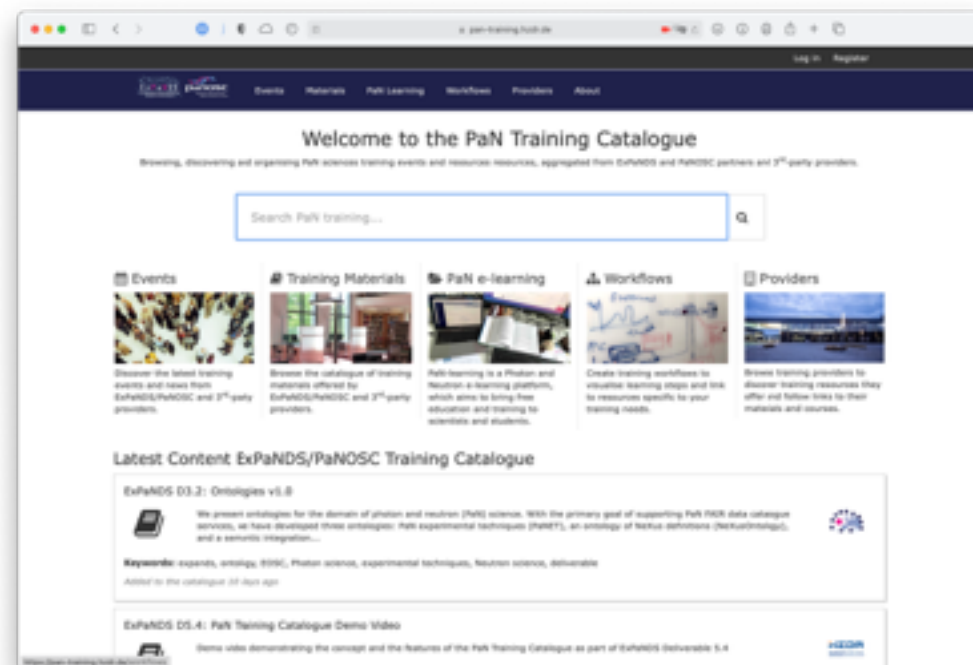
# Software Publication

- One of the best places for collaborative software development is GitHub.
- A software can be published with a release on GitHub
- With an additional uploaded to a data repository it is also possible to:
  - Add additional metadata (authors, title, description, license, ...),
  - Obtain DOI to be citable and
  - Increase visibility.



# The PaN (Training) Catalogue Supports Workflows

- Bringing all components of a research experiment together in a **comprehensible** and **reproducible** description is a challenge...
  - In the ExPaNDS project a PaN Training catalogue was developed to provide an easy-to-use environment to:
    - Provide references (URLs) to external training material (website, video, e-learning courses, ...) and
    - (Training) Workflows to show dependencies between different kinds of materials.
- ↳ A workflow used to describe an experiment is also in a certain way a training workflow!



# PaN Training or Experiment Workflow

**Pulmonary arterial hypertension research**

**Introduction**

**Dataset for Synchrotron Imaging Experiment**

**Local Jupyter Notebooks**

**Cloud Resources EOSC**

**Pulmonary arterial hypertension research**

This training workflow shows the combination of the demonstrations from the EXPANDS end-term review.

The workflow consist of three steps:

- Basic material to introduce the problem.
- The associated dataset with the related scientific publications and
- Source code and reference to the Jupyter notebooks.

**keywords:** synchrotron, imaging, jupyter notebooks, python, pulmonary arterial hypertension

**Target audience:** research data scientist, life scientists

**Primary user:** Medics

**License:** Open Data Commons Attribution License 1.0

**author:** Keesha Wess [ORCID], David Kneller [ORCID]

**Contributors:** Raymond Makoa [Lund University], Gabe Hirsch [PSE]



This project receives funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641

# Roadmap and Next Steps

- The PaN Training catalogue is intended for all PaN facilities, scientists, students, ...
- AAI is coming late 2021, but user accounts are already possible upon request.



This project receives funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641

# PaN Training Catalogue

<https://pan-training.hzdr.de>

## Demo Video

<https://pan-training.hzdr.de/about>

# And now: Live Demo of our Catalogue



This project receives funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641