

HeFDI Data Talk

Date	Topic	Presenter
18. November 2022	DAPHNE4NFDI. Data from Photon and Neutron Experiments	Lisa Amelung (Deutsches Elektronen-Synchrotron DESY, Hamburg)



Abstract:

DAPHNE4NFDI is the first (inter-)national attempt to bring together users and large-scale research facilities to create a comprehensive infrastructure to process research Data from PHoton and Neutron Experiments (DAPHNE) according to the FAIR principles. Our community faces a common need for high-level, rapid data analysis and the challenge of implementing research data management for increasingly large and complex datasets. All this involves not only a broad range of scientific disciplines and stakeholders, but also the connection to complex instrumentation and IT. Within the talk, Lisa Amelung would like to inform about the DAPHNE4NFDI consortium and discuss (current) challenges and future tasks.

About the HeFDI Data Talks:

The HeFDI Data Talks are a bi-weekly open information and discussion event focused on data management in the context of science, in which relevant NFDI consortia as well as research data management services present themselves. The series discusses current topics and presents numerous – including local and regional – tools and services. The HeFDI Data Talks are an offer of the HeFDI Initiative (Landesinitiative HeFDI), which is funded by Hesse's Ministry for Science and Arts (HMWK).

DOI: <https://doi.org/10.5281/zenodo.7414563>; License information: Creative Commons Attribution 4.0 International ([CC BY 4.0](#))



DAPHNE4NFDI

DAta from PHoton and Neutron Experiments

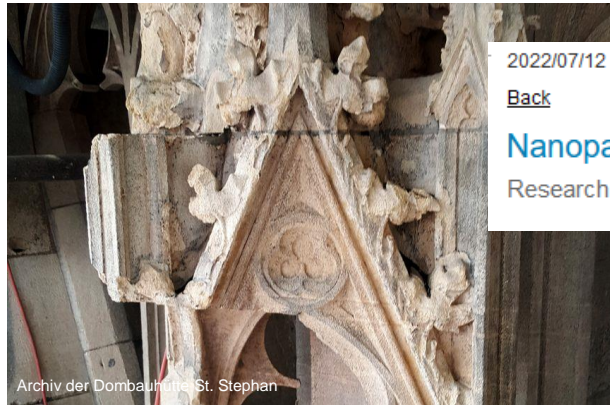
HeFDI Data Talk, 18.11.2022

Lisa Amelung | Deutsches Elektronen-Synchrotron DESY, Hamburg



What is „DAta from PHoton and Neutron Experiments“?

The basis for many scientific discoveries – in many different fields!



2022/07/12

[Back](#)

Nanoparticles can Save Historic Buildings

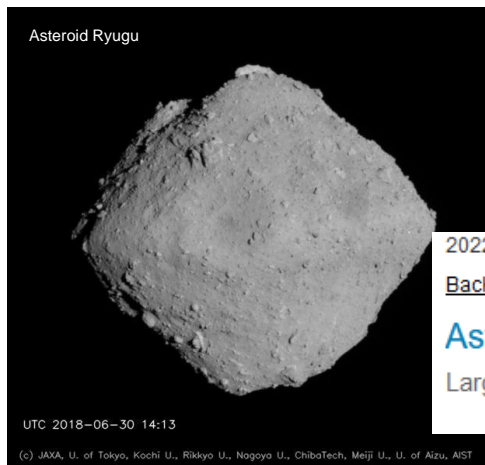
Research team investigates nano-crystals for greater strength of sandstone at DESY

2021/04/02

[Back](#)

DESY X-ray lightsource identifies promising candidates for COVID drugs

Existing active substances bind to key virus protein



2022/09/23

[Back](#)

Asteroid dust in the X-ray beam

Large international campaign analyses samples from asteroid Ryugu

2016/08/26

[Back](#)

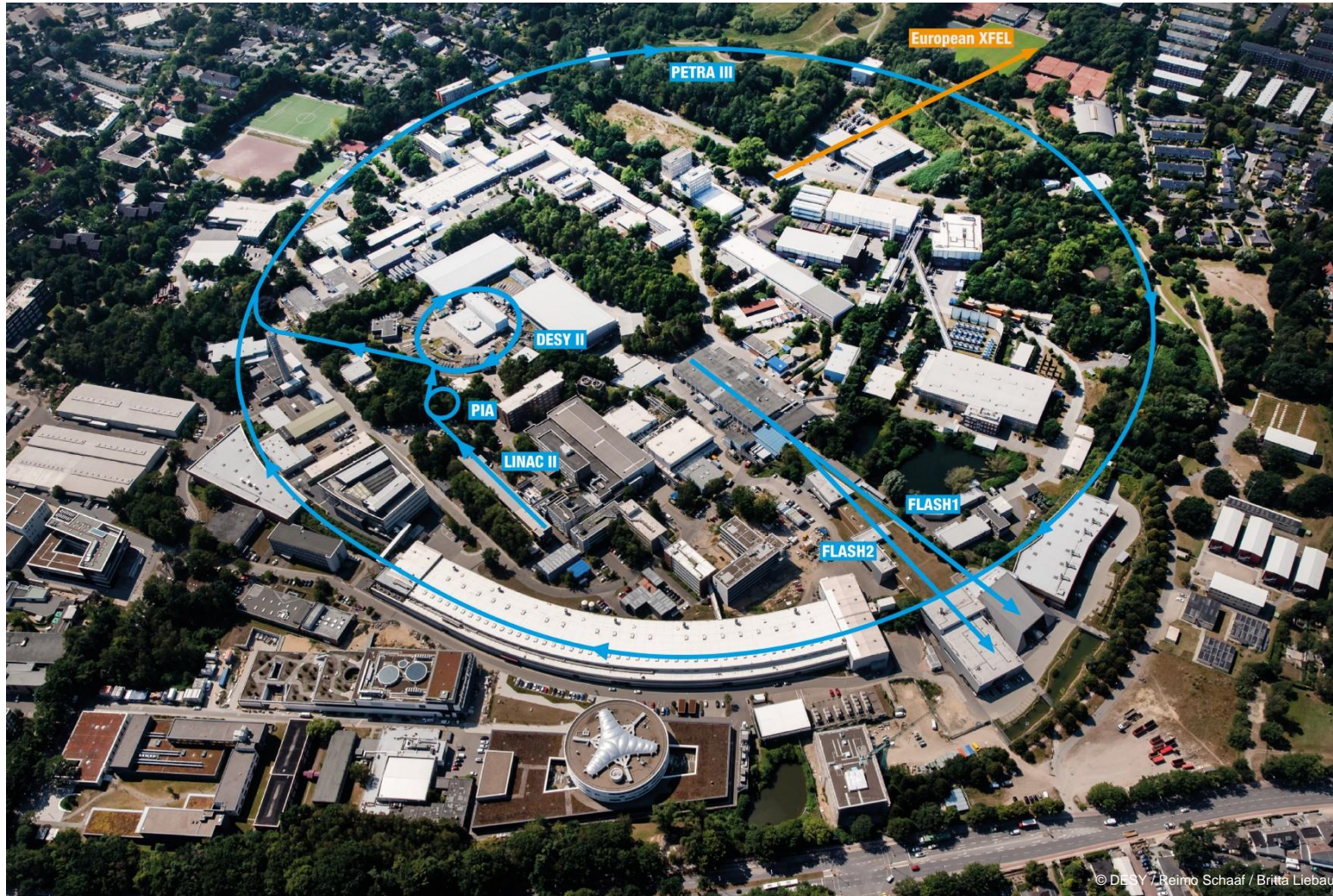
'The Scream' and the secret of the white spots

Why researchers look for bird droppings on Edvard Munch's masterpiece with DESY



Deutsches Elektronen-Synchrotron DESY, Hamburg

One example of a large scale research facility



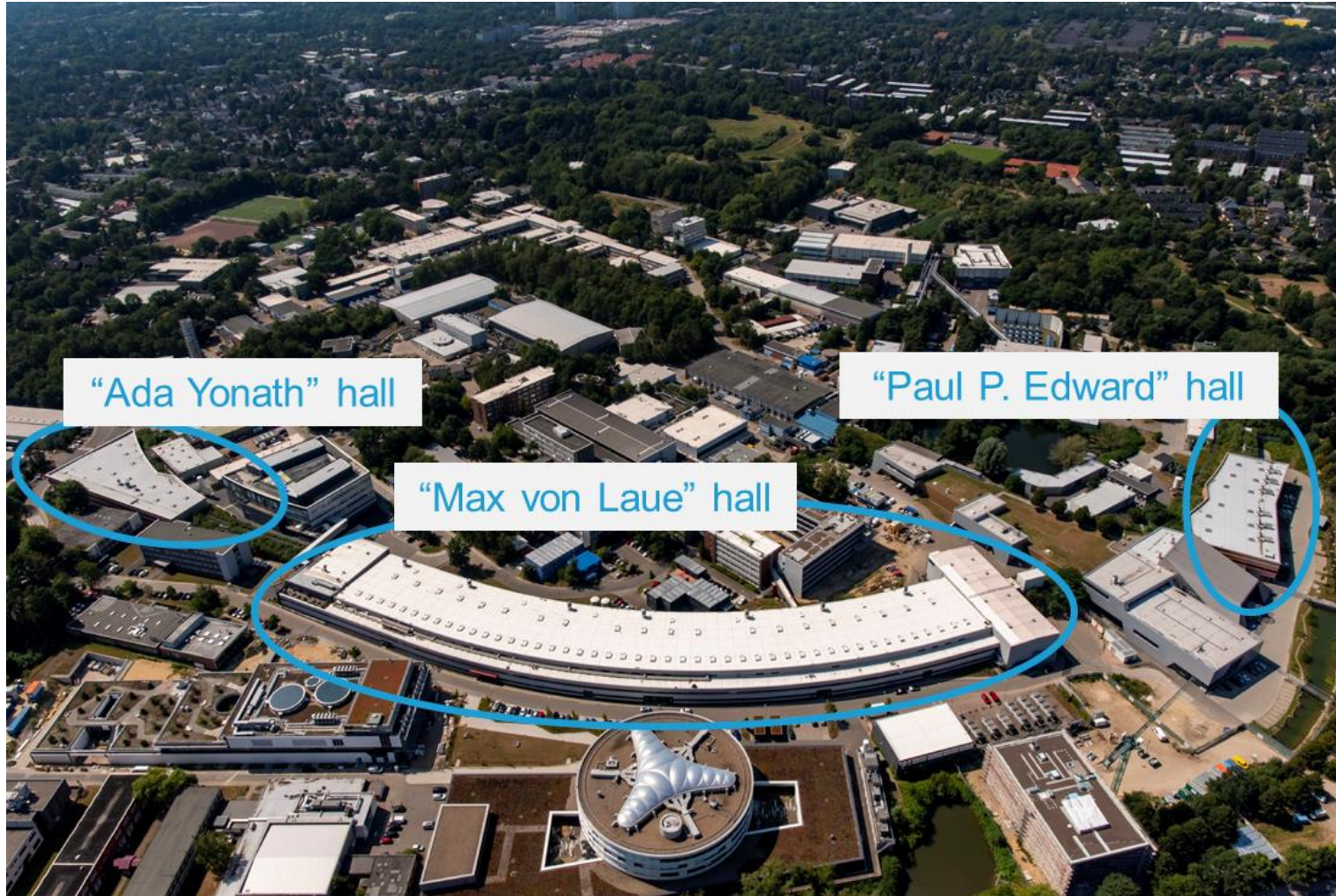
Large in terms of “scale”

- Length **European XFEL**: 2.1 km
- Length FLASH: 315 m
- Radius PETRA III: 2.3 km

Staff: 2,700 (300 in Zeuthen)
+3000 external researchers/year

PETRA III @ DESY

X-ray radiation source

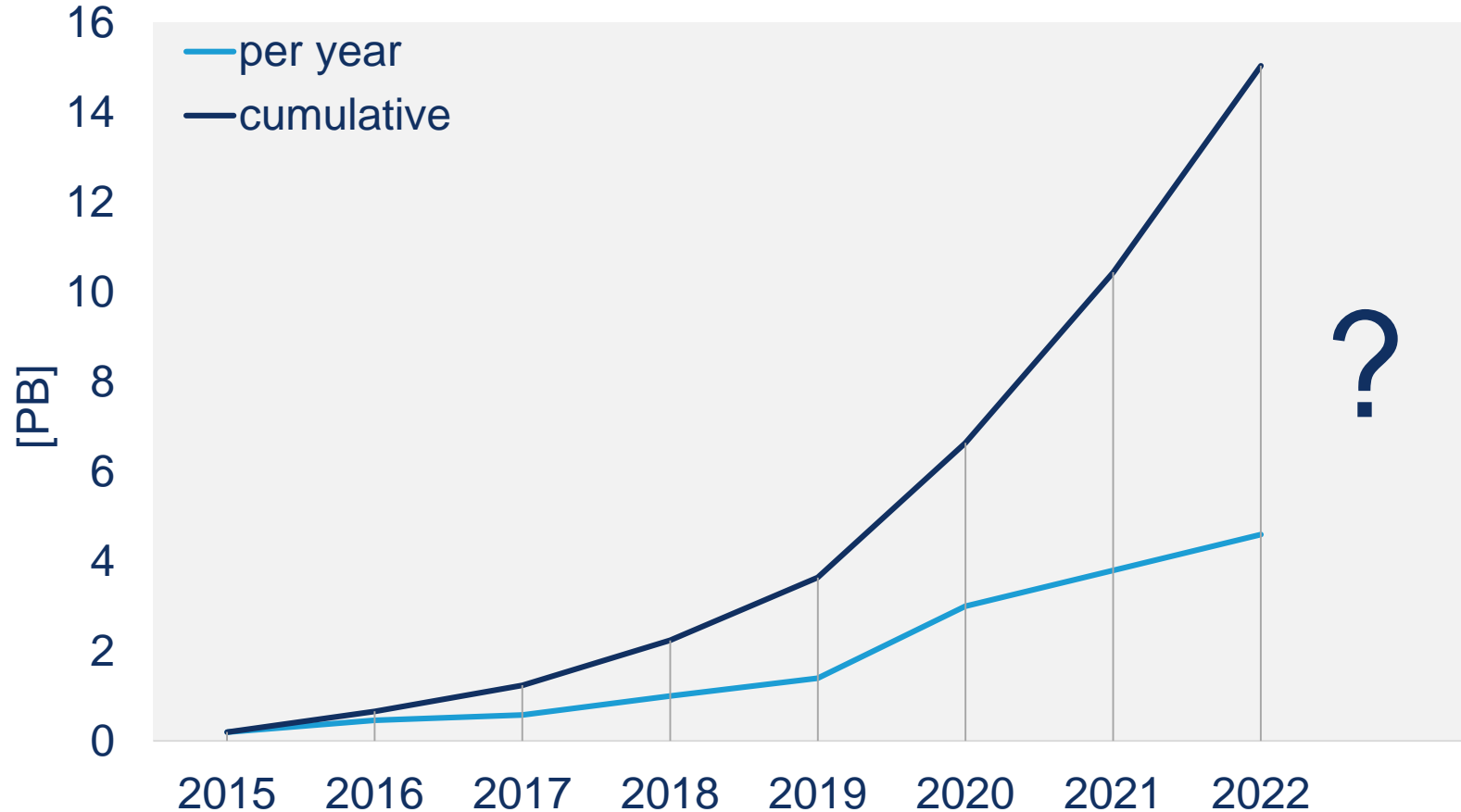


Large in terms of “opportunities”

- 24 different beamlines: e.g. diffraction & scattering, spectroscopy, bio-imaging, crystallography
- on-site visits/remote users or mail-in & data only

Yearly amount of data produced at PETRA III

It's a lot...

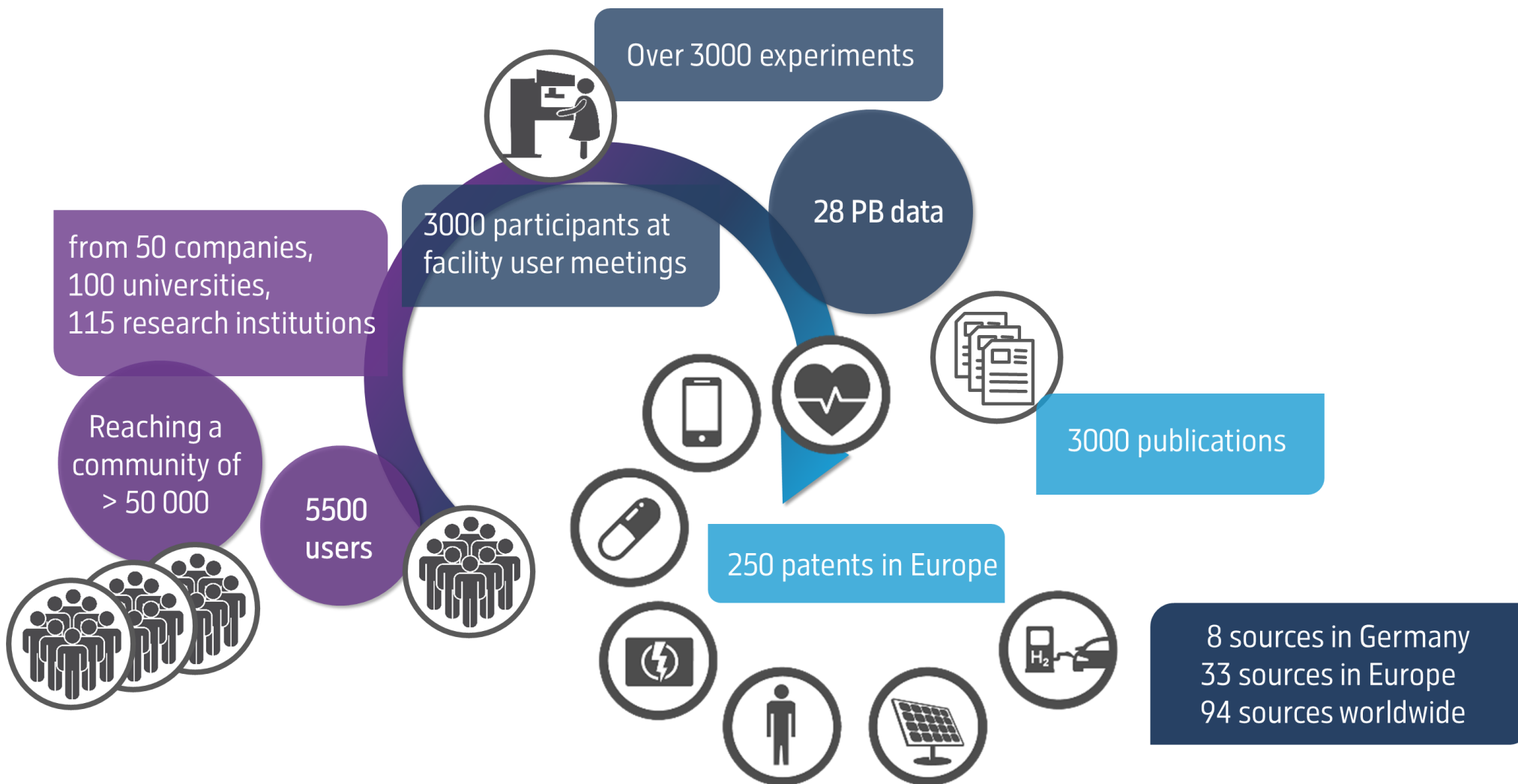


Large in terms of “data”

- 2021: ca. 3.8 PB of data
- 2022: ca. 4.6 PB expected
- 2022: ca. 15 PB total storage
- increasingly large and complex datasets
- when/if to delete data sets?

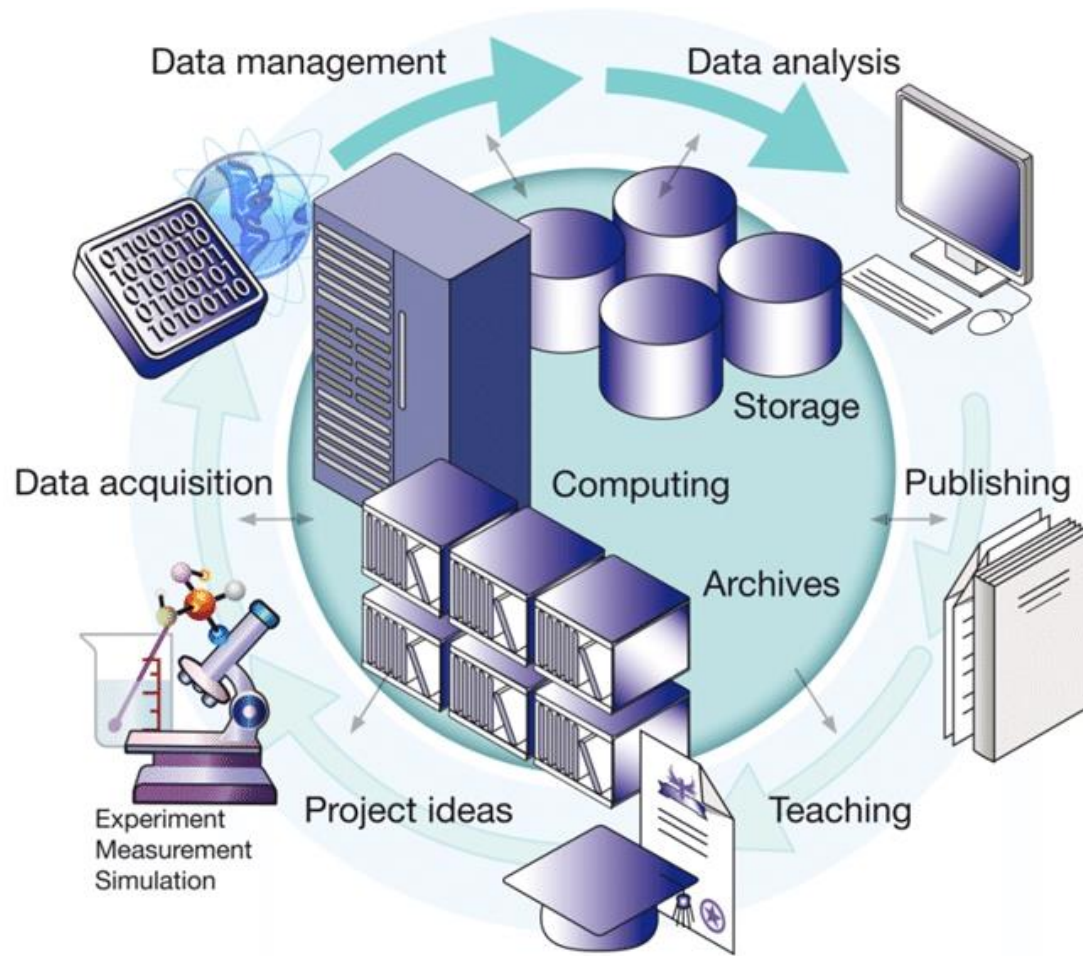
Research with photons and neutrons in Germany

Numbers per year



The common needs and challenges

And where the fundamental idea of (DAPHNE4)NFDI comes into play



Jung, C. et al. (2014). Optimization of data life cycles. Journal of Physics: Conference Series. 513. 032047. 10.1088/1742-6596/513/3/032047.

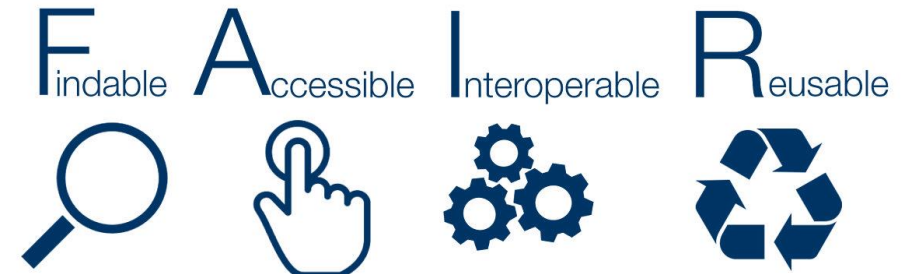
Status

- increasingly large and complex datasets
- high-level rapid data analysis
- a broad range of scientific disciplines and stakeholders
- the connection to complex instrumentation and IT



Goal

- research data management (in retrospect)



SangyaPundir (https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg), Colour, CC-BY SA 4.0

DAPHNE4NFDI

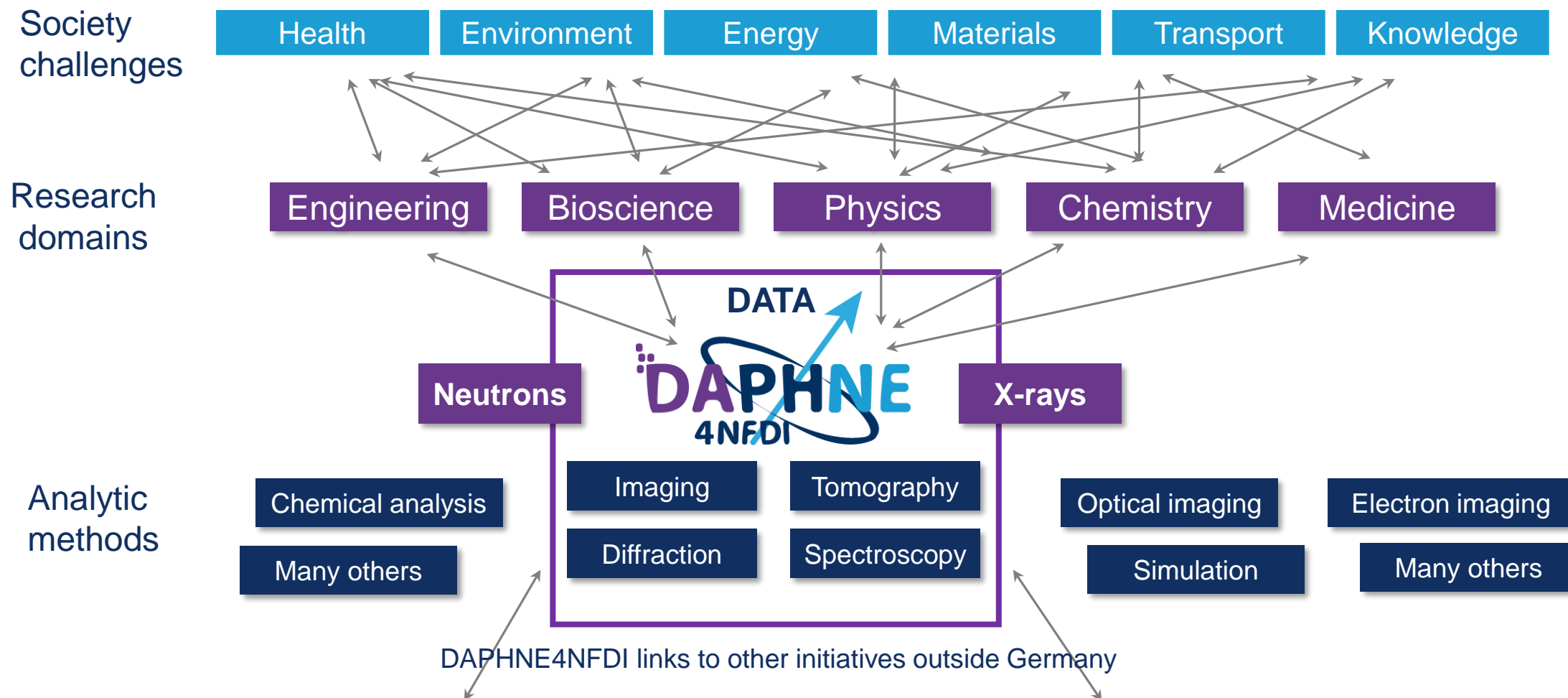
Who we are

- DAPHNE4NFDI is funded by the DFG as part of the NFDI
- 18 funded partners: 11 universities and 7 large scale facilities
- more than 100 participants within Germany and Europe
- Scope of 5+ years (2021-2026) and €16.7M
- Coordinated by DESY in collaboration with the user community



Our research community impacts on global challenges

Impact extends far beyond physics or materials science community



DAPHNE4NFDI Set-up

Who is involved?

Executive Board - Task Area leaders

- TA1: Bridget Murphy (CAU) & Wiebke Lohstroh (TUM)
- TA2: Anton Barty (DESY, Speaker) & Frank Schreiber (U Tübingen)
- TA3: Sebastian Busch (Hereon) & Tobias Unruh (FAU Nürnberg)
- TA4: Jan-Dierk Grunwald (KIT) & Astrid Schneidewind (FZJ)
- TA5: Astrid Schneidewind (FZJ) & Christian Gutt (U Siegen)
- TA6: Anton Barty (DESY, Speaker)

Technical Advisory Board (TAB - tba)

- state-of-the-art solutions: technical solutions of software management, roll-out, version control etc. and IT infrastructures hosted at the facilities
- synchronize overarching strategy across all facilities & wider landscape of European partners



Infrastructure HIFIS

- Cloud service (Sync & Share)
- Software development platform (GitLab)
- Communication (Mattermost etc.)
- Event announcement (HIFIS events)

Community

- Scientist from large scale facilities, universities, research institutions basis: use cases
- Universities and school kids (curricula)
- Industry

International Advisory Board (IAB - tba)

- issues relating to the international cooperation and ongoing alignment with strategic developments taking place in Germany
- embedding our activities outside of the domestic setting, gain international visibility

Our task areas

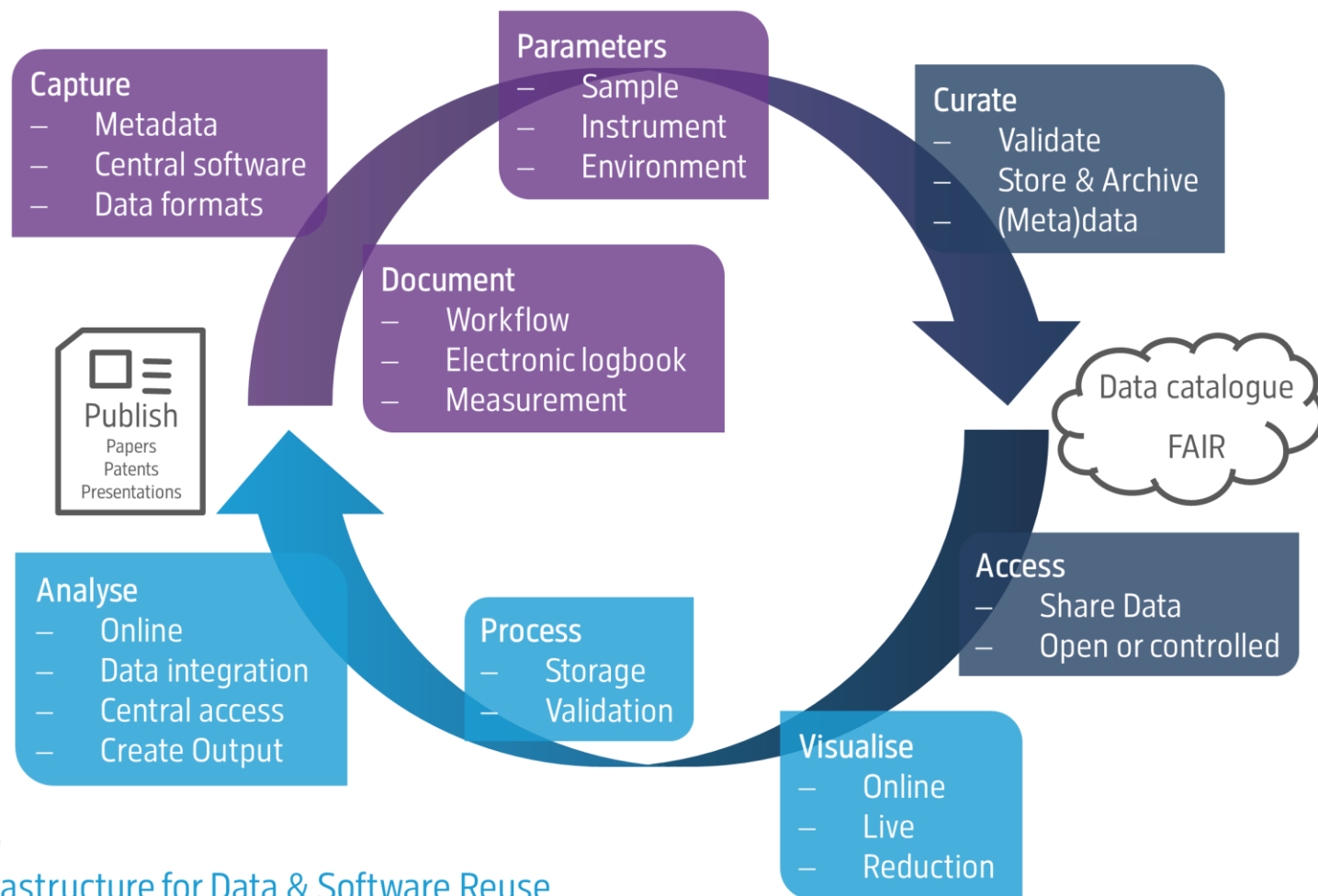
TA1-TA6

Collect

TA1: Managing Data Production

Store

TA2: (Meta)data Repositories & Catalogues



Outreach and Dissemination

TA4: Establish & enhance awareness of FAIR principles

External Communication and Policy

TA5: Common data policy & alignment with European partners

Project Management

TA6: Project coordination & Administration

Evaluate

TA3: Infrastructure for Data & Software Reuse

Task Area 1: Managing Data Production

Enabling re-use and repeatability of results, ideally searchable

Before experiment

Proposal

- Proposal number (ID)
- Science Motivation
- Experiment concept
- Technique(s)
- Sample & environment
- Instrumentation
- Science team

Facility

- Which facility?
- Beam parameters
- Instrumentation
- Detectors
- Proposal

Digital Sample ID

- 'DOI for samples'
- Barcoding?
- Cross-link to other consortia

During experiment

Facility logs

- Experiment ID (directory)
- Beam parameters
- Motor positions
- Instrument configuration
- Sample environment
- Detector calibration

User record

- Instrument configuration
- Actual samples used
- Ideally a digital sample ID
- Actual sample environment
- Changes to original plan
- What happened when
- Run log (data lookup table)
- Currently: Paper logbooks, google sheets etc.

After experiment

Analysis

- Data analysis steps
- Not all data is useful (runs)
- Intermediate data, code, scripts

Publication

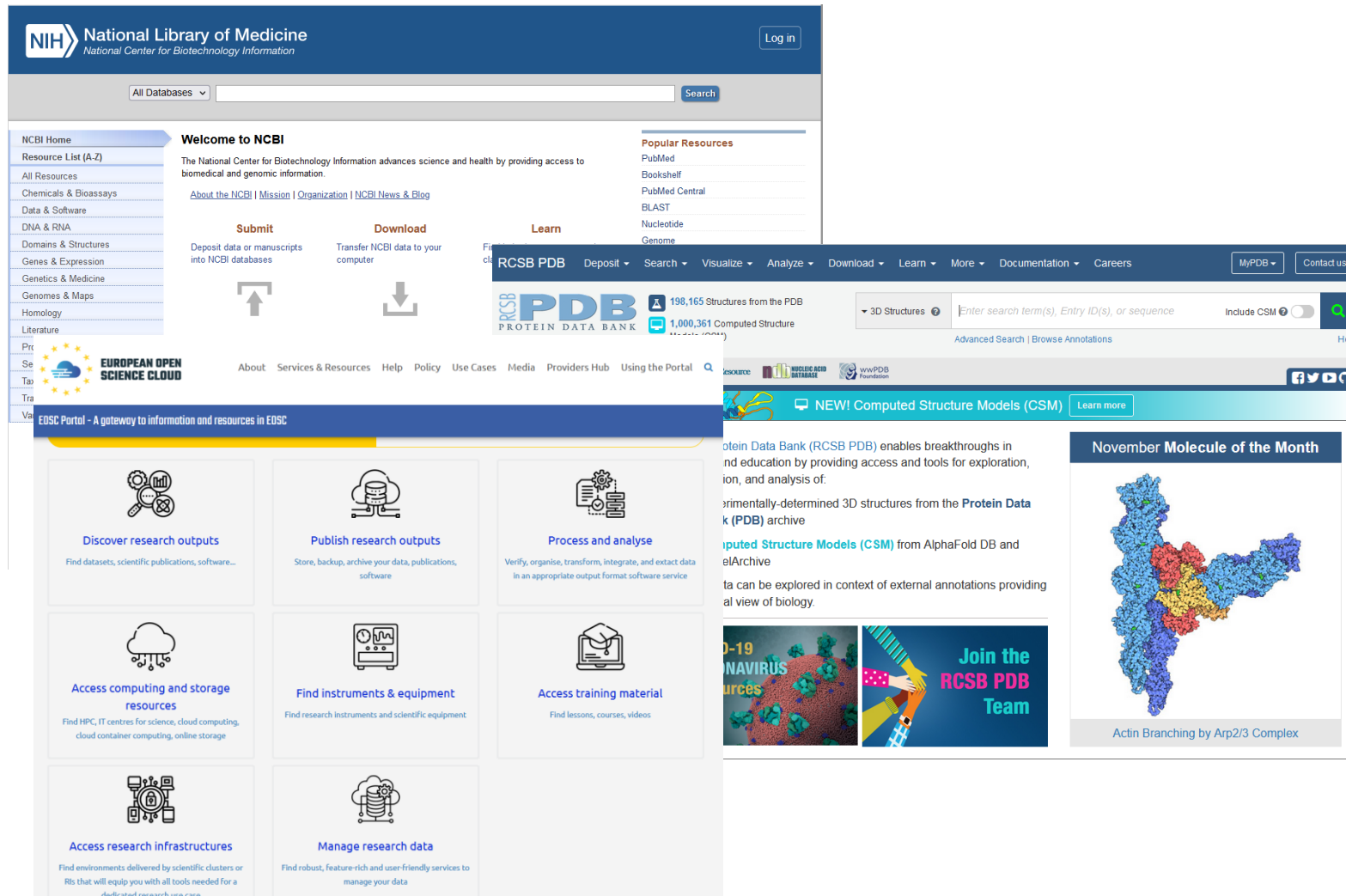
- Findable and searchable
- Description of process
- Citation, DOI reference
- Sometimes data is deposited (PDB, CXIDB)
- May use a subset of data, or data from many experiments

Re-use

- Check and verify results
- Improve the analysis
- Re-use code for new work
- Build on past data

Task Area 2: (Meta)data repositories & catalogues

A place to find published data - and in some cases the ability to reprocess data



Status

- No (common) solution

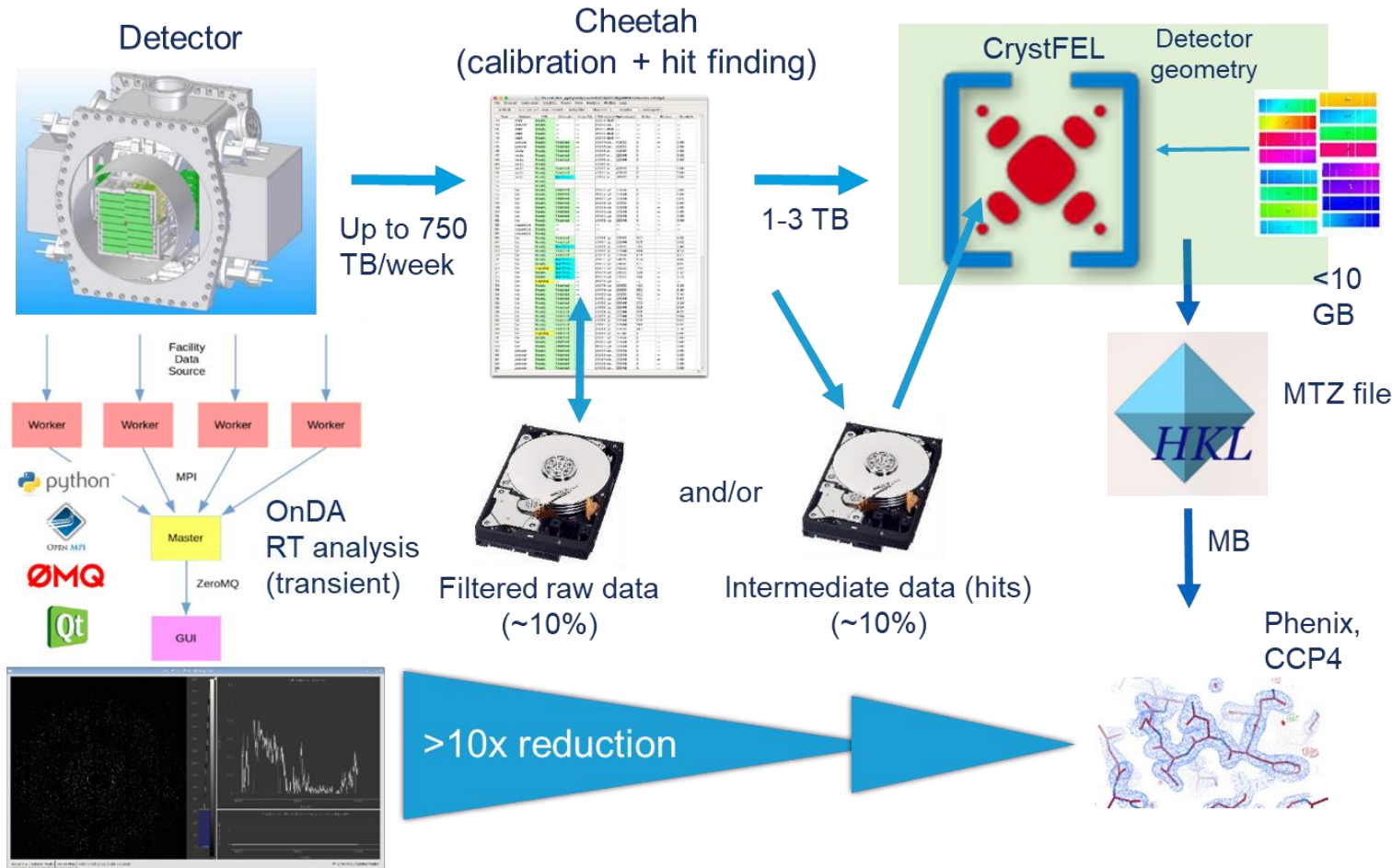


Goals

- Specification of metadata standards, sample description
- Establish federated, interlinked data repository for participating facilities
- Reference data bases: use cases (e.g. EXAFS)

Task Area 3: Infrastructure for data & software reuse

Analysis software infrastructure is currently fragmented and hard to re-use



Status

- Too much reliance on undocumented code and scripts

Goals

- Accelerate science outcomes and sustainability by leadership in a user software ecosystem
- Remotely available user software for re-use by all on facility infrastructure

Task Area 4: Outreach and Dissemination

The NFDI consortium as a role model and educator

Inside DAPHNE

- Explain, discuss and disseminate
- Organize
- Provide an exchange platform
- Support pilot projects and Use Cases

Actions

- Create awareness: conferences, fairs, (public) events
- Transfer knowledge: facility and (future) user
- Support juniors: network for young scientists
- Encourage teenies: (summer) schools

DAPHNE4NFDI within NFDI

- Connect specialists, use experiences
- Organize and support common Use Cases

Outreach to society and industry

- Inform industry about
- Connect society and industry to highlight developments
- Encourage to re-use

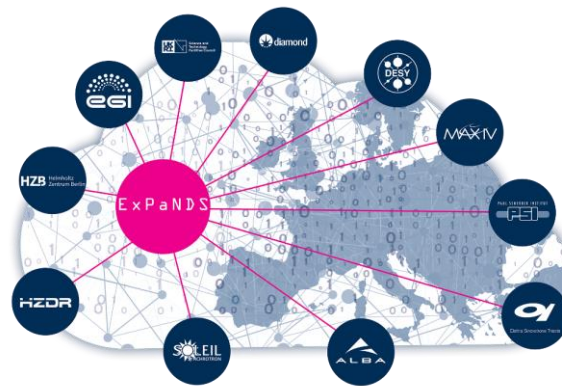


Task Area 5: External Communication and Policy

Integration into the international research community



**EUROPEAN OPEN
SCIENCE CLOUD**



Status

- embedded in a worldwide network of > 30.000 synchrotron and neutron users



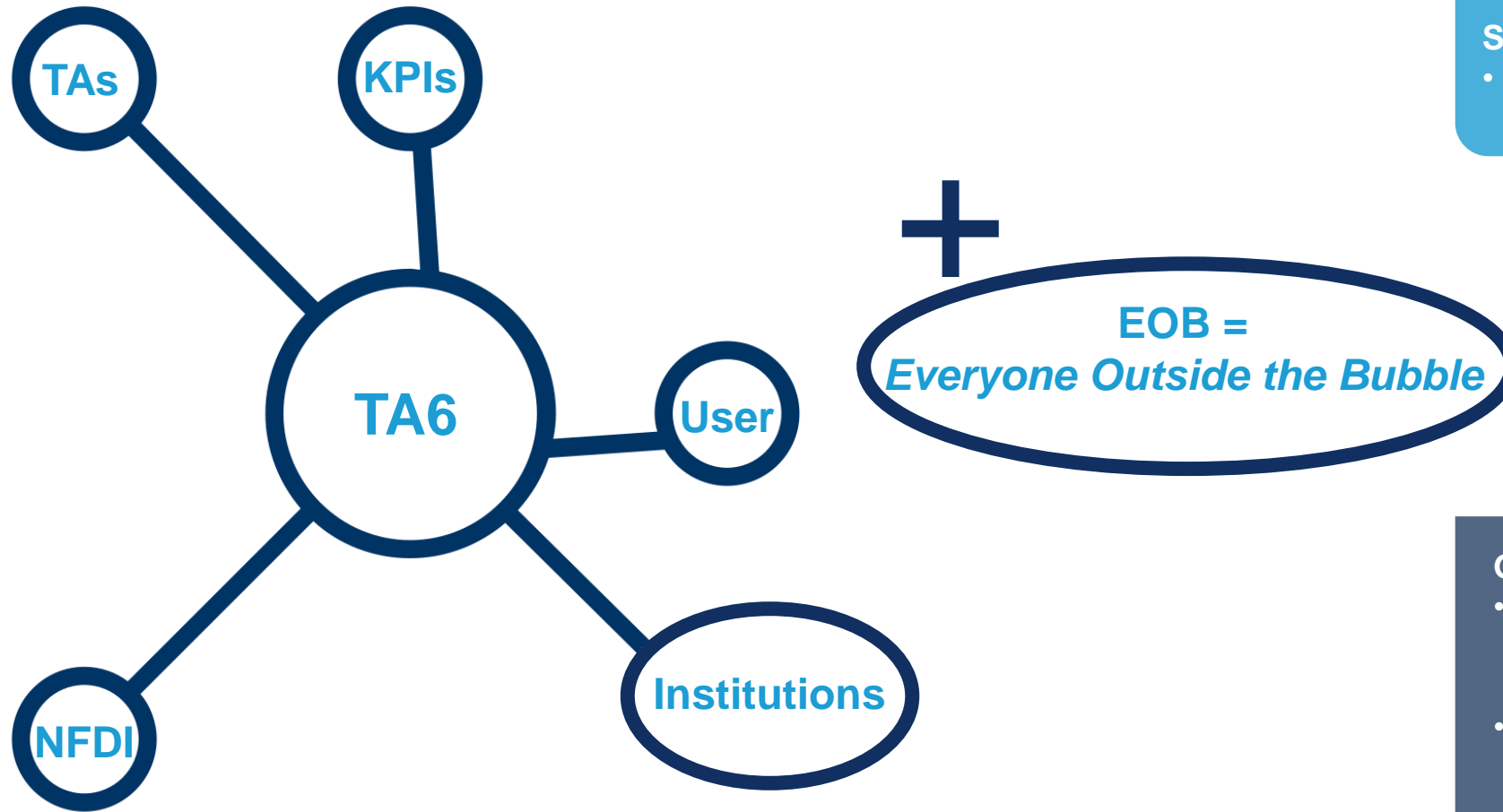
Goals

- Strengthen connections to (European) user organizations and facility organizations
- Sustain already existing outcomes of former/other initiatives
- Long-term:

Perpetuate DAPHNE4NFDI

Task Area 6: Project Management

Managing, evaluating and steering



Status

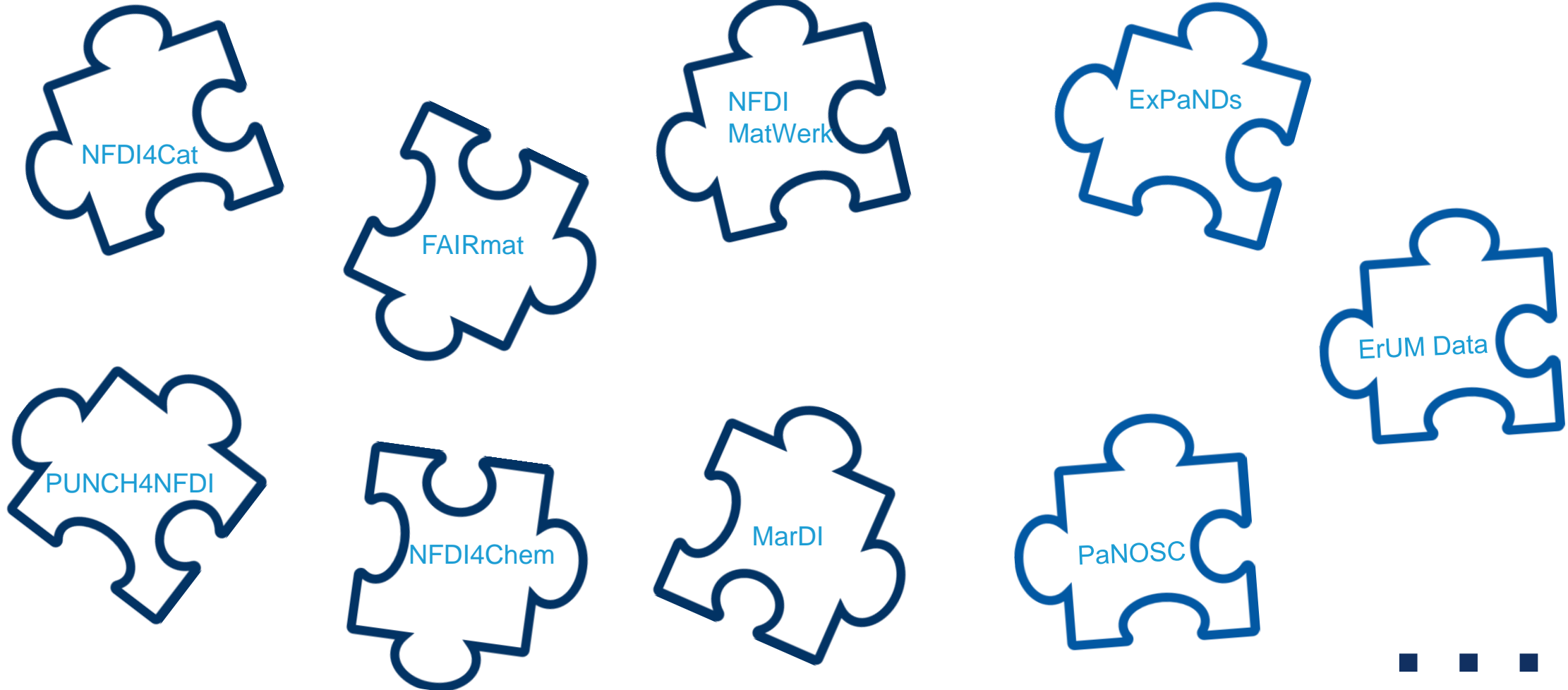
- Stakeholders and general project set-up are clear

Goals

- Provide (more) infrastructure for a successful project performance
- Create a team out of individual(s) groups
- Constantly evaluate project development

Collaborations (within NFDI)

A broad network gives a broader view



Upcoming event

Feel free to join @ DESY or remotely – Networking lunch included



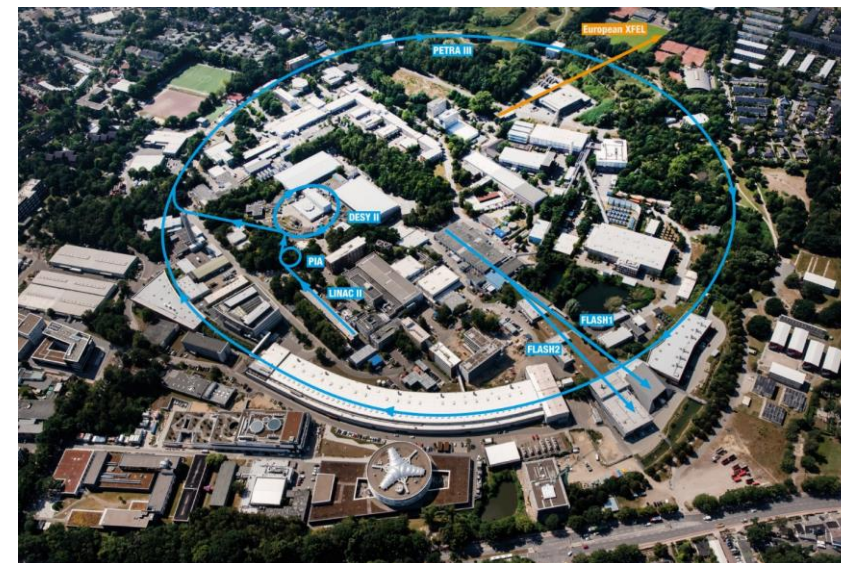
NFDI Physical Sciences Joint Colloquium

7 December 2022 @DESY, Hamburg



John R. Helliwell

*Applying the FAIR Principles to Crystallography Data Publication
– a use case for DAPHNE4NFDI?*



<https://events.hifis.net/e/Helliwell>

Thank you for your attention!

Contact us!

www.daphne4nfdi.de

Lisa Amelung

Project Coordinator DAPHNE4NFDI

Deutsches Elektronen-Synchrotron DESY
Notkestrasse 85 - build. 49a, room 216
22607 Hamburg

Lisa.Amelung@desy.de | +49-40-8998-5866