

# EXPANDS

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

## **Establishing PaNET at all RIs**

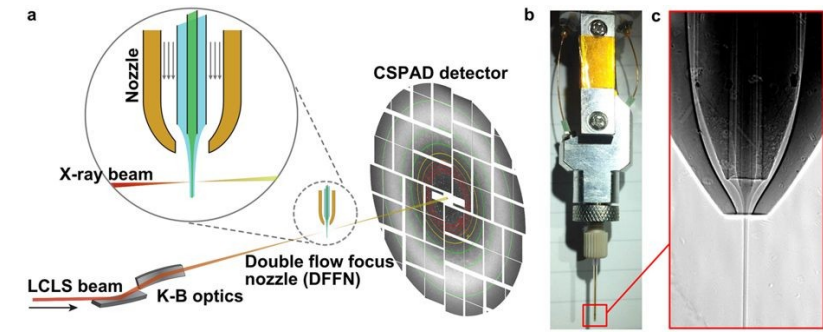
**Paul Millar**

2022-06-14

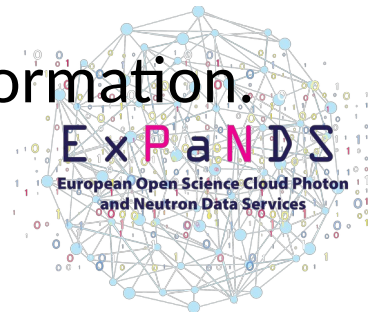
*PaNOSC & ExPANDS face-to-face meeting  
Prague, Czech Republic*



# What is PaNET?



- A set of **opaque PIDs**, each representing a scientific technique; e.g.,  
<http://purl.org/pan-science/PaNET/PaNET01168> → serial femtosecond crystallography
- Techniques have **a label** (currently just English).
- Techniques are organised into **a hierarchy**:
  - “x-ray tomography” has less specific terms: “tomography”, “x-ray probe”.
  - “tomography” has more specific terms: “x-ray tomography”, “fluorescence tomography”, “absorption tomography”, ...
- Each technique can have a **definition** and **links** to other sources of information.



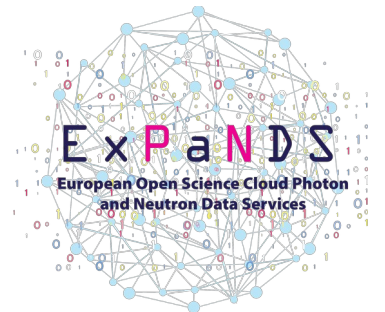
# Making PaNET useful

- Building ontologies is sometimes a useful exercise in itself.
- However, for PaNET, we want to use it to **identify things**:  
Primarily datasets, but other things may be tagged, too.
- PaNET information is difficult to add by an external agent.  
We really need the **facilities** to add PaNET information.
- Therefore, we need your (the facilities') help in making PaNET useful!



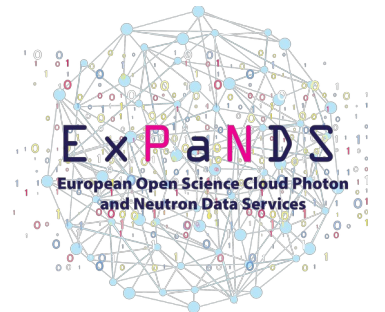
# Tagging datasets with PaNET

- Identify which technique (PaNET term) created the dataset.
- Make it easier to find FAIR data: more robust than a natural language search. The title or description fields:
  - Might not mention the technique.
  - Might be written in a different natural language.
  - Might use a different (equivalent) term to describe the technique.
  - Might use a more specific term, which doesn't match the search term.
- Machine-readable description of technique makes it easier to build derived data:
  - Which technique used at some specific instrument / beam-line / facility.



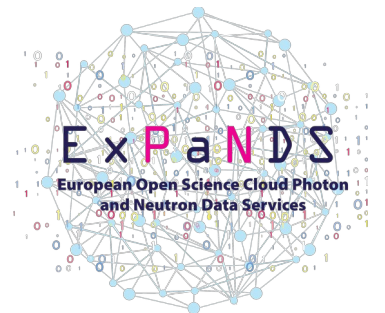
# What are the general steps?

- Figure out what is the **procedure** to deduce the PaNET term.
  - Each facility will (likely) need to do something different here.
  - This might involve adjusting your existing procedures.
- Update how datasets are added to the dataset catalogue:
  - Include PaNET information when creating a new dataset.
- Update dataset catalogue configuration to expose PaNET information.



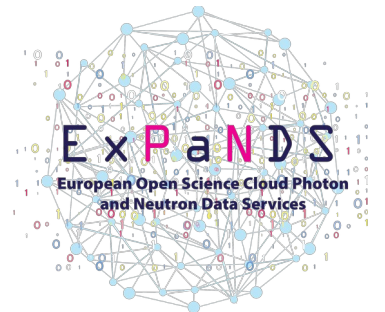
# Possible procedure:

- Go through each beam-line / instrument:
  - Look for the best PaNET term for the data it produces.
  - Is there no suitable PaNET term? Request a new term.
- Is the problem poorly constrained?
  - An instrument (or beam-line) might service multiple experimental techniques.
  - Additional information may need to be capture, perhaps:
    - from researchers, as part of the proposal entry process.
    - from beam-line scientist, as part of their normal activities



# What we're doing to help

- We're putting together a series of documents on PaNET.
  - These are called “Technical Notes”.
  - These have a narrow scope: offering advice on a specific problem.
- Some documents focus on popular catalogue software:
  - Currently two: one for ICAT, the other for SciCat.
  - These are (something like) “HOW TO”.
  - These documents exist, but are currently too immature to share.



# Summary

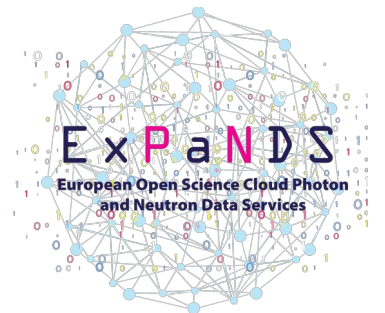
- We're looking for facilities that want to start adopting PaNET.  
We need datasets that have PaNET terms.
- Should we identify, for each facility, an individual who feel responsible for PaNET adoption.
- Should we establish a PaNET adoption discussion forum?

DISCUSS!

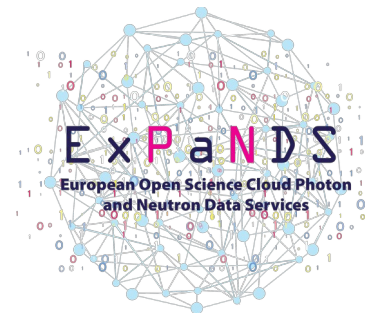




# Thanks for listening!

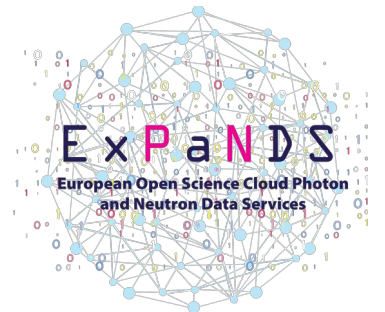


# Backup material



# PaNET can also be used for ...

- Identifying:
  - **Training material** – which technique(s) are covered?
  - **Facilities** – where is this technique used?
  - **Events and workshops** – which techniques are being discussed?
  - **Software** – which kind of data is processed?
  - **Services** – what kind of analysis?
  - **Pipelines** – what kind of data?
  - ...
- Linking to datasets:
  - Find suitable public dataset.



# Exposing PaNET

- PaNET information exposed via:
  - Querying datasets via PaN search API.
  - OAI-PMH supports PaNET via DataCite’s “subject”:  
Would allow B2FIND / OpenAIRE to support dataset selection via technique
  - Landing page: learn the technique from the DOI

