How SciCat could help to achieve goals of DAPHNE ?

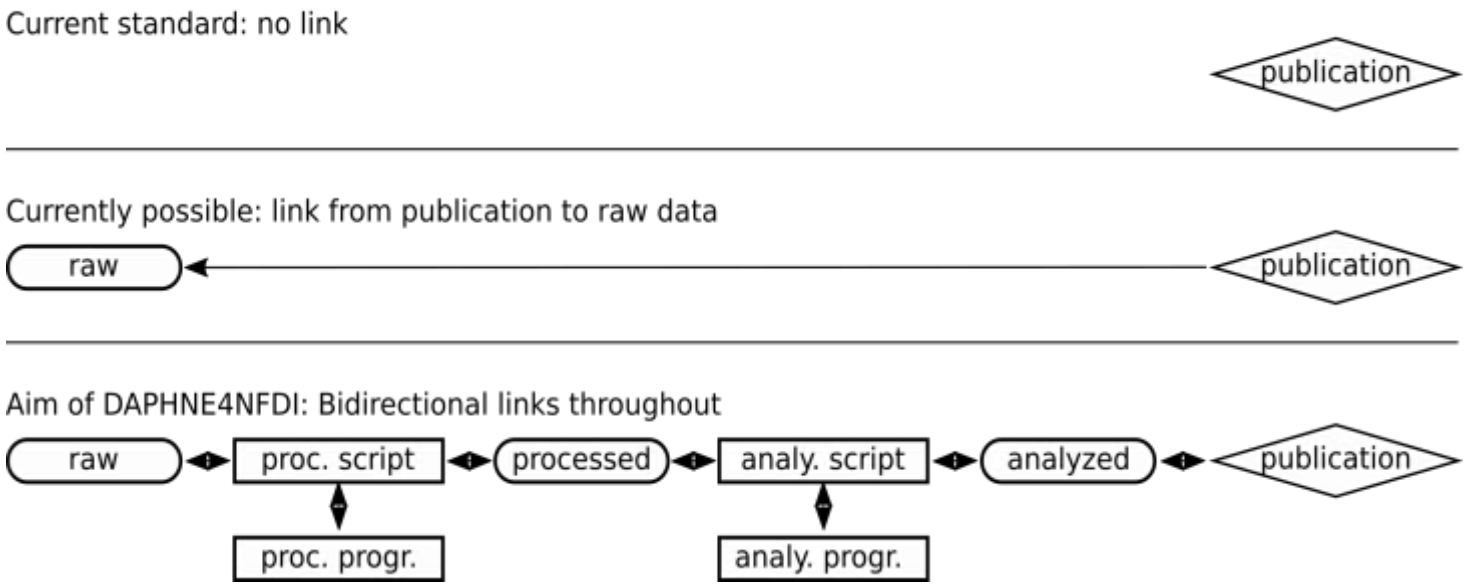

Figure from DAPHNE proposal: linking data and publication

How could DAPHNE contribute to the SciCat Project ?

Which contributions could we imagine?

uni lab perspective, non exhaustive list!

- Data ingestion
  - Web frontend to create dataset entries in SciCat
  - Upload tool for instrument control PCs in labs
    (work in this direction done at Rosalind Franklin Institute [1])
- Metadata Schema management
  - How to structure metadata of datasets and samples
- Simple production deployment
  - Provide easy-to-use SciCat installation
    that only needs minimal additional configuration to work in production
  - Provide access management that works for most DAPHNE participants
    (one option: hifis AAI, similar to what is used in Sync & Share @ DESY)

[1] https://www.rfi.ac.uk/projects/data-management-at-the-franklin/

Further features that could be thought of…

again, non exhaustive list …

- Integration of NeXus / hdf5 web-based data-viewer
  - (e.g., developed at ESRF/Panosc)
- Publish datasets to external services
  - (e.g., direct transfer from SciCat to Zenodo, useful if SciCat is only used internally without public interface)
- Prepare Jupyter environments for data-analysis on university hardware (containers)
  - (SciCat web-interface already offers Jupyter-button)
- Data conversion tools and strategies to migrate older/other data formats to NeXus

How we see the presented development efforts in Tübingen

- Just a “usable” prototype
  - To get a better idea about additional necessary features
  - No final decision on technical level (technology stack, architecture …)
    - We are prepared to redo things that have been worked on so far
    - Streamline development to be easily integrated into SciCat project for sustainable development
- Used in “production” in Frank Schreiber’s group
- Depending on how much momentum SciCat picks up within DAPHNE, we would be committed to transfer the project “from Tübingen to DAPHNE”

Current state (ongoing work)

- Online documentation
  - https://schreiber-lab.github.io/SciCat4daphne/
- Git repository
  - https://github.com/schreiber-lab/scicat4daphne
- Docker-compose file that aims to be “production-ready” (still some work to do)
- Demo virtual machine
  - http://scicat-demo.sytes.net (original SciCat interface)
  - http://scicat-demo.sytes.net/upload (ingestion frontend extension)
  - Credentials: User: admin Passwd: 2jf70TPNZsS (same as in SciCatLive)
  - Online until 15.06.2022

How could participation be organised?

- Collaborative software development within DAPHNE
  - First collaborative software dev. effort within DAPHNE that could serve as blueprint for other developments that will still follow in DAPHNE
  - Strengthen collaboration within DAPHNE on working level
  - Start to use central infrastructure within DAPHNE (e.g. hifis gitlab etc.)
  - Through involvement of different DAPHNE groups:
    Get a better idea about necessary features
- Participate in SciCat dev meetings through a representative
  - To communicate needs and results.

To start the discussion

- Where is the common ground in DAPHNE regarding the SciCat use-cases? E.g. :
  - will the installation rather be used to organise data internally or to provide public access?
  - how should the data enter into SciCat?
  - which data sources do you see for your local catalogue?
  - what will be the benefit to have a data catalogue in place, a.k.a. how will the data be used once it is available in the catalogue?
- Which features would be most interesting for you?
- Who would be interested in using a small-scale installation of a data catalogue?
- Who would potentially be interested in contributing to the development of additional SciCat features within DAPHNE?