Meeting Notes
1st of October 2024

- Release Jobs
  - Few issues were found that needs to be addressed
  - Despina implemented BE endpoint fullquery and fullfacet
  - Jobs interface in FE was changed to support changes

- Search UI
  - next PRs are not ready

- Documentation
  - SciCatLive:
    - asked collaborators to review documentation and share findings
    - Regina will contribute. She will get in touch with Laura
  - No updates for BE or FE

- Dataset schema and validation
  - more discussion is needed
  - What do we mean but schema validation:
    - structural validation: the structure and the values match the rules provided in the schema
    - semantic validation: we check the meaning of the entry
  - Example for semantic validation: checking that a measure of length is using a unit that is associated with length, aka meters.
  - Linus, Dylan and Spencer will explore tools and options and will provide feedback
  - In our case, it will be easier to implement structure validation first
  - three possible settings / use cases:
    - disabled: BE accepts the scientific metadata without checking and create the dataset
    - warning: BE accepts the scientific metadata and will warn if they do not validate, nevertheless will create the dataset
    - strict: BE validates the scientific metadata and will create the dataset only if they are valid
  - where is the full schema saved?
  - where the schema reference is saved: in the dataset?
  - Allow enough flexibility to be able to assign different and multiple schemas for each dataset

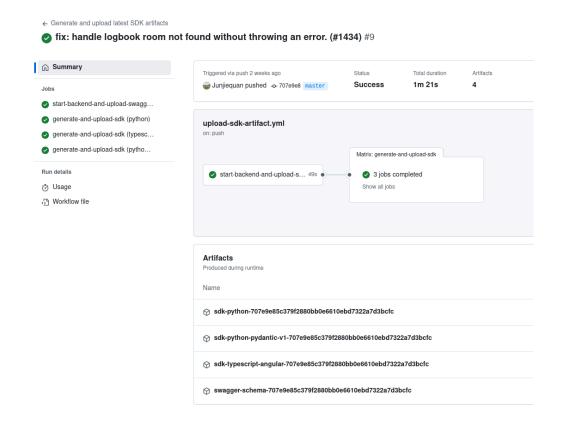- who controls which schema is applied to the dataset: user, ingestor or scicat?

- Data display
    - the project cannot include in core the visualization code
    - how to integrate external tool to visualize data?
    - micro frontends do not seems the right solution
    - connect the dataset visualization with external services implementing the specialized visualization tailored to the data and the science field
    - possible technology: streamlit (https://streamlit.io/)
    - data examples:
        - Crystallography
        - Tomography
        - others?

- SDK
    - packages are created as artifacts on every PR merge to Master Example:
    https://github.com/SciCatProject/scicat-backend-next/actions/runs/11174676786

- At every release, new sdk packages are created and posted on the public repository related to the language.

Deployments 16

✅ **python-pydantic-v1-sdk-package**
6 hours ago

✅ **python-sdk-package** 6 hours ago

✅ **npm-sdk-package** 6 hours ago

+ 13 deployments

- links to official repositories
  - https://www.npmjs.com/package/@scicatproject/scicat-sdk-ts
  - https://pypi.org/project/scicat-sdk-py/
  - https://pypi.org/project/scicat-sdk-pydantic/

SDK roadmap proposal

Following the discussion during the last collaboration meeting, I propose the following roadmap forward regarding the SDK:

- *Finalize auto generation of javascript and python SDKs CI/CD.* **COMPLETED**
- *Create SDK and openapi file as artifacts at every merge to master.*
  COMPLETED
  Artifacts should be named according to the following shema:
  > **sdk-<language>[-<additional-options>]-<git-commit-id>.zip**
- Publish them automatically on the language specific platforms with the following names:
  - javascript: scicat-sdk-js
  - python: scicat-sdk-py
  - python-pydantic: scicat-sdk-pydantic
  COMPLETED
- Publishing platforms are:
  - javascript: npm-js
  - python: pypi
  COMPLETED
- Clearly state that SDKs are provided as they are, with no guarantee.
  No testing is done on SDKs during CI/CD.
- Include new languages as needed
- refactor pySciCat to extend scicat-sdk-py, with backward compatibility, additional functionalities and validation and include testing like the ones included in discarded PR.
- Proposed pySciCat as the low level SciCat python library, providing the following:
  - additional functionalities,
  - extensive testing
  - backward compatibility
  - different release cycle from the SDK due to human revisions
- After 6/12 months evaluate if having scicat-sdk-py and pyScicat is sustainable and the most sensible solution.
  We will adapt the road map forward based on the outcome.

This roadmap is recorded in the following github issues:
https://github.com/SciCatProject/scicat-backend-next/issues/1403

Proposed Agenda
17th of September 2024

- Current releases updates
  - Jobs
  - Search UI
    - Metadata key and type
      https://github.com/SciCatProject/scicat-backend-next/pull/1428
  - Documentation
    - where should we save the documentation specific for BE and FE?
    - which folder?
  - SDK
  - Branding
    - Tagline: https://github.com/SciCatProject/scicatproject.github.io/issues/55
    - Mission statement
      https://github.com/SciCatProject/scicatproject.github.io/issues/56
  - other efforts
- Discussions
  - Project leadership and governance
    - https://opensource.guide/leadership-and-governance/
    - http://oss-watch.ac.uk/resources/benevolentdictatorgovernancemodel
    - https://www.linuxfoundation.org/resources/open-source-guides/building-leadership-in-an-open-source-community
  - Data visualization
    - which use case to address first?
    - how to integrate external tools?
  - Dataset Schema Validation
    - Semantic vs structural validation
    - Input vs output schema
- PR and issues
- Additional topics

Open discussions for future features

currently not scheduled on any release

- Proposed types for Scientific Metadata, Issue #984
  - https://github.com/SciCatProject/scicat-backend-next/issues/984

- Metadata entry schema
  - The overall metadata schema is free for the user to design their own structure
  - SciCat will impose a schema at the entry level to be able to interpret and visualize the values reducing uncertainty and wrong interpretation
  - https://github.com/SciCatProject/scicat-backend-next/issues/939#issuecomment-1874033672
  -
- Dataset hierarchy
  - Proposed solution:
    - unify the dataset schema
    - add dataset type: collection (specialized type of derived dataset)
    - add relationships and leverage them to build a hierarchy
  - https://github.com/SciCatProject/scicat-backend-next/issues/805

- Independent endpoint for datasets scientific metadata management
  - https://github.com/SciCatProject/scicat-backend-next/issues/954
  - As we use SciCat as our platform to curate our dataset, it is becoming more important to be able to add, update or delete individual scientific metadata entries
  - High priority

- Dataset history and update history
  - should we create a stand alone collection for update history?
  - should we track any changes made to datasets?
  - should we track changes made to other objects?

SciCat Collaborators Zoom meeting details

Topic: SciCat Collaborators Meeting
Time: Feb 6, 2024 04:00 PM Copenhagen
    Every 2 weeks on Tue, until Dec 24, 2024, 24 occurrence(s)

Join Zoom Meeting
https://ess-eu.zoom.us/j/61405959531?pwd=MllyYm12RS9RNEdGTzR3bXRvNGp5Zz09

***Meeting ID:* 614 0595 9531**
***Passcode:* 507998**

One tap mobile
+46850163827,,61405959531#,,,,*507998# Sweden

Dial by your location
• +46 8 5016 3827 Sweden
Find your local number: https://ess-eu.zoom.us/u/cbwrVAG268

Join by SIP
61405959531@109.105.112.236
• 61405959531@109.105.112.235

Join by H.323
• 109.105.112.236
• 109.105.112.235