Meeting Notes

11th of December 2024

Merry Christmas and Happy New Year

- Christmas break
  - Meeting scheduled on December 24th, 2024 is canceled due to Christmas break
  - Next meeting is scheduled for January 7th, 2025
- SciCatCon 2025
  - ESS final approval came through
  - It will be held on-site at the ESS DMSC located in Lyngby, located north of Copenhagen city center.
  - Doodle poll is available to specify availability https://doodle.com/meeting/participate/id/ej5QkPvd
  - With the new year, ESS will start the organization process, including creating the web presence.
- Branding
  - Laura presented the latest updates regarding SciCat identity.
  - Logo is confirmed to be this one:

        - github issue: https://github.com/SciCatProject/scicatproject.github.io/issues/54
        - new proposed logo in github issue: https://github.com/SciCatProject/scicatproject.github.io/issues/54#issuecomment-2500964064
    - Tagline.

      ## Advance your data to advance your science

        - Updated tagline based on feedback received
        - github issue: https://github.com/SciCatProject/scicatproject.github.io/issues/55
        - Selected tagline in github issue: https://github.com/SciCatProject/scicatproject.github.io/issues/55#issuecomment-2531473335

    - Project mission:

      Advance your data to advance your science. SciCat is a community driven scientific metadata catalog designed to simplify metadata management, enabling data sharing, discovery, and collaboration. By providing a central hub for metadata, SciCat makes data easily findable and accessible for the entire community. With flexible integration into diverse infrastructures and a clear path to data publication, SciCat empowers organizations to foster an open and innovative data culture, driving high-quality science forward.

        - github issue: https://github.com/SciCatProject/scicatproject.github.io/issues/56
        - selected project mission in github issue: https://github.com/SciCatProject/scicatproject.github.io/issues/56#issuecomment-2531467928

- Release jobs
    - On-going internal refactor to improve code management and maintenance
    - There is a PR open to review
    - We are in the final push to release job to master
    - Regina asked about documentation
    - PR #56 in documentation is open with the jobs updates https://github.com/SciCatProject/documentation/pull/56
    - README file has been added in code with info for developers

    - Regina asked if MaxIV is using jobs.
      MaxIV is using BE v4 and is not using jobs at the moment.
    - RFI is currently rabbitmq with jobs in BE v3, but they are intend to migrate to kafka when they switch to BE v4.0

- Dataset subsystem V4
    - Max introduced the work that Martin is doing on Datasets under the DAPHNE contribution to SciCat
    - Martin presented the work in details
    - Dataset V4 aims to:
        - uses only Unified dataset schema
        - simplify the code
        - offer functionalities related only to datasets
        - leverage mongodb aggregation functionality
    - Dataset v4 will co-exists with Dataset v3
    - Spencer asked if any endpoints can be offered on multiple API versions.
      According to the documentation available at the following url:
      https://docs.nestjs.com/techniques/versioning
      The same code and route can be offered under multiple API versions:
        - api/v3/proposals -> GET proposals
        - api/v4/proposals -> GET proposals

- SDK
    - Emil asked the reason to create the python and python-pydantic packages, as the python package leverages pydantic v2, while python-pydantic uses pydantic v1.
    - Given that the python SDK are relatively new and they are not widely used yet, It was decided to remove python-pydantic with pydantic v1 and keep the python package based on pydantic v2.

- Carlo introduced new PSI collaborators: Fredric and Omkar

- Bjorn reported on PR.
    - Most of the outstanding PR from dependabot have been resolved and merged.
    - FE PR #1626 to update angular to v18 is still outstanding. He will test if it can be done without to much effort

- Performance
  Igor presented quickly the results of the performance tests that he performed.

- Discussion about when to run such tests?
- Laura suggested to run them only on major releases
- Jay proposed to run on each release and, also, continuously on the individual instances
- Igor would like to establish a baseline in a controlled environment and then promote continuous monitoring
- Max is concerned about forcing a specific monitoring tool, which might make it hard to adopt scicat specifically if an IT infrastructure is already present
- Can we de-couple monitoring tools from monitoring protocols, so we can use libraries tight to the protocol?
- Do well defined and well adopted monitoring protocols exists? Can such protocols be consumed by different monitoring tools?
- We will need to be conscious of the hardware were the tests are run on, as different hardware might influence tests' results.
- Bjorn suggested to look in open telemetry protocol, which can de-couple the scicat from the monitoring tools. https://opentelemetry.io/
- There is a library to integrate nestjs application with open telemetry systems: https://github.com/open-telemetry/opentelemetry-js-contrib/tree/main/plugins/node/opentelemetry-instrumentation-nestjs-core#readme

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:
  - typescript with angular: ***scicat-sdk-ts***
  - python: ***scicat_sdk_py*** (with pydantic v2)
  - python-pydantic: ***scicat_sdk_pydantic*** (with pydantic v1), obsolete and removed

  COMPLETED
- Publishing platforms are:
  - javascript: npm-js
  - python: pypi

  COMPLETED
- Renamed python packages to address naming issues: ***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.
  Planned:
  - Typescript with fetch: ***scicat-sdk-ts-fetch***
- 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.
- Refactor FE to use new SDK: ***COMPLETED. bug fixing***

This roadmap is recorded in the following github issues:

https://github.com/SciCatProject/scicat-backend-next/issues/1403

Proposed Agenda

7th of January 2025

- Current updates
  - Jobs
  - SDK
  - Datasets v4

- Performance test discussion

- SciCatCon 2025
  - Poll: https://doodle.com/meeting/participate/id/ej5QkPvd

- 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: Jan 7, 2025 04:00 PM Copenhagen
Every 2 weeks on Tue, until Dec 23, 2025, 26 occurrence(s)
Please download and import the following iCalendar (.ics) files to your calendar system.
Weekly:
https://ess-eu.zoom.us/meeting/u5Upd-yhqzIvGdU6_6kfQjr_Gr3-A_FaoUcT/ics?icsToken=98tyKu-pqjsvE9GctRqGR_McBoigZ_PwmFhYjadoyiCzCCZiaBHJEfARf5NANtnf

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

Meeting ID: 614 0595 9531
Passcode: 507998

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

Meeting ID: 614 0595 9531
Passcode: 507998