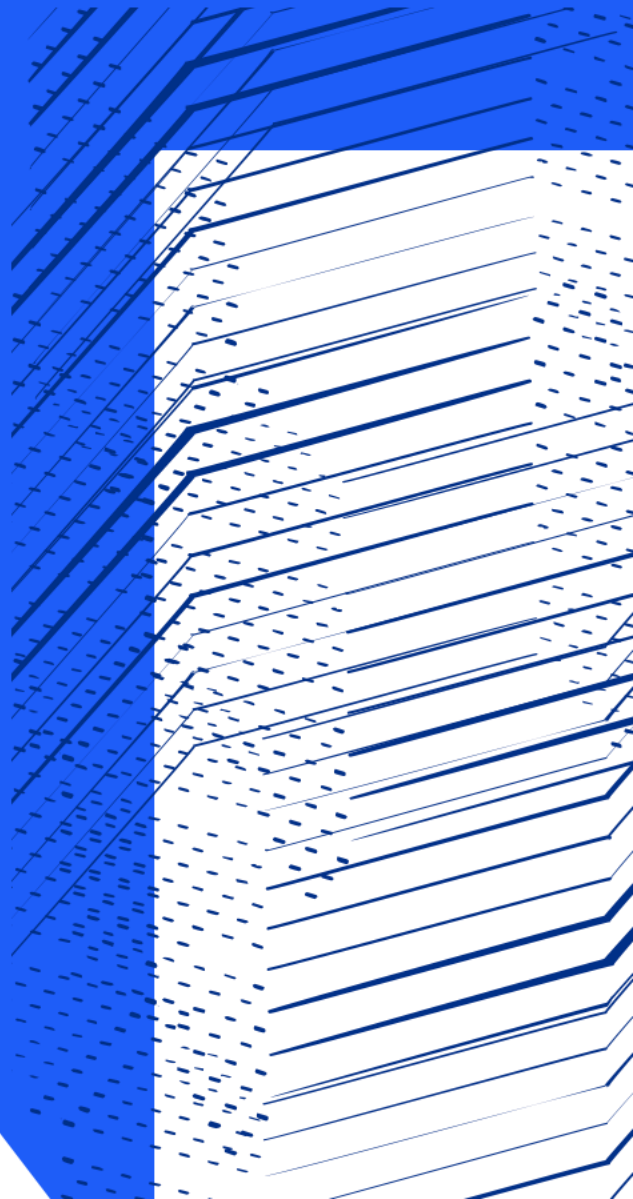
Science and Technology Facilities Council

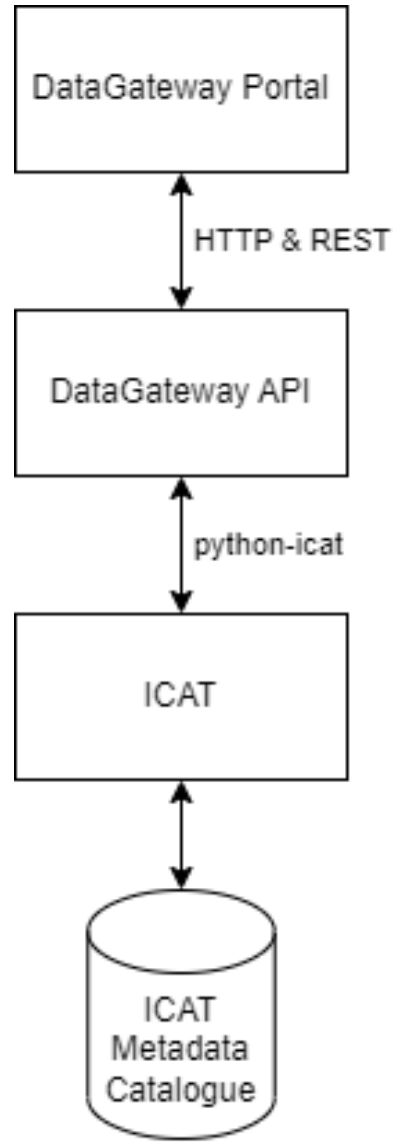
DataGateway API & Search API

Viktor Bozhinov



DataGateway API

- Flask-based Python REST API application
- Fetches data from ICAT
  - Using python-icat (ICAT backend) or directly from an ICAT database (DB backend)
- Interfaces with DataGateway Portal
  - Provides data discovery and access functionality for large-scale science facilities
  - Allows for experimental data to be browsed, explored, and visualized
  - Currently used at Diamond Light Source and ISIS
- Its own endpoint specifications and schema relationships
- Requires Authentication



Science and Technology Facilities Council

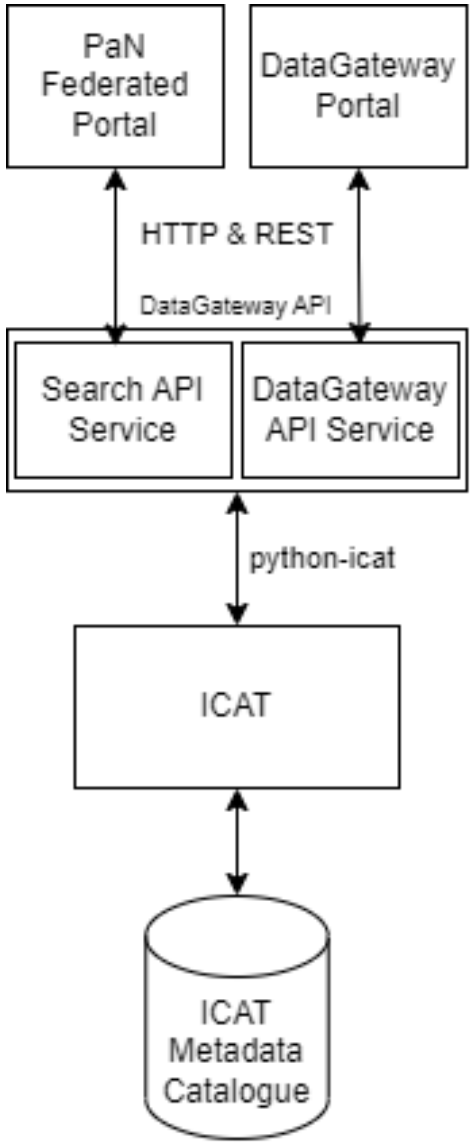
Search API

- Part of the ExPaNDS and PaNOSC European projects
- Search API allows:
  - Public data from ICAT catalogues to be found
  - ICAT facilities to connect their data repositories with the European Open Science Cloud (EOSC)
- EOSC:
  - Federated portal allowing researchers to publish, find and use research data
  - A big community of researchers

Science and Technology Facilities Council

Search API

- Implemented as part of the DataGateway API codebase
- python-icat used for interfacing with ICAT
- Common DataGateway API logic reused
- Three endpoints for searching public documents, datasets, and instruments
  - Support Loopback REST query filters
  - Return data in the format specified
  - Do not require authentication
- Configurable Search API to ICAT entity mappings to allow facilities with differently configured ICATs to run their own instance of the API



Science and Technology Facilities Council

DataGateway API & Search API

- Codebase now consists of two APIs
- Application modes were added to allow running both APIs as a single instance or as separate instances
  - Modes can be enabled in the configuration file
- Two sets of endpoints served on different URL paths
- Separate Swagger Docs

Recent Work

- JSON configuration file was replaced with YAML file
- New DataGateway API endpoints were created for the new ICAT 5 entities
- Generator script was updated to generate test data for the new entities and insert data into the database using the ICAT backend
- More tests were added to increase the code coverage for DB backend and new ICAT 5 entities
- Support for scoring Search API results was added
  - Sadly, the PaNOSC Scoring component does not work with a large dataset
- The application was Dockerised and can now run in a container

Future Work

- **Update Postman Collection**
  #411 opened last month by MRichards99 · 6 tasks
- **Database returns timezone indicators but does not accept them** bug db backend
  #392 opened on Nov 22, 2022 by Reillyhewitson
- **Pydantic 1.10.2 does not support python 3.6** 1
  #384 opened on Oct 5, 2022 by Reillyhewitson · 3 tasks
- **Get API working with Poetry 1.2** 2
  #379 opened on Sep 15, 2022 by Reillyhewitson · 1 task
- **Investigate Composite Actions for CI** refactor
  #371 opened on Sep 1, 2022 by MRichards99 · 3 tasks
- **Configure Release workflow to run only if CI passes**
  #361 opened on May 16, 2022 by VKTB · 1 task
- **Update DataGateway API OpenAPI/Swagger Docs to ICAT Backend** db backend icat backend
  #347 opened on Feb 28, 2022 by MRichards99 · 2 tasks
- **Tests for ICAT 5 Mappings** expands-search-api 1
  #310 opened on Jan 31, 2022 by MRichards99 · 1 task

Science and Technology Facilities Council