Latest news from SciCat ecosystem:

- ElasticSearch integration
- Job sub-system

SciCat

Max Novelli
max.novelli@ess.eu
SciCat Project Leader

Senior Data Curation Scientist
European Spallation Source, DMSC
Copenhagen, Denmark

NoBUGS 2024, ESRF, France, 2024/09/26

SciCat Project

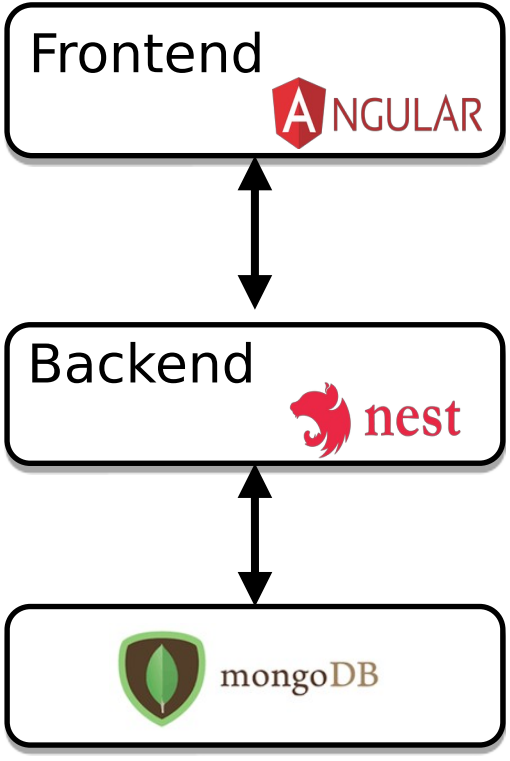
Open Source Scientific Metadata Catalog

Project

- International collaboration
- Flexible and Extensible
- Easy integration
- Young and Dynamic
- Cool technologies

Community

- Committed and Diverse
- Active and Responsive
- Supporting and Accepting



https://scicatproject.github.io/

https://github.com/scicatproject

https://scicatproject.github.io/documentation

European Spallation Source

Paul Scherrer Institut

MAX IV

Rosalind Franklin Institute

Heinz Maier-Leibnitz Zentrum

Advanced Light Source

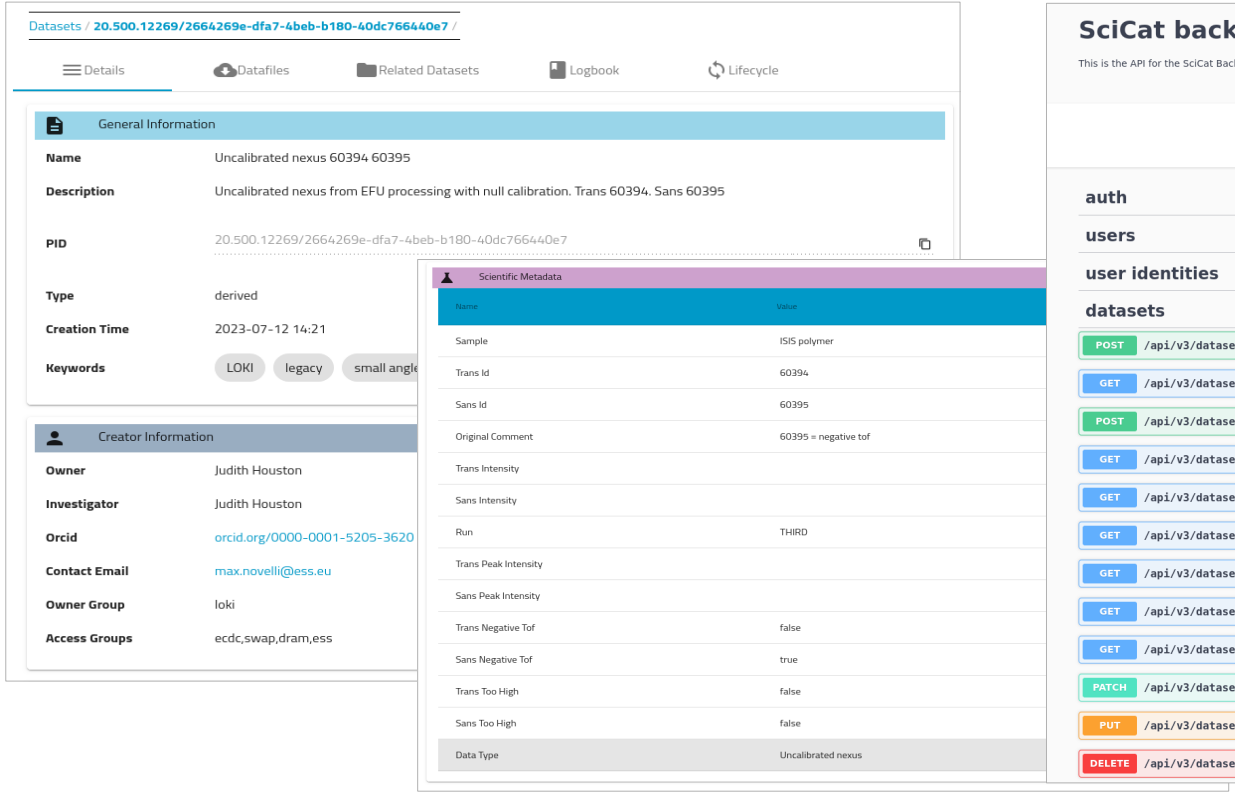
Deutsches Elektronen-Synchrotron DESY

Synchrotron SOLEIL

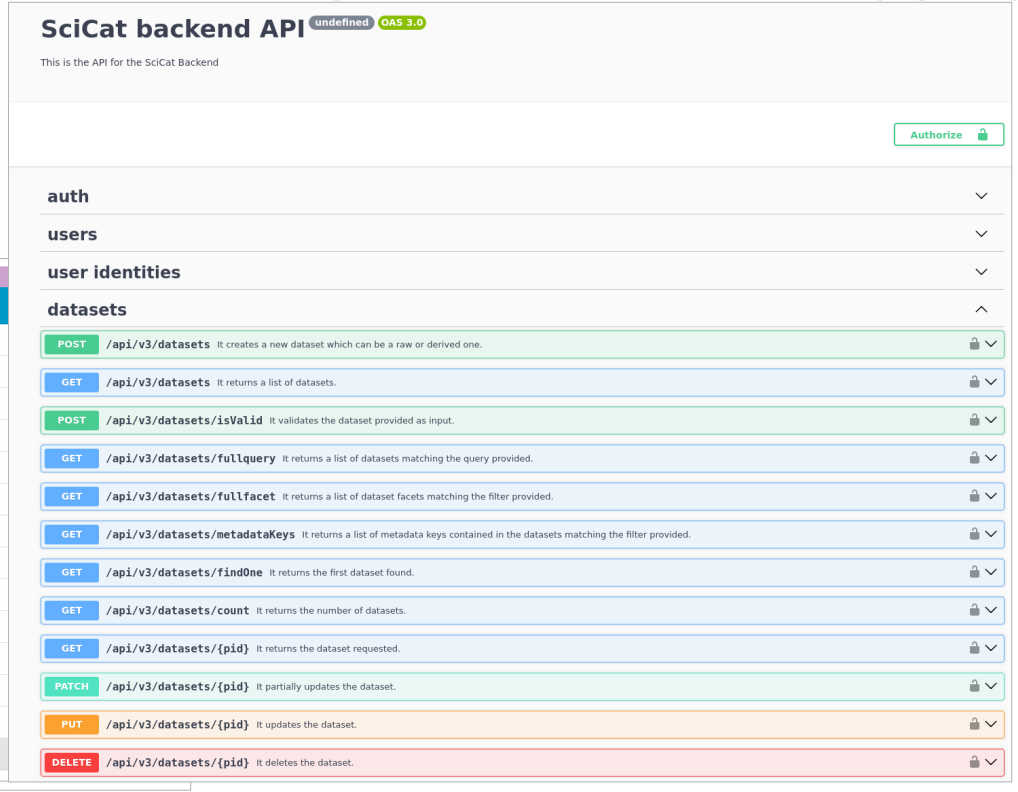
DAta from PHoton and Neutron Experiments

SciCat Interfaces

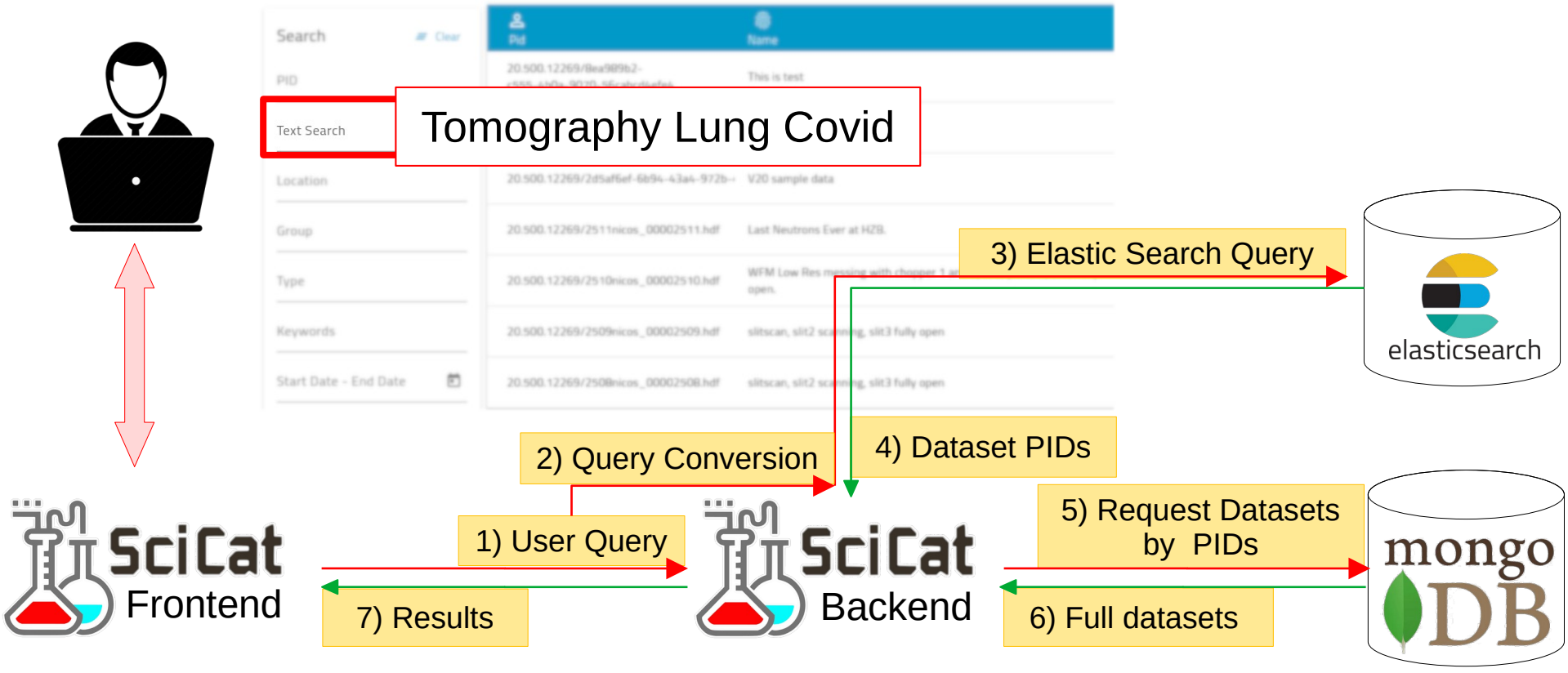
Frontend



Backend (REST API, Swagger)



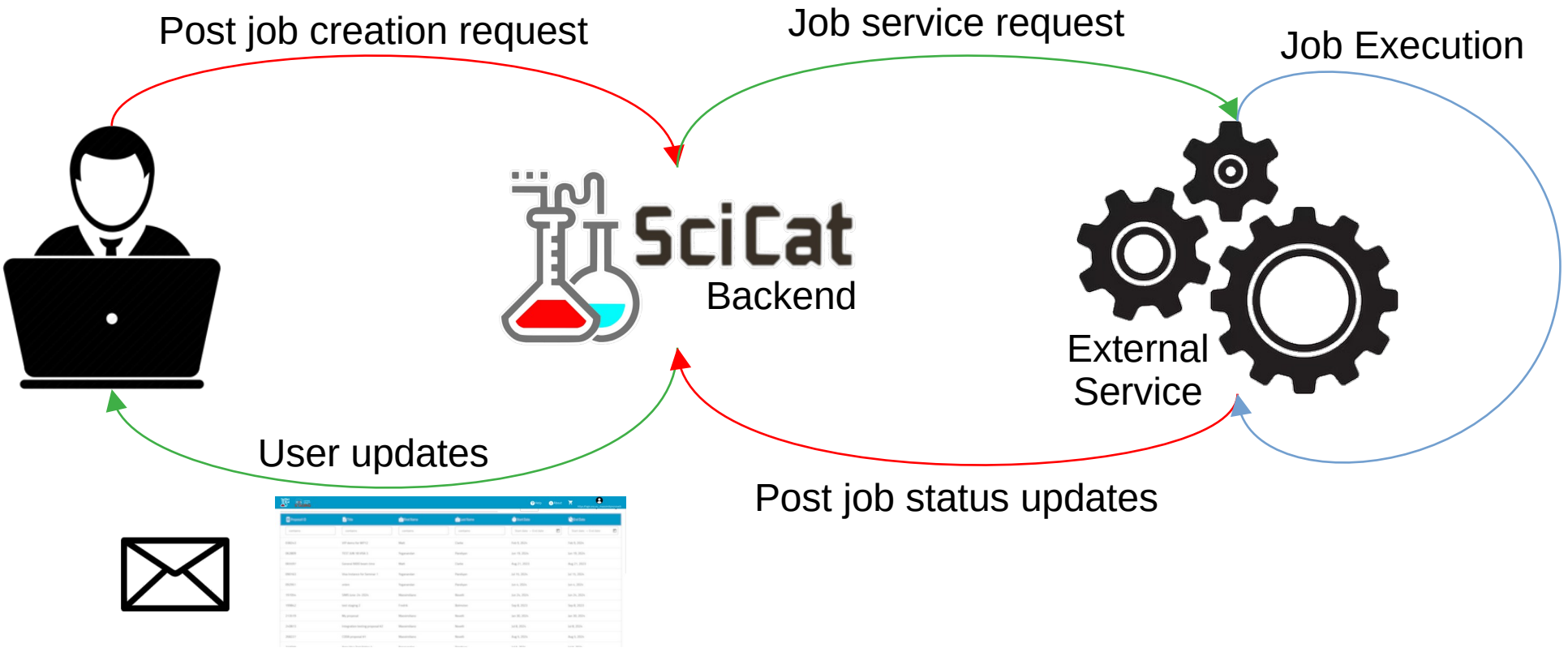
Elastic Search Integration



Elastic Search: benefits

- Elasticsearch is optional.
  It can be enabled in configuration.
- Provides advanced search
  through a dedicated endpoint
- Transparent data syncronization
  between MongoDB (main database) and ElasticSearch.
- Enhance search functionalities
  by leveraging ElasticSearch Tokenizers and Analyzers
- Results ranking
  with relevancy score provided by ElasticSearch

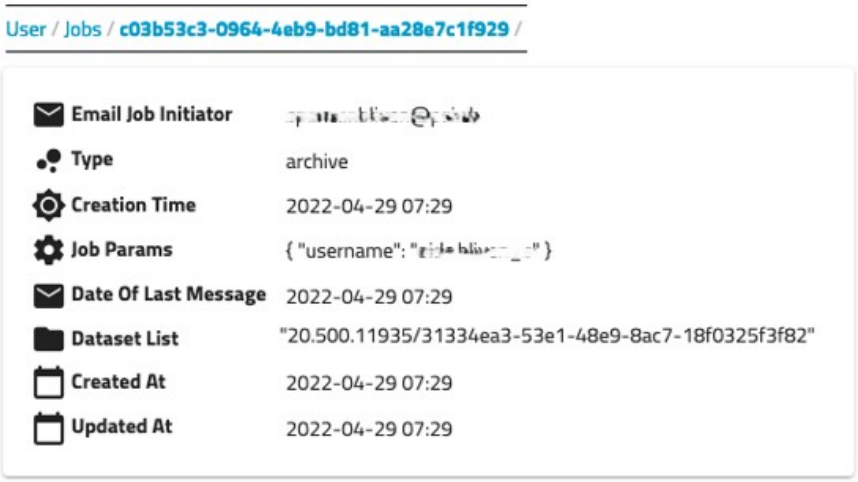
Jobs


Post job creation request
Job service request
Job Execution
SciCat
Backend
External
Service
User updates
Post job status updates

Moving forward

Version 3

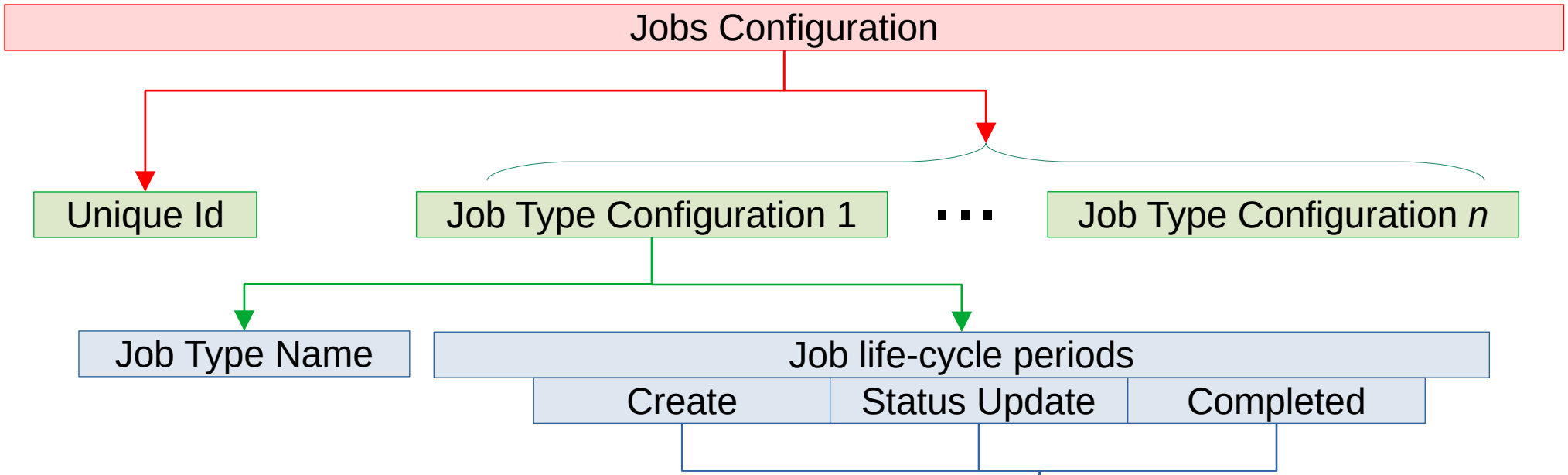
- Fixed job types
- Hard-coded job actions
- Customization requires rebuilding
- Code not following best practices



Version 4

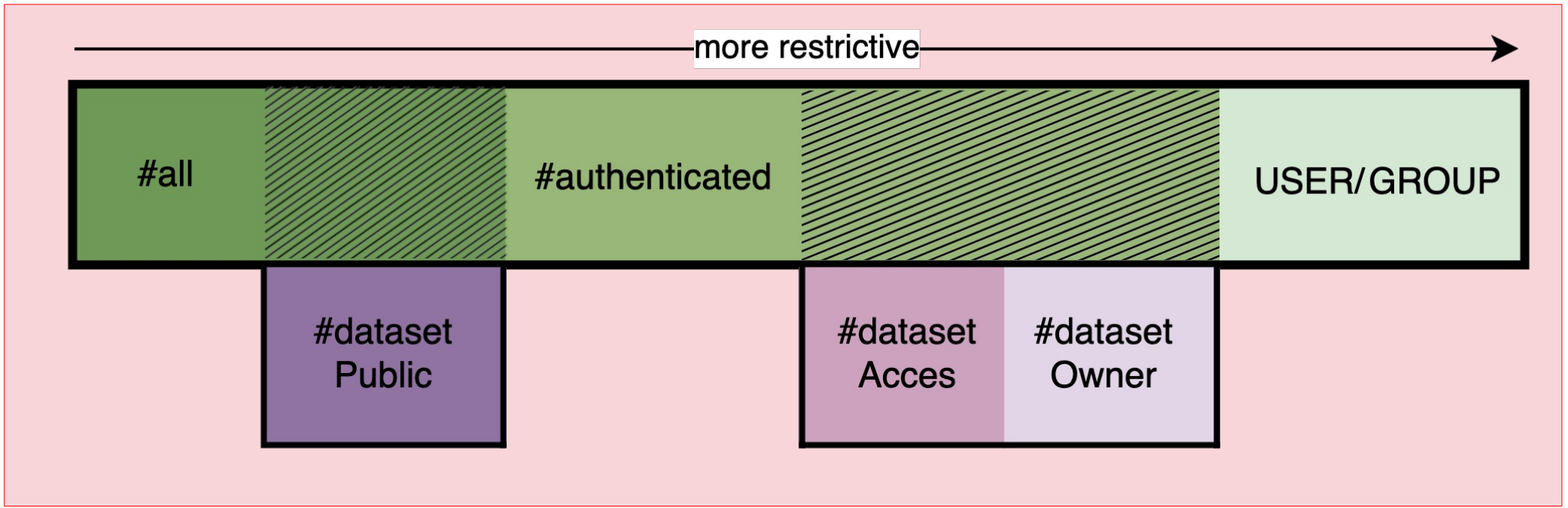
- Configure jobs without code changes
- Code refactoring: Modularity
- Use of official SciCat version containers
- Better testing
- Better code coverage
- Develop plugin system to foster contributions and new functionalities *(Future effort)*

Jobs Configuration




```
{
  "actionType": "email",
  "to": "{{job.contactEmail}}",
  "from": "dacatmail@psi.ch",
  "subject": "[SciCat] Your {{job.type}} job submit
    successfully",
  "bodyTemplate": "job-template.html"
}
```

Jobs Permissions



https://scicatproject.github.io/documentation/Development/v4.x/backend/authorization/authorization_jobs.html

Status

Elastic Search Integration

- Released Minimum Viable Product
- Automated API tests
- Currently in production at ESS
- Collecting feedback from internal users
- Allocating resources to:
  - Improve integration configuration
  - Improve ElasticSearch configuration
  - Implement extensive configuration

Jobs

- Unit tests
  - every user type
  - job configuration
  - dataset permissions
- API tests covering authorization (~200)
- tests are integrated in CI
- Integration tests with scicatlive (ongoing)
- Currently been tested at PSI

Thank you

Acknowledgments

Elastic Search Integration

- Junjie Quan (ESS)

Jobs Release

- Spencer Bliven (PSI)
- Sofya Laskina (EPFL)
- Despina Adamopoulou (EMPA)

Massimiliano (Max) Novelli

max.novelli@ess.eu

EUROPEAN SPALLATION SOURCE

Thanks also to:

- Carlo Minotti
- Dylan McReynolds
- Igor Khokhriakov
- Regina Hinzmann
- Laura Shemilt
- Linus Pithan