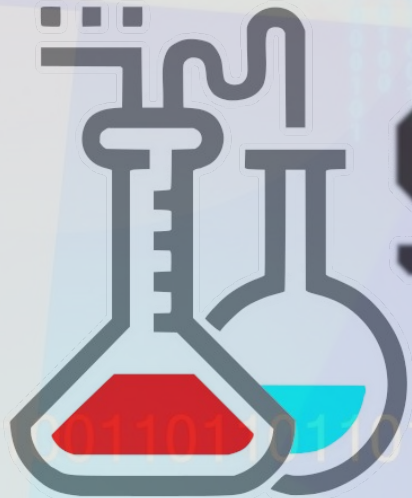


Latest news from SciCat ecosystem:

- 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

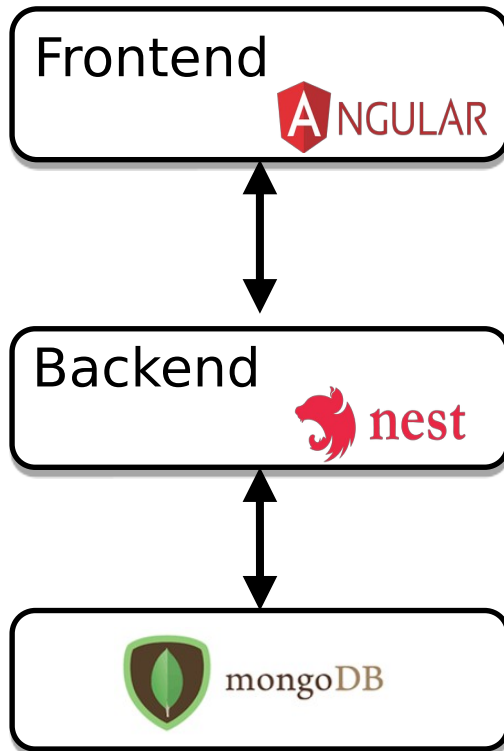
Open Source Scientific Metadata Catalog

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>





SciCat Interfaces

Frontend

Datasets / 20.500.12269/2664269e-dfa7-4beb-b180-40dc766440e7 /

Details Datafiles Related Datasets Logbook Lifecycle

General Information

Name Uncalibrated nexus 60394 60395

Description Uncalibrated nexus from EFU processing with null calibration. Trans 60394. Sans 60395

PID 20.500.12269/2664269e-dfa7-4beb-b180-40dc766440e7

Type derived

Creation Time 2023-07-12 14:21

Keywords LOKI legacy small angle

Creator Information

Owner Judith Houston

Investigator Judith Houston

Orcid orcid.org/0000-0001-5205-3620

Contact Email max.novelli@ess.eu

Owner Group loki

Access Groups ecdc,swap,dram,ess

Scientific Metadata

Name	Value
Sample	ISIS polymer
Trans Id	60394
Sans Id	60395
Original Comment	60395 = negative tof
Trans Intensity	
Sans Intensity	
Run	THIRD
Trans Peak Intensity	
Sans Peak Intensity	
Trans Negative Tof	false
Sans Negative Tof	true
Trans Too High	false
Sans Too High	false
Data Type	Uncalibrated nexus

Backend (REST API, Swagger)

SciCat backend API undefined OAS 3.0

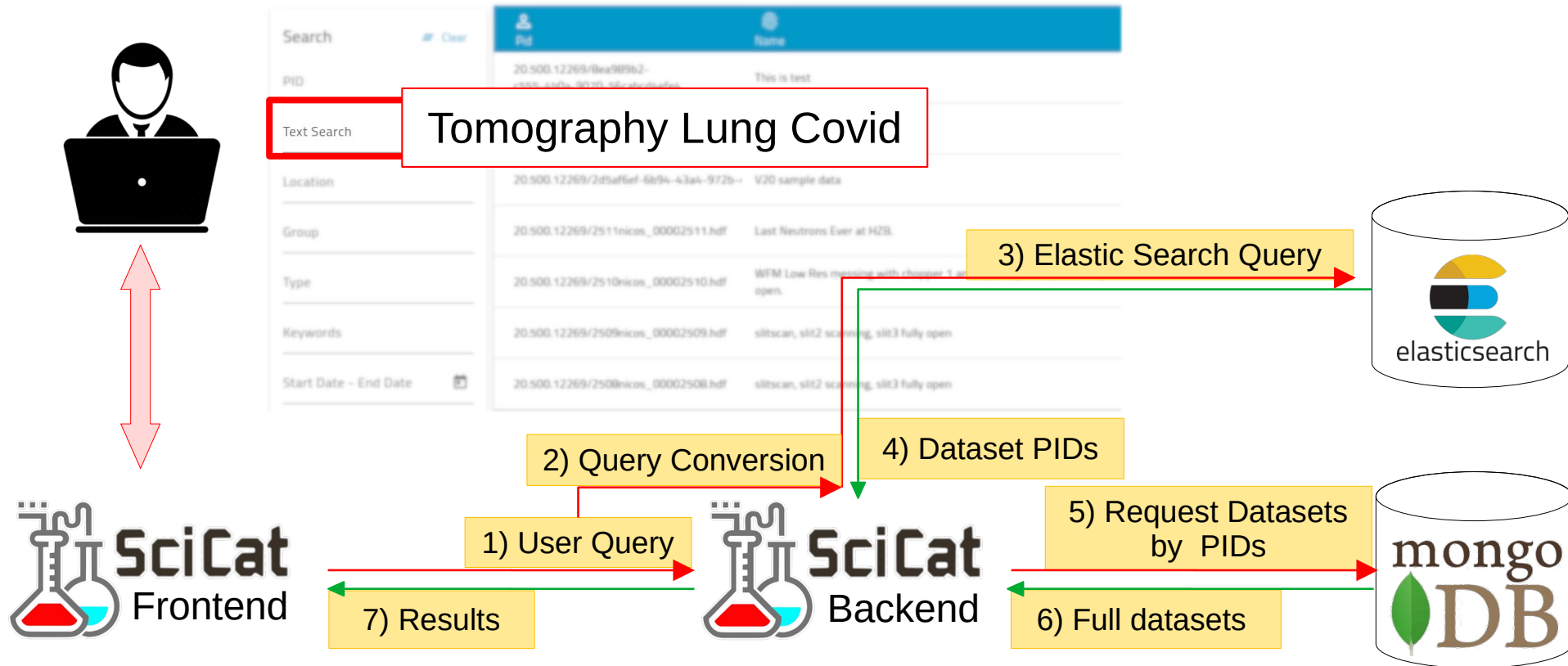
This is the API for the SciCat Backend

[Authorize](#)

- auth**
- users**
- user identities**
- datasets**
 - POST** `/api/v3/datasets` It creates a new dataset which can be a raw or derived one.
 - GET** `/api/v3/datasets` It returns a list of datasets.
 - POST** `/api/v3/datasets/isValid` It validates the dataset provided as input.
 - GET** `/api/v3/datasets/fullquery` It returns a list of datasets matching the query provided.
 - GET** `/api/v3/datasets/fullfacet` It returns a list of dataset facets matching the filter provided.
 - GET** `/api/v3/datasets/metadataKeys` It returns a list of metadata keys contained in the datasets matching the filter provided.
 - GET** `/api/v3/datasets/findOne` It returns the first dataset found.
 - GET** `/api/v3/datasets/count` It returns the number of datasets.
 - GET** `/api/v3/datasets/{pid}` It returns the dataset requested.
 - PATCH** `/api/v3/datasets/{pid}` It partially updates the dataset.
 - PUT** `/api/v3/datasets/{pid}` It updates the dataset.
 - DELETE** `/api/v3/datasets/{pid}` It deletes the dataset.



Elastic Search Integration





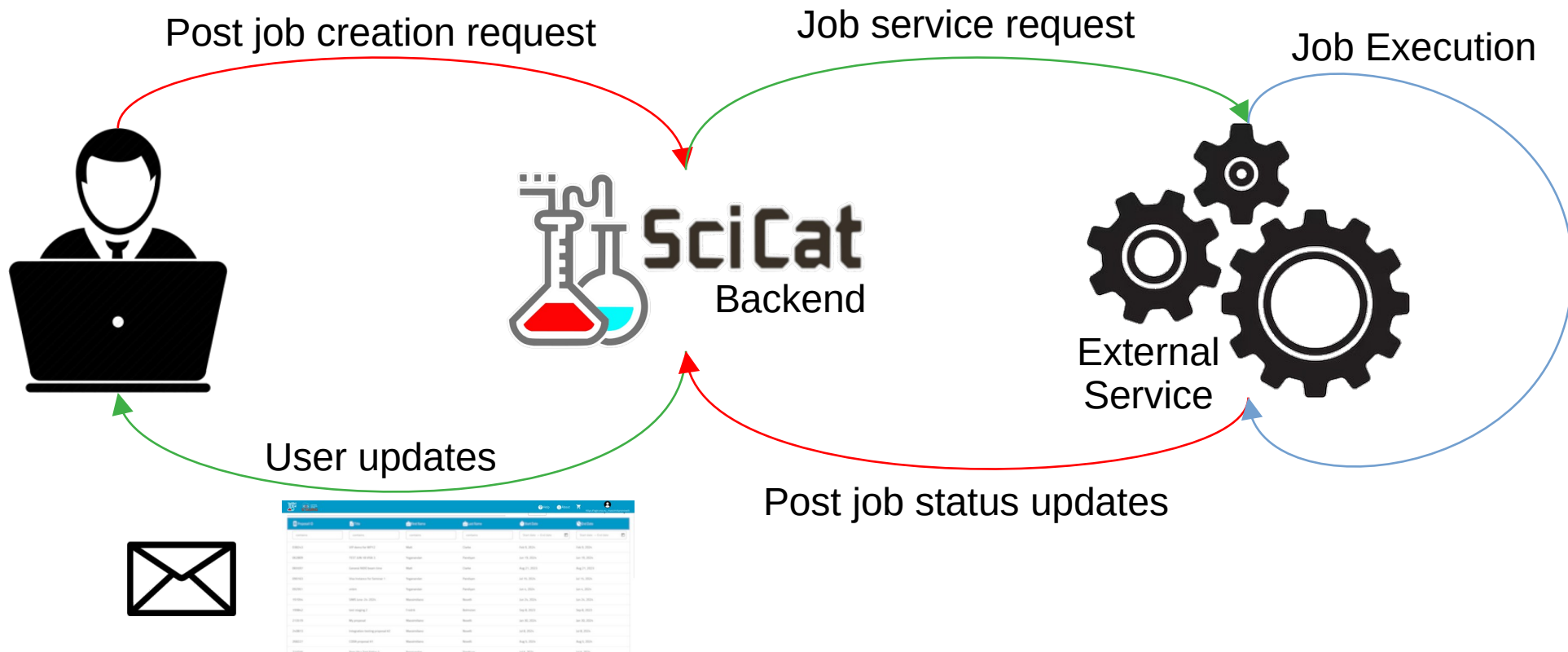
Elastic Search: benefits



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



Jobs





Moving forward

Version 3

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

User / Jobs / c03b53c3-0964-4eb9-bd81-aa28e7c1f929 /

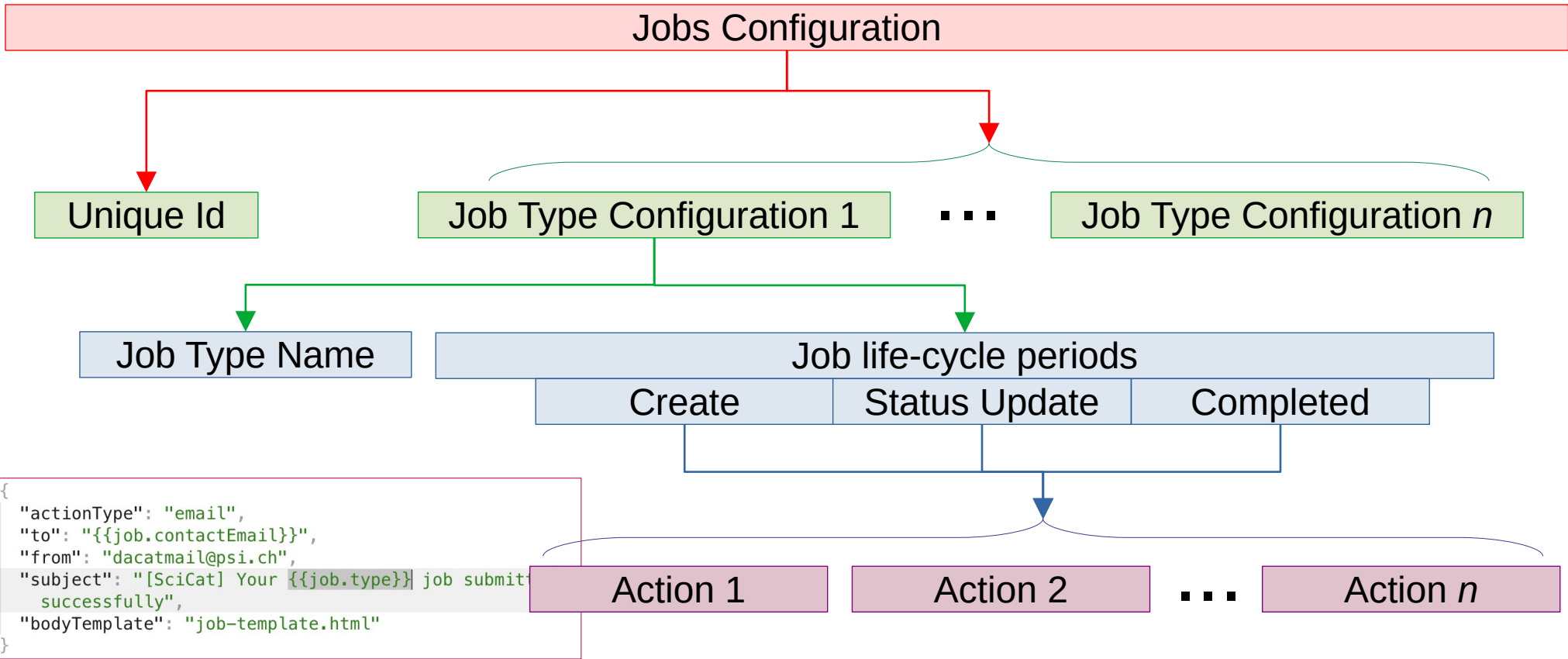
✉ Email Job Initiator	op@scicat@scicat
🔗 Type	archive
⚙️ Creation Time	2022-04-29 07:29
⚙️ Job Params	{ "username": "zika@scicat" }
✉ Date Of Last Message	2022-04-29 07:29
📁 Dataset List	"20.500.11935/31334ea3-53e1-48e9-8ac7-18f0325f3f82"
📅 Created At	2022-04-29 07:29
📅 Updated At	2022-04-29 07:29

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)



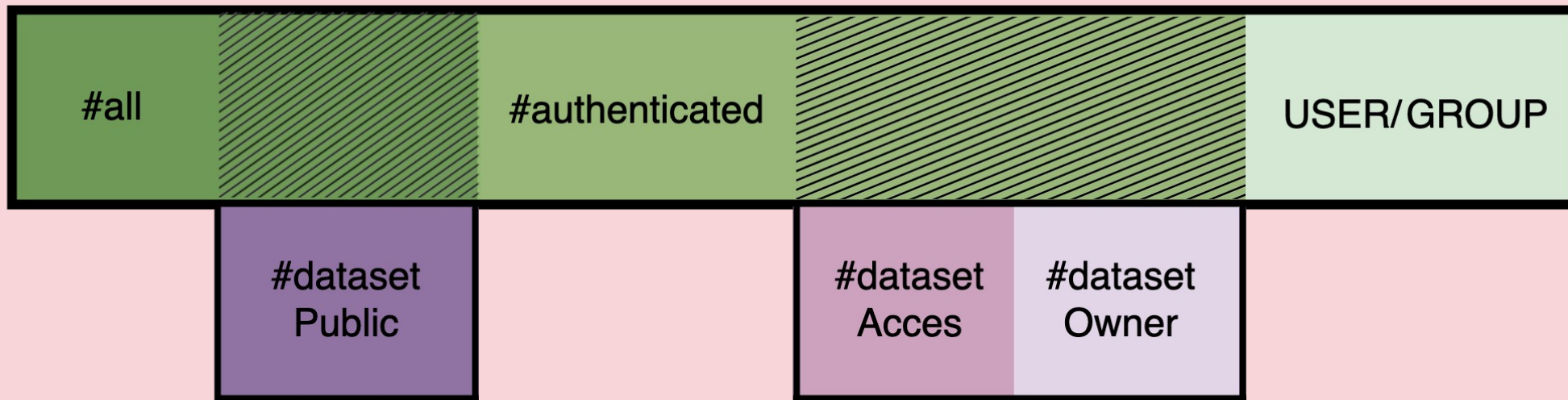
Jobs Configuration





Jobs Permissions

more restrictive →





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



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



Thanks also to:

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