PSI Center for Scientific Computing, Theory and Data

PSI@SciCatCON24

Carlo Minotti, Spencer Bliven, Leonardo Sala
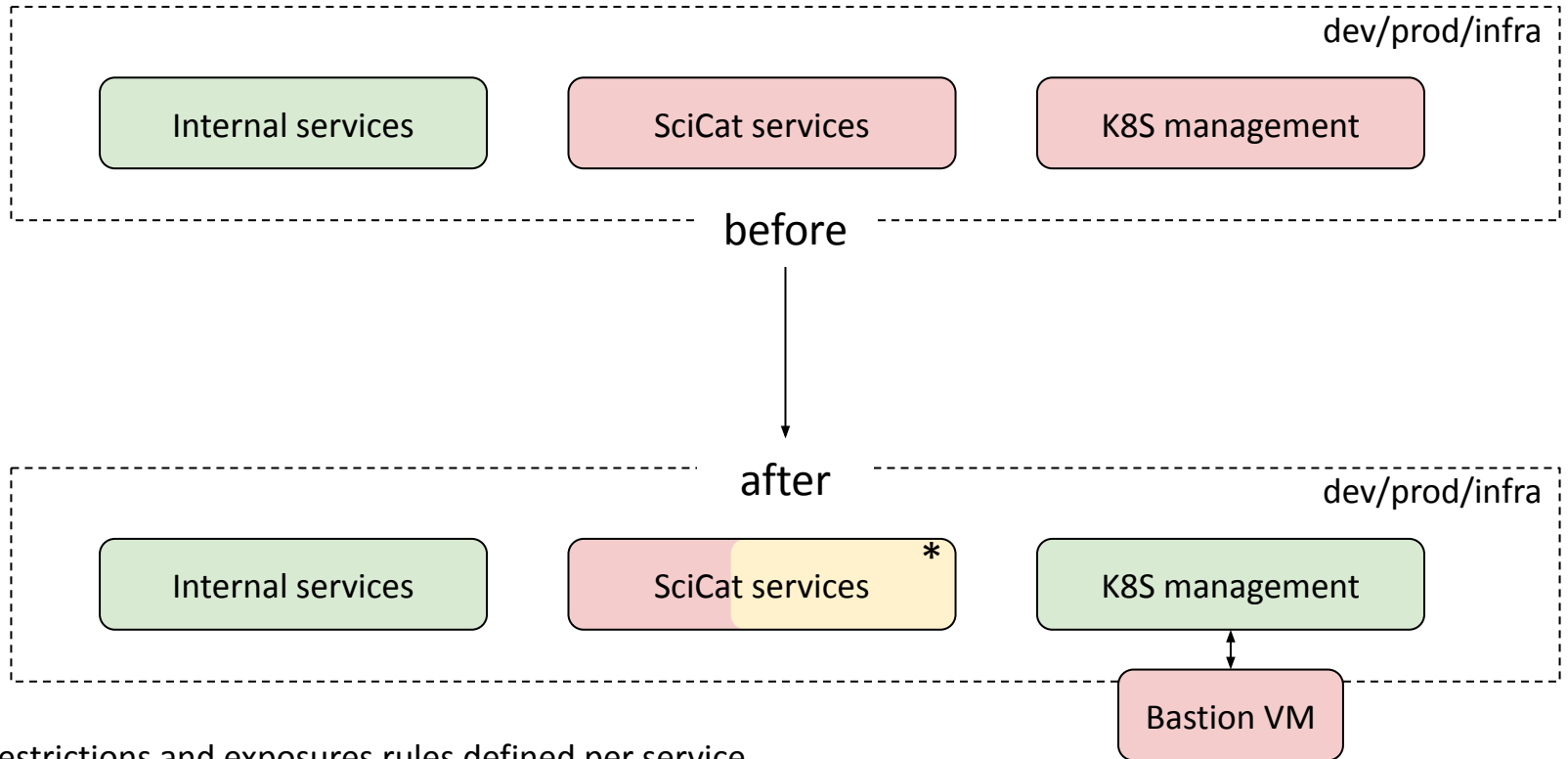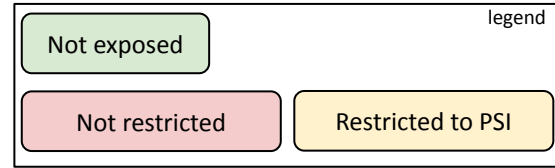PSI, 02/07/24

Agenda

PSI

1   Infrastructure updates

2   User-facing updates

3   Grants

4   Future plans

Infrastructure 1/2

Security review

legend

| Not exposed |
| Not restricted | Restricted to PSI |

dev/prod/infra

| Internal services | SciCat services | K8S management |

before

after

dev/prod/infra

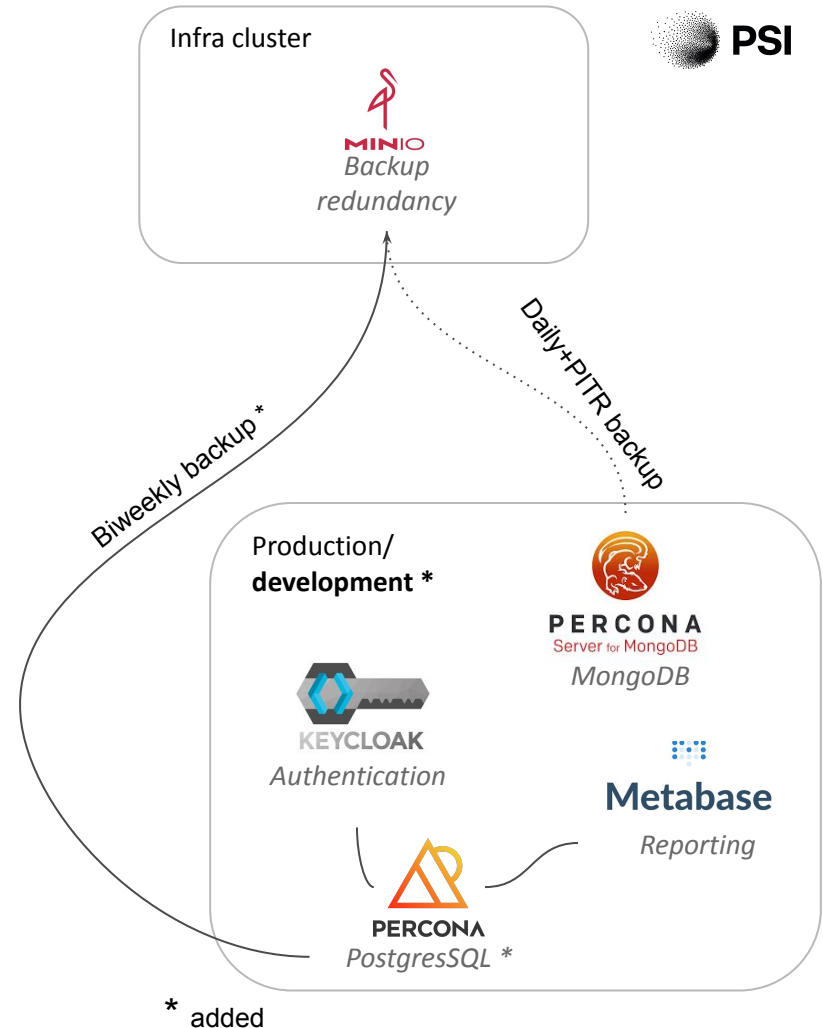| Internal services | SciCat services * | K8S management |

Bastion VM

* restrictions and exposures rules defined per service
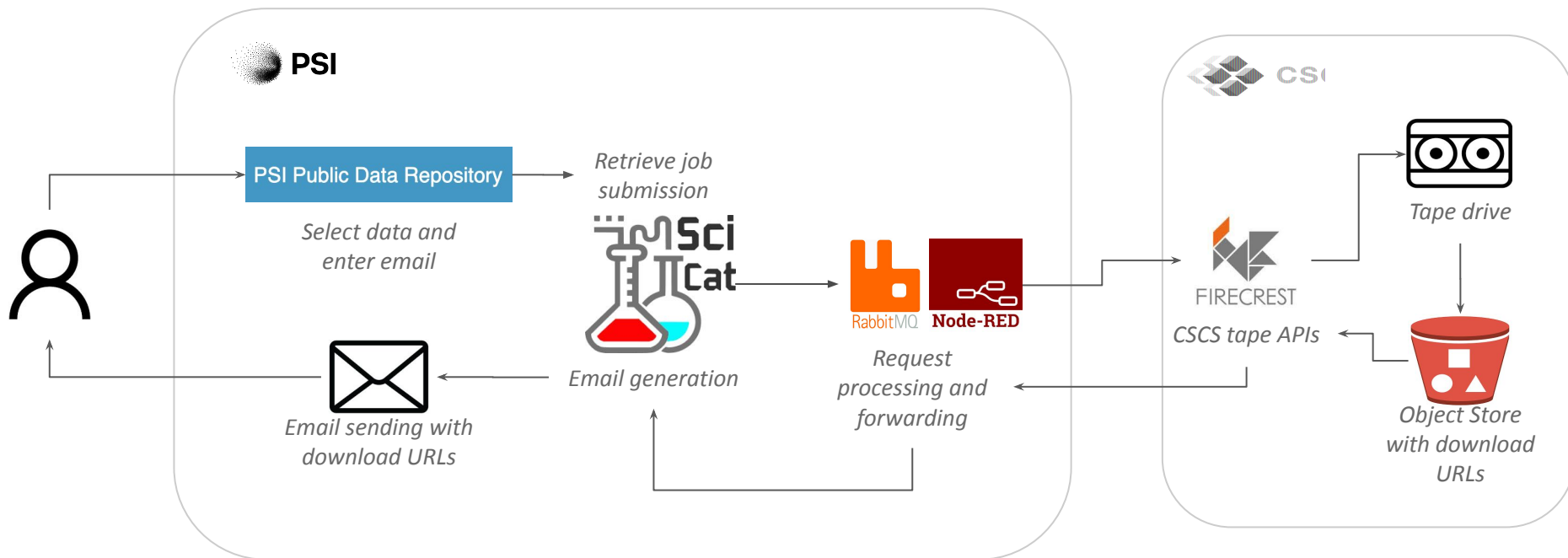
Infrastructure 2/2

Improved infrastructure and backups

1. Upgrade all to latest stable, supported version (dated Feb 24)

2. Reproducibility, atomicity and idempotency, through recommended devops tools, e.g.:
   - Terraform
   - kubeone
   - CRDs
   - Helm charts
   - K8S manifests
   - Krew plugin manager
   - OLM

3. Added backup of development mongoDB and Keycloak and Metabase DB

Infra cluster

Backup redundancy

Daily+PITR backup

Biweekly backup *

Production/
development *

MongoDB

Authentication

Reporting

PostgresSQL *

* added

PSI

User-facing updates 1/3

Retrieve from landing page to LTS object store

PSI Center for Scientific Computing, Theory and Data

User-facing updates 2/3

Golang SciCat CLI

PSI

The PSI scicat-cli is a multiplatform CLI wrapper around the SciCat REST APIs and with light interactions with the PSI infrastructure
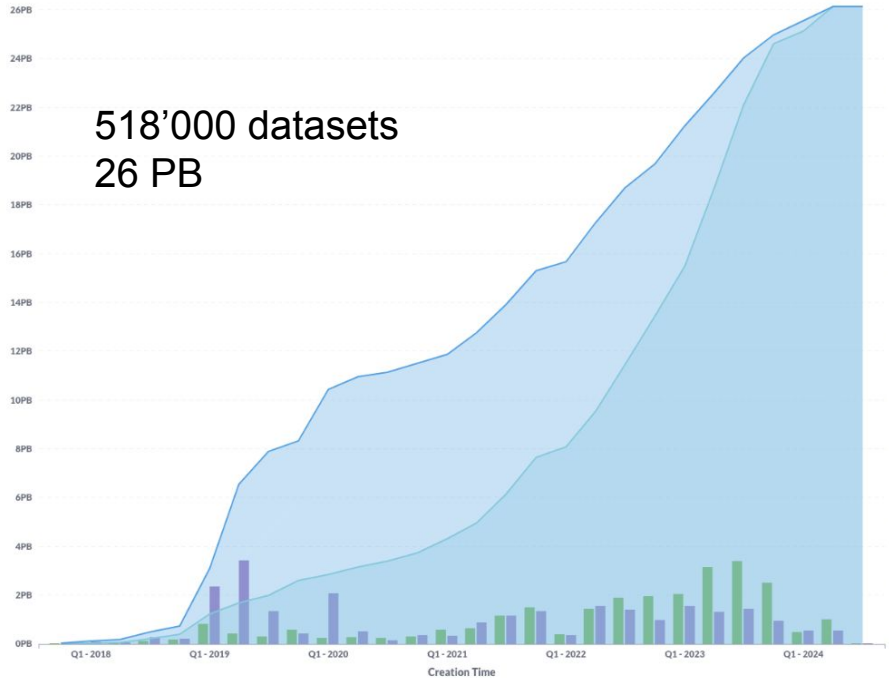
1. Repository publicly available on <u>GitHub</u>

2. Binaries automatically published on GitHub as <u>releases</u> assets (with semantic versioning)
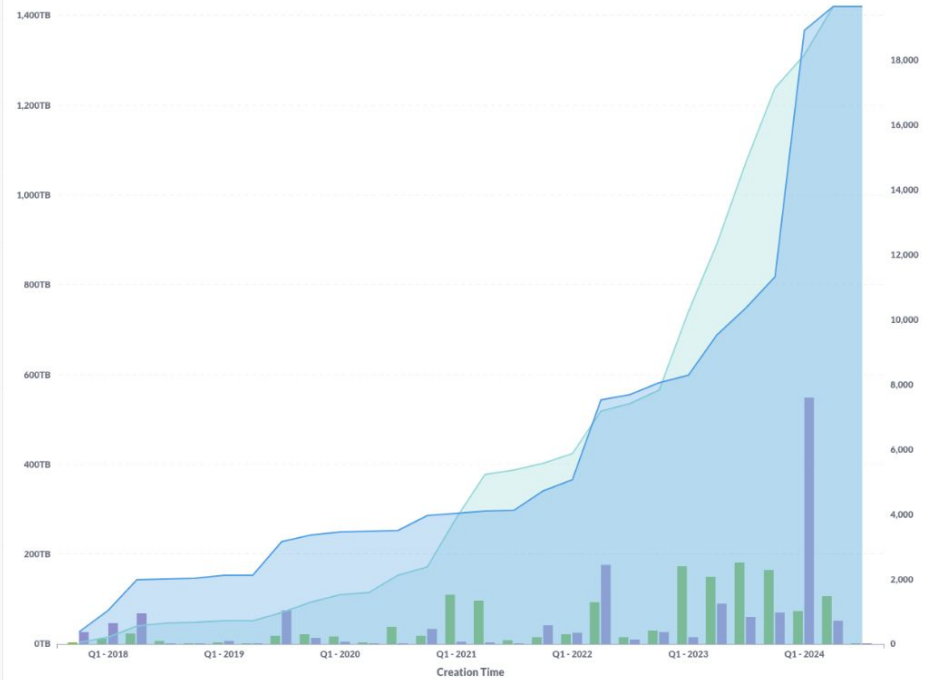
3. Almost fully unit-tested

4. <u>V3</u> in alpha

PSI Center for Scientific Computing, Theory and Data

User-facing updates 3/3
Users' statistics

518'000 datasets
26 PB

Grants

Submitted grants where SciCat is involved

1. Swiss Open research Data Measure 1:
   a. **OpenEM:** [...] a consortium of **Swiss electron microscopy** facilities working together to improve **Open Research Data** practices, provide an open and **FAIR** repository, **standardize** EM metadata & automate collection, follow a consistent data lifecycle, streamline publication of EM data into **open repositories** like EMPIAR, EMDB, and the PDB [...]

2. Swiss Open research Data Measure 2:
   a. **API01**: Improving the interoperability of Electronic Lab Notebooks and data repositories by enhancing API integration
   b. **API02**: Towards a common storage access API
   c. **API03**: Interoperability between the ETH Domain Repositories (IntER)
   d. **API04**: Reproducibility and Interoperability in Active Data Publication (RI-ADP)

Future plans

- Migration to BE v4
- Federated authentication
- Multiple backend storages
- Federated access group management
- UI for dataset creation and ingestion
- Extend the scicat-cli for better data upload
- Automatic metadata extraction
- Metadata schemas
- Auto publishing after embargo period
- API integration with ELNs, workflow managers, e.g. for data redundancy, metadata import/export, Swiss repositories publishing, ...
- Object store compliant SciCat API layer
- Integration with Beamline Experiment Control system at PSI
- (scicatlive for K8S, i.e. a SciCat operator)
- <u>Open positions</u>
- …

PSI Center for Scientific Computing, Theory and Data