ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Managed Database

Pierre Chanial, EGO

Why?

EOSC
1.Access to all European research data
2.Access to world-class data services
3.Clear rules of use and service provision
4.FAIR data tools, training and standards
5.Sustainable after the grant

Why?

- Databases provide space to store and **index** user data
  - To do further analysis
  - Easy to be combined with other data sources through the data lingua franca: SQL
  - Cross-fertilization: new science
- But database admin takes a lot of time
  - Setting up, backups, redundancy

Current status:
Access to isolated tabular data sources

SQL

ESAP
Analysis Platform

SQL

User database connector

- Managed by ESAP-DB
- On-demand creation of dbs, table sets and tables
- Owned by the user

ADQL

ESFRI APIs

SQL

IVOA connector

- Use of column names as-is
- Complex queries

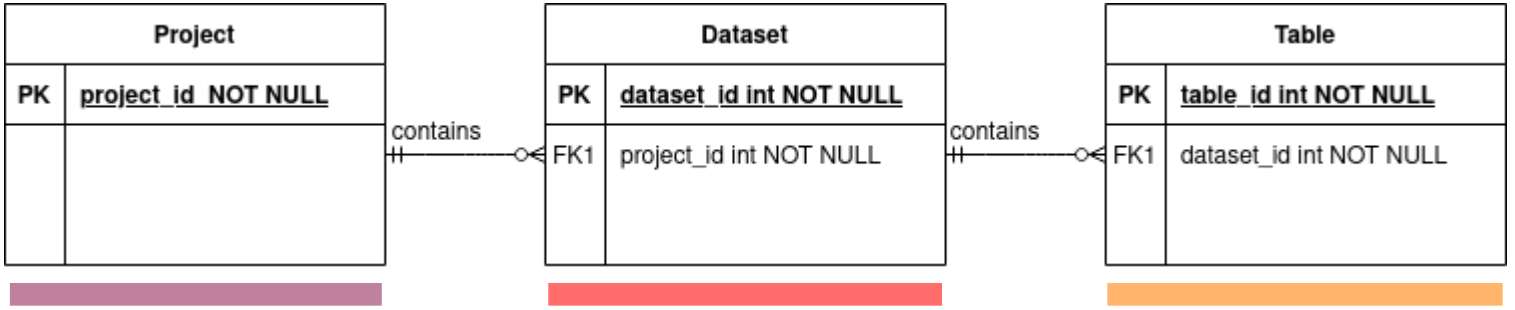
ESFRI connector

- Standardized, pre-formatted queries

External resource connector

- CSV files (through http)
- RDBMs (SQLAlchemy)
- Access with read-only role

ESAP-DB data model



SELECT did FROM `project_id.dataset_id.table_id` WHERE metadata='xyz'

ex: 'escape_datalake'

Rucio's Scope

Rucio's Dataset or Container

ESAP-DB stack

- Back-end (docker-compose):
  - ESAP-DB: FastAPI / Gunicorn (uvicorn)
  - DB admin (postgresql 12)
  - 2 x DB projects (postgresql 12)
  - PGAdmin
  - soon: Celery & Flower
- ESAP-DB client (Python)

Next step: federation of heterogenous data sources

SQL

ESAP
Analysis Platform

Trino / openLookEng: SQL-to-everything federation

openLooKeng

Data-lake connector

- Access to metadata
- NoSQL Elastic Search?

User database connector

- Managed by ESAP-DB
- On-demand creation of DBs, table sets and tables
- Owned by the user

IVOA connector

- Use of column names as-is
- Complex queries

ESFRI connector

- Standardized, pre-formatted queries

External database connector

- Most RBDM
- Most NoSQL DB

If you want to go fast, go alone

If you want to go far, go together