

Basic Metadata Catalogues



TA2 Meeting Göttingen

Hubert Simma
DESY

Jan 30, 2025

FAIR data: A **local** (private) implementation

- ❑ Each data object is a row in a **database** (table) with **fields**:

*PID	metadata	data
------	----------	------

- ❑ Choose some database implementation

e.g. buy a big disk with POSIX file system and use standard tools (ls, grep, ...)

```
-rw----- 1 simma punch 439873 Jan 30 11:55 file.hd5
```

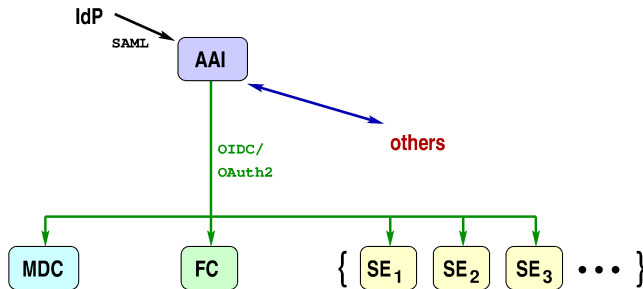
administrative metadata

data
+
scientific
metadata

FAIR data: A **distributed** (shared) implementation

Distinct **web services**:

- Metadata Catalogue (MDC)
- File Catalogue (FC)
- Storage Element (SE)
- AAI



where

- separate MDC and SE becomes mandatory for **large data objects** (cost of search)
- multiple SE may become mandatory in practice (replication, funding, ownership)
- FC becomes mandatory if there are multiple SE or varying storage locations (SURL)

→ **modular distributed architecture**

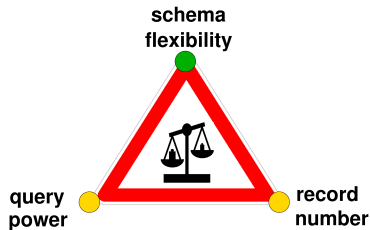
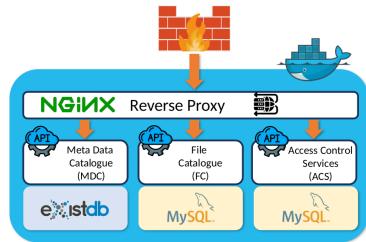
✓ MDC and FC running (v1)

- Access Control Service (ACS) as sidecar

✗ VOMS + X509 certificates

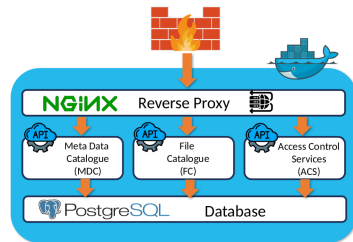
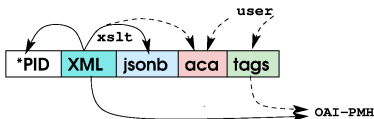
✗ MDC tradeoff

✗ No access control for SE



- ✓ MDC and FC running (v2)
- ✓ dedicated IAM instance at INFN/CNAF (X.509 free, registered as eduGAIN SP)
- PostgreSQL

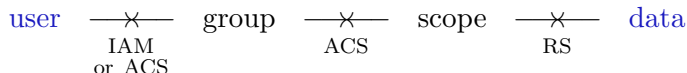
XML ↔ jsonb
Xpath ↔ “quick-search” (JSONpath)



- ✓ Fine-grained token-based access control for MDC, FC, and SE
- ➔ Modular building blocks, reusable in other use-cases, e.g. in PUNCH and beyond

Authorization Policies

- ❑ Policy **Decision** Point → Access Control Service ([ACS](#))
- ❑ Policy **Enforcement** Points → Resource Servers (RS = [MDC](#), [FC](#), [SE](#))
- ❑ Composition of (smaller) many-to-many relations



- ❑ Path matching

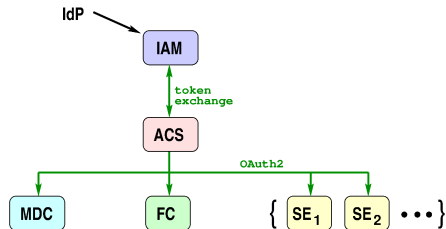
SURL = <https://dcache.ifh.de:2880/a/b/c/d/e/f>

scope	access
storage.read: /	yes
storage.read: /d	yes
storage.read: /d/e	yes
storage.read: /d/e/f	yes
storage.read: /x	no

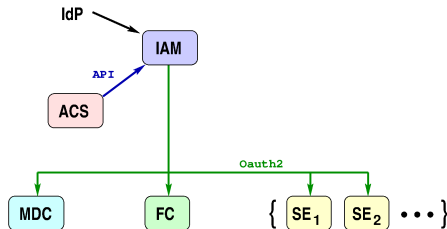
Access Control Service

- ❑ 3rd “metadata catalogue”

- ❑ ACS in front



or beside/behind of IAM



- ❑ ILDG setup is in certain aspects ahead of (or interesting for) other experiments

Missing

- ☐ Ironing of MDC and FC (32+11 issues)
 - ☐ Completion of (passiv) ACS
 - ☐ Clients (GUI, CLI)
 - ☐ Deployment (UK)
 - ☐ Documentation + hands-on
-
- ☐ Integration with FTS
 - ☐ S3 integration
 - ☐ Kubernetes
 - ☐ ...

Web pages

- ❑ [ILDG home](#) (to be improved and moved)
- ❑ [IAM](#) (user registration)
- ❑ [MDC interface](#) (simple listings)
- ❑ [Gitlab](#) (QCDml draft, API, client tools, containerized SW environment, examples, ...)
- ❑ [Hands-on workshop](#)
- ❑ [Markup tool](#)