



Carlo Minotti & Ionuț Iosifescu-Enescu

Interoperable REST API for (meta)data exchange

The interoperability, cooperation and the connections among the ETH Domain services and repositories themselves as well as to larger global community portals and aggregators are inherently strengthened

Introduction (1)

Metadata = "data about data" [<https://howtofair.dk/how-to-fair/metadata/>, 2023.06.15]

- Metadata is defined as the data providing information about one or more aspects of the data including its provenance
- It is used to summarize basic information about data that can make tracking and working with specific data easier.
- It is very important for making your data FAIR!

Sources:

- <https://howtofair.dk/how-to-fair/metadata/>, [2023.06.14]
- <https://en.wikipedia.org/wiki/Metadata>, [2023.06.14]

Introduction (2)

Data repositories:

- Used for data sharing: documentation, data publication, long-term preservation
- E.g. Materialscloud, EnviDat, ETH Research Collections, SciCat (see report)

Generic data search platforms and aggregators:

- Used for searching data across different fields, i.e. support highlevel searches
- E.g. opendata.swiss, data.europa.eu, EOSC Portal, Google dataset search ...

Data sharing APIs:

- Automating querying and accessing data and metadata from repositories
- SciCat <-> Envidat, SciCat <-> Materialscloud, any system <-> any system

The Interoperability/Missing API Problem

- ETH Domain scientists publish datasets all over the place:
 - ETH Domain repositories (Materialscloud, EnviDat, ETH Research Collections ...)
 - Global external repositories (Dryad, Zenodo, ...)
 - even on their own Websites or FTP servers (which will disappear)
- No data sharing data from different ETH sources, following the same rules and FAIR standards! Basically it is a mess!
- Little visibility of open data published by the ETH Domain scientists
- Not all datasets have a DOI assigned by DataCite!

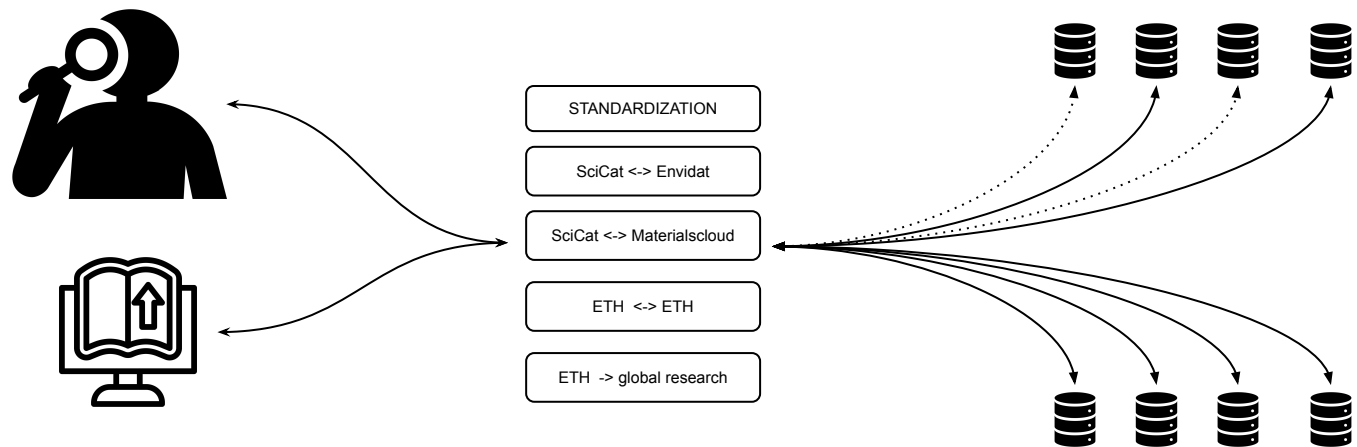
Initial Use Cases (And what else?)

- A scientist needs to:
 1. Find relevant data by searching in one place instead of a wide range of different sources:
 - various ETH Domain repositories
 - global external repositories (Dryad, Zenodo, ...)
 - larger aggregators (opendata.swiss, data.europa.eu, EOSC Portal, DataCite Commons, Google Datasets Search + Google/Web search)
 2. Data follows curator -> export (meta)data from one system to another
e.g. SciCat <-> Envidat, SciCat <-> MaterialsCloud
 3. Increase his dataset visibility by making sure it can be found in large aggregators (e.g. data.europa.eu, EOSC Portal, Google Datasets Search) irrespective of what ETH domain repository is using

Key challenges

- Metadata format standardization: define and agree on a common metadata format for non field-specific search platforms based on the requirements from: DCAT-AP (opendata.swiss); DataCite Schema; etc.. (DC is not enough!)
- API standardization: implement or agree to adopt common API standards for participating data services and repositories in the ETH domain
 - move datasets & (meta)data across different ETH domain repositories
 - Dataset follows researcher principle, e.g. SciCat to EnviDat and vice versa (other repositories very welcome!)
- API to connect to existing global data search platforms (e.g. opendata.swiss, data.europa.eu, EOSC, etc...) based on the same standards in order to increase the visibility of the ETH open data

Conclusion: Fostering ETH Domain Interoperability



ETH facility

— Data system or repository APIs