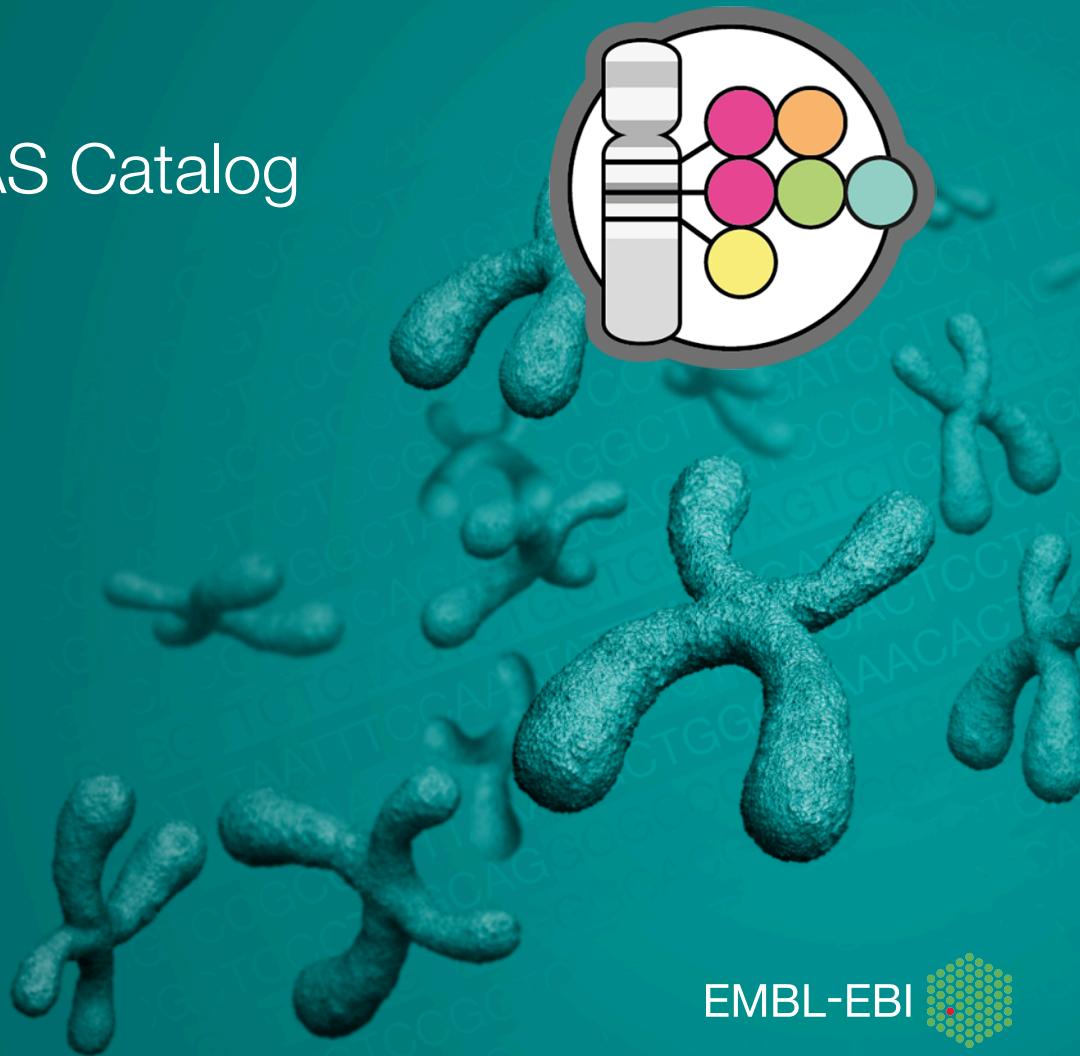


Programmatic data access and retrieval

The API of the GWAS Catalog

Daniel Suveges
Bioinformatician

www.ebi.ac.uk/gwas



API Workshop

- **Presentation**
 - What is an API, why should I use it?
 - How to extract data via API
 - Understanding the returned data
 - GWAS Catalog API
 - GWAS Catalog Summary statistics API
- **Interactive session**

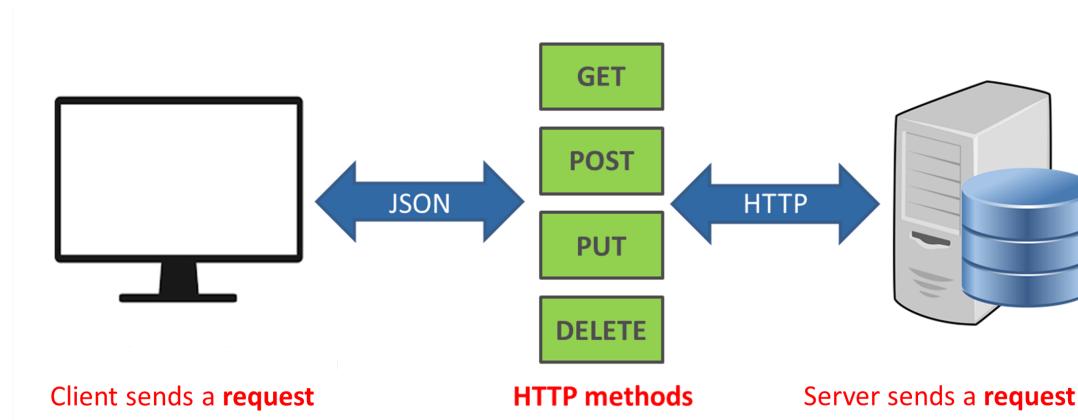
High-throughput access

- Web interface: optimal for low-throughput, exploratory data access.
- How to check 100s or 1000s of variants?



What is an API?

- Interface between application instead of users



- Client: script, other application or user
- Requests are formulated as URLs
- Standardized, machine readable responses

Extracting data

- Requests are URLs:

`https://www.ebi.ac.uk/gwas/rest/api/studies`

API address

endpoint

- Further arguments:

`.../api/studies`

`.../api/studies/GCST000854`

`.../api/studies/GCST000854/associations`

`.../api/studies/search/findByPublicationIdPubmedId?pubmedId=21041247`

Understanding the returned data

- The returned data is formatted in standardized, machine readable format (JSON):

Key-value pairs:

```
{  
    "key1": "value1",  
    "key2": "value2"  
}
```

Lists:

```
[  
    "value1",  
    "value2"  
]
```

- Extract required field eg. study accession.
- User has to be aware of the design of the returned data.
- Documentation of the GWAS Catalog API:

<https://www.ebi.ac.uk/gwas/rest/docs/api>

The GWAS Catalog REST API

- **API address:** <https://www.ebi.ac.uk/gwas/rest/api/>
- **Endpoints:**

.../api/studies/GCST000028/associations

.../api/singleNucleotidePolymorphisms/rs224073/studies

.../api/efoTraits/EFO_0001360

.../api/associations

- **Search by...**

.../api/[endpoint](#)/search

- Search study by pubmed ID etc.
- Search variant by gene, genomic region etc.
- Search association by pubmed ID etc.

Further reading

- Workshop material online:
https://github.com/EBISPORT/GWAS_Catalog-workshop/
- Documentation on the APIs:
 - GWAS Catalog: <https://www.ebi.ac.uk/gwas/rest/docs/api>
 - Sum stats: <https://www.ebi.ac.uk/gwas/summary-statistics/docs/>

Interactive session

- **Exercises:**
 1. Query the REST API using a browser
 2. Programmatical access of the same data
 3. Retrieve a data for multiple entries
 4. Pooling data together from different endpoints (optional)
- Python and interactive Jupyter notebooks: