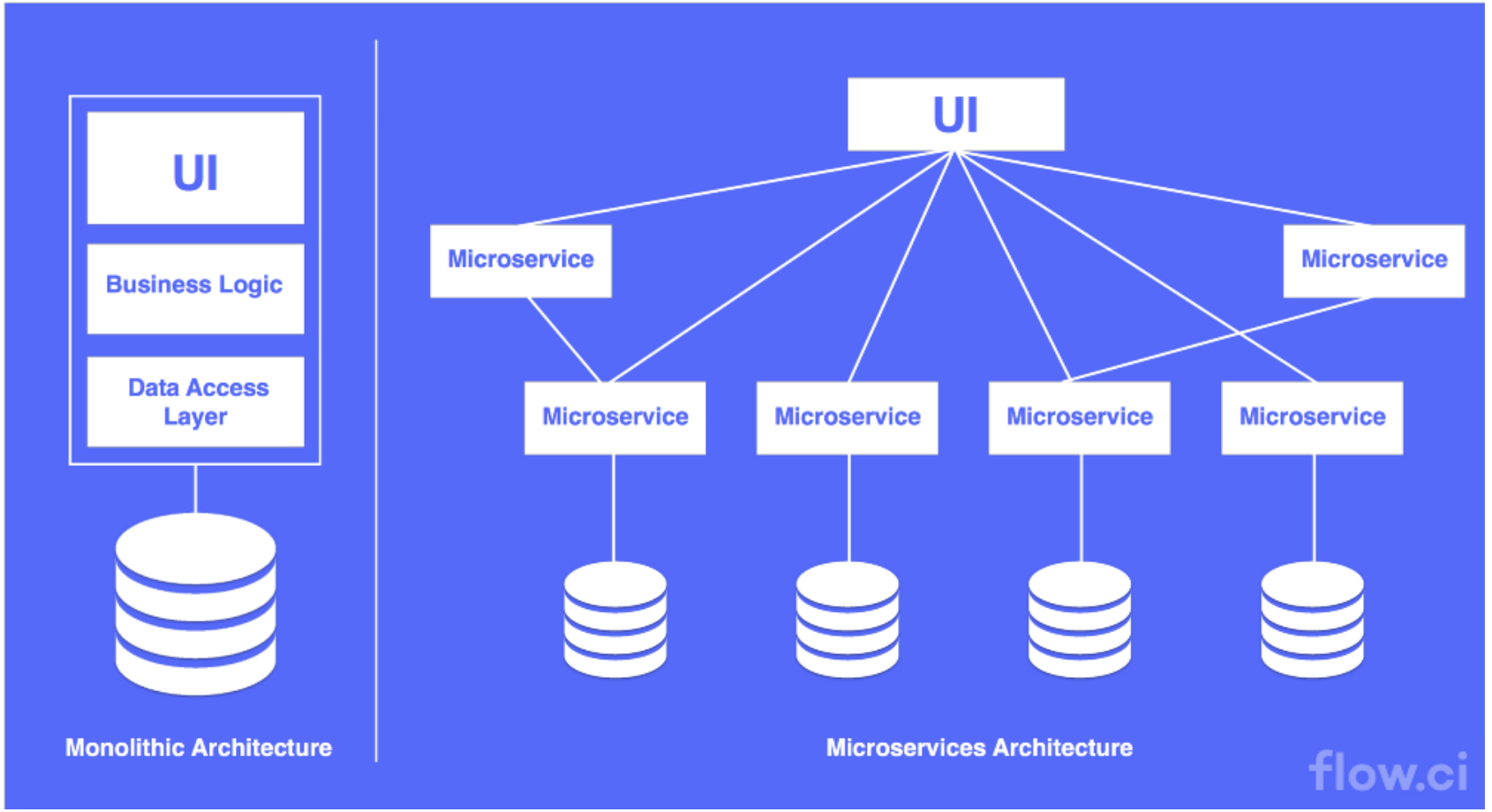
DBW - APIs

APIs – Programmatic access

- API = “Application programming interface”
  1. Set of routines, functions, or procedures/methods, offered as a library to be used inside other software (API = “software library”)
  2. Any web service providing remote functionalities to be used inside other software or most usually as data provider (API = “web service”)
- In bioinformatics the approach is generally used to access data or to allow communication among application components (“microservices architecture”)
- Strategies: SOAP/XML, XML-RPC (remote executions), **REST**
- Authentication implicit via Oauth2 / openIDConnect

Microservices architecture



REST (REpresentational State Transfer)

- REST-ful web services
  - Used primarily to serve data
  - Data can be pre-processed at the server-side (so becoming a kind or RPC)
  - Controlled through HTTP and called using standard URLs
    - /api/{store}/{id}/option.format?option
  - HTTP interfaces to Data repositories
- HTTP based (GET, PUT, POST, DELETE)
  - Allow to GET, PUT (update), POST (insert), and DELETE a resource in the data portal (i.e. a DB)
- Using REST APIs, application components can be independent (and distributed) as long they communicate using HTTP and a known format

Data exchange languages

- Data exchange formats
  - XML: Most traditionally used by web services (SOAP, RPC)
    - Same structure as HTML, but with no fixed tags
    - Requires XML-schema to specify tags and check coherence

```
<Course id="DBW">
<Acronym>DBW</Acronym>
<Title>Databases and Web applications</Title>
</Course>
```

  - JSON: Data interchange format replacing XML (most popular)
    - Natively understood by Javascript

```
Course: {Acronym: 'DBW', Title: 'Databases and Web applications'
```

- Both require "schemas" to validate data model

Programmatic Access (client side)

- Perl

```
use LWP:Simple;
use JSON;
my $content = decode_json(get('http://...'));
```

- PHP

```
$data = json_decode(file_get_contents("http://...."));
```

- Python

```
import requests
data = requests.get('http://...').json()
```

Web service : server side

- Usual web applications but…
  - Output is not meant to be shown in browsers (no HTML, CSS, JS)
  - Headers required
    - `Content-type: text/xml | text/plain | application/json | application/x-gzip | image/png`
      - Define the type of data being sent
    - `Content-Disposition: attachment; filename=file_name`
      - Force download (when seen from a browser)
    - `Access-Control-Allow-Origin: *`
      - Allow access from any client (to avoid security checks on JS/AJAX)
  - Formats can change
    - In theory should be requested via HTTP (Accept…) but normally are included in the URL
  - Error handled via HTTP codes
    - 200 ok, 404 not found, …
  - Prevent caching
    - Header: `Cache-Control: no-cache`
- Programming frameworks are very useful here due to the complex routing

- A “quite complex” backend : http://mmb.pcb.ub.es/gitlab/MMBData/MMBApi

RESTful URLs

- No standard
  - A typical schema is

    /api/{store}/{id}/option.format?options

    ex. `http://mmb.pcb.ub.es/api/pdb/2ki5/entry.json`

- Documented via
  - Ad-hop help pages
  - OpenAPI (a.k.a. Swagger) (recommended)

and Bioinformatics

Home

MMB Data repository API Interface

Statistics

/info[.xml|.json]	PDB/Uniprot data repository information (default json)

Protein Data Bank

/pdb/{id}/entry[.xml|.json]	PDB full entry metadata: Ascession Date, Experiment type, re chain ids, ligand data, remarks, chain sequences, sequence c json). Individual fields can be recovered, completing the URI
Parameters	

Usage example	`/pdb/2ki5/entry?[.xml|.json]`
	
	SwissProt Hit on chain 0 (A) sequence.
	`/pdb/2ki5/entry/chains/0/swpHit/idHit`

/pdb/{id}[_bn{n}] /headers[.gz]	Headers from PDB file
Options	**{id}_bn{n}**: Biounit {n} instead of assymetric unit

ICGC DCC Docs

Home | Portal | Download | Submission | Dictionary | Software | PCAWG

API ENDPOINTS

Overview

ICGC API is a set of RESTful endpoints -- programmable interfaces over the Web -- that allows third-party developers to build automation scripts and apps. This documentation describes ICGC API in details, including data model information of both inputs (parameters) and outputs (response records). It also allows you to interact with and test out each API directly on this page, which shall provide clear insights into how the API responds to different parameters.

Endpoints

https://dcc.icgc.org

browser — Show/Hide | List Operations | Expand Operations

PQL — Show/Hide | List Operations | Expand Operations

analysis — Show/Hide | List Operations | Expand Operations

POST	/v1/analysis/enrichment	Submits an asynchronous enrichment analysis request. Users must poll the status using the GET resource
GET	/v1/analysis/enrichment/{analysisId}	Retrieves an enrichment analysis by id
POST	/v1/analysis/phenotype	Creates a new Phenotype analysis by providing IDs of Donor entity sets.
GET	/v1/analysis/phenotype/{analysisId}	Retrieves the result of a phenotype analysis by its ID.
POST	/v1/analysis/survival	Creates a new Survival Plot analysis by providing IDs of Donor entity sets.
GET	/v1/analysis/survival/{analysisId}	Retrieves the result of a phenotype analysis by its ID.
POST	/v1/analysis/union	Creates a set analysis asynchronously. Status can be retrieved by polling the /{id} GET endpoint.
POST	/v1/analysis/union/preview	Retrieves a sample data of a set analysis as preview.

Full entries and sequences

- /api/pdb/{id}/entry/ /api/pdbMonomer/{id}/entry/
  - Full data in XML or JSON

- /api/pdb/{id}.fasta

- /api/uniprot/{id}/entry

- /api/uniprot/{id}.fasta

PDB search options

- /api/pdb/ Search on PDB
  - **resmin=value, resmax=value** Min Max for resolution (XRAY only)
  - **qcompType=(prot, nuc, prot-nuc, carb, other)** Compound types.
  - **qexpType=(ELECTRON_CRYSTALLOGRAPHY, ELECTRON_MICROSCOPY, FLUORESCENCE_TRANSFER, INFRARED, NEUTRON_DIFRACCTION, NMR, SOLID-STATE_NMR, X-RAY)** Type of Experiment.
  - **query=txt** Text query
    queryOn=(header, compound, sources, authors)
  - **sequence=seq** Sequence match
    - **molTy=(protein | na)** Sequence type
      seqType=(exact | regex) Type of sequence match (exact | regular expression)

PDB options for structures

- /api/pdb/{id}/ /api/pdbMonomers/{id}/
  - Default: standard PDB coordinates(possible .gz)
  - Available filters
    - **bunit={n}** Show Biounit n instead of the Assimetric Unit
    - **noheaders=1** Skip PDB headers (implicit in the following filters)
    - **group=(ATOM | HETATM)** PDB label selection. (HETATM includes CONECT)
    - **groupRes=[!](POLAR | APOLAR | NUC | PROT)** Residue type selection. "!" negates
    - **groupAt=[!](POLAR | APOLAR | NOH | BACK | NABACK)** Atom type selection, "!" negates
    - **filter=[!][RES]nres:chain.atom/model** Atom filter using J(s)Mol format ("!" negates selection)