THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society

DATA
CULTURE
SOCIETY

API USE FOR RESEARCH

SESSION 1

Course tutors

Dr. Somya Iqbal, Alex Crest, & Dr. Aybuke Atalay

AYBUKE ATALAY

ALEX CREST

SOMYA IQBAL

Housekeeping and itinerary

Session 1 - Day 1 - 26/11/25

Remit: Understanding fundamentals and demonstration

- [14:00-14:20] - **Talk**: Introduction to APIs and use cases
- [14:20-14:30] - **Q & A** - main goals from attendees for API use and domain
- [14:30-14:40] - **Talk** & brief **demo**: Web based access for APIs (searchable interface) a short demonstration
- [14:40-15:00] -**Talk**: Understanding API documentation
- [15:00-15:20] - **Talk** : Understanding file outputs with JSON/XML structures

10-minute break

- [15:30-15:45] - **Talk** and **demo**: Understanding API user keys/authentication
- **Demo**nstration (Google APIs)

Final 15 minutes

- Discussion: **Q & A**
- Next session overview for practical & exercise
- Record attendance & feedback

Introduction

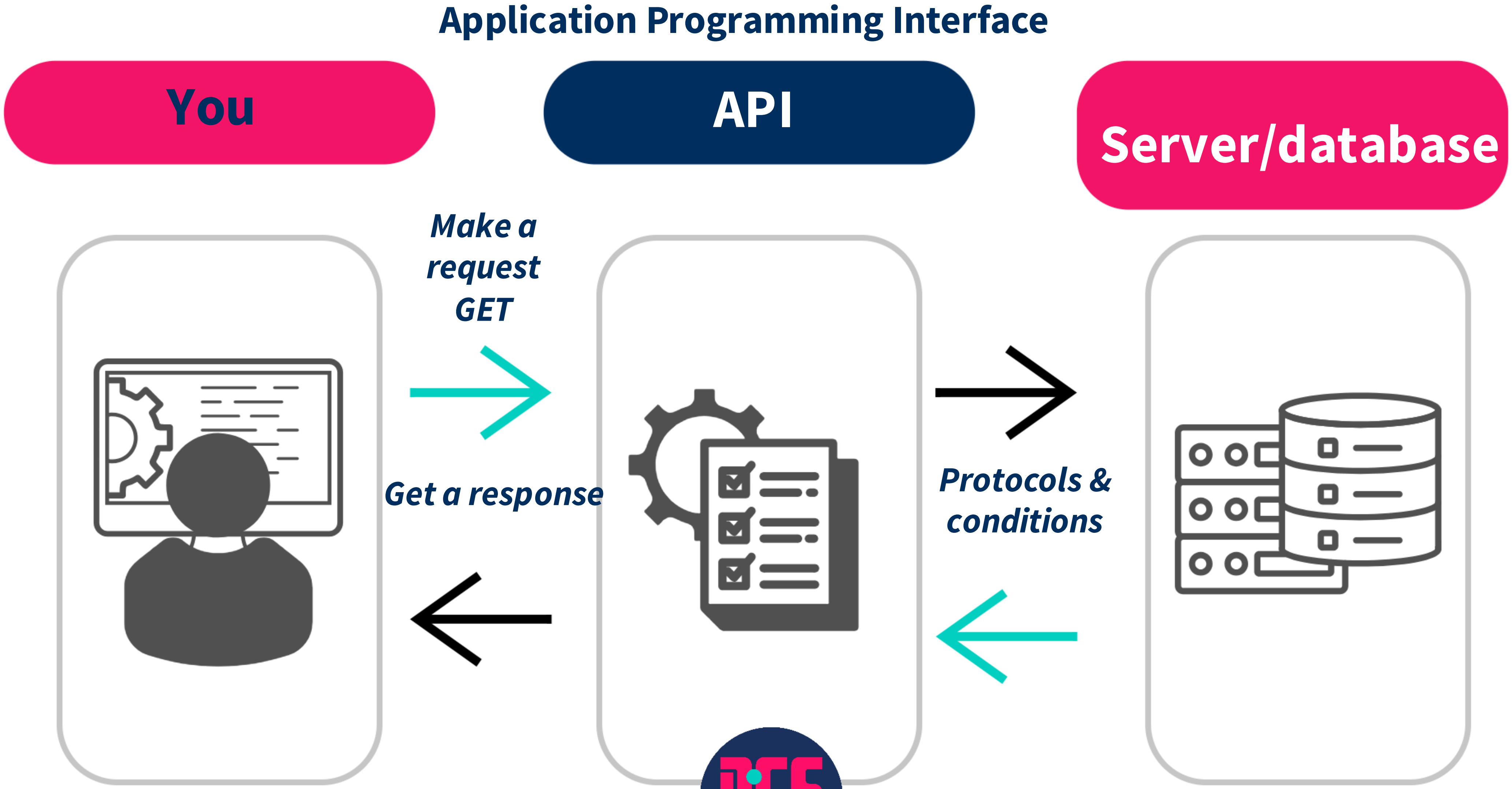

Application Programming Interface

Image: SI

API Types & uses

Are they open? public? Private? Internal?

Differing architectures in how these are setup

REST

Clients use HTTP verbs--> GET (read), POST (create), PUT/PATCH (update), DELETE (remove). For example, GET /api/objects/123 will get object #123

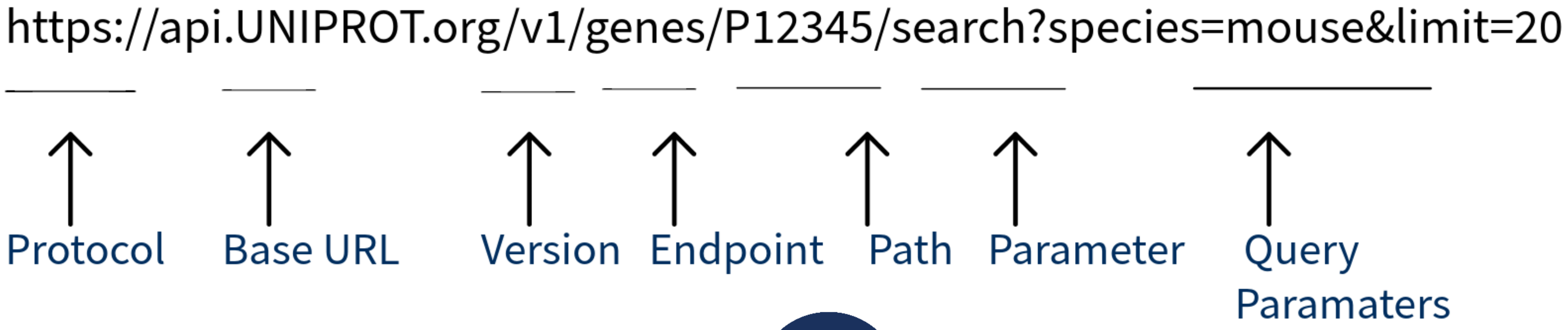
SPARQL

(Sparkle). This is a query language specifically for RDF (Resource Description Framework) data (the semantic web standard). SPARQL queries graph-like data (subject-predicate-object triples) and returns results.

GraphQL

Instead of fixed endpoints, GraphQL lets clients request exactly the data they need via a query language. It operates over a single HTTP endpoint.

HTTPS and the web...



https://api.UNIPROT.org/v1/genes/P12345/search?species=mouse&limit=20

Protocol

Base URL

Version

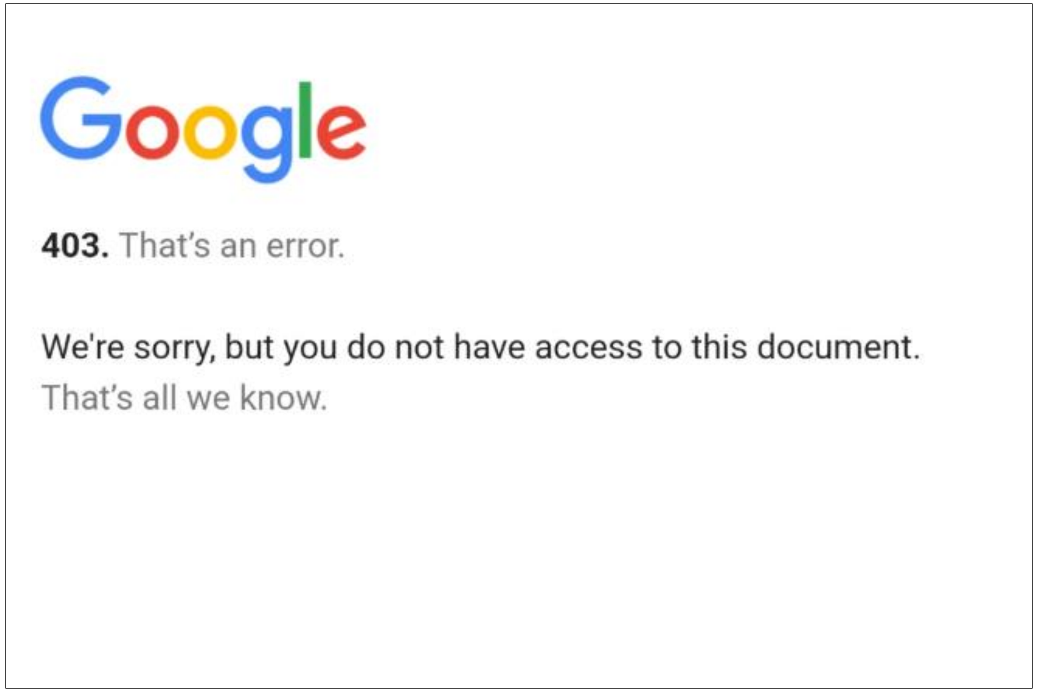
Endpoint

Path

Parameter

Query Paramaters

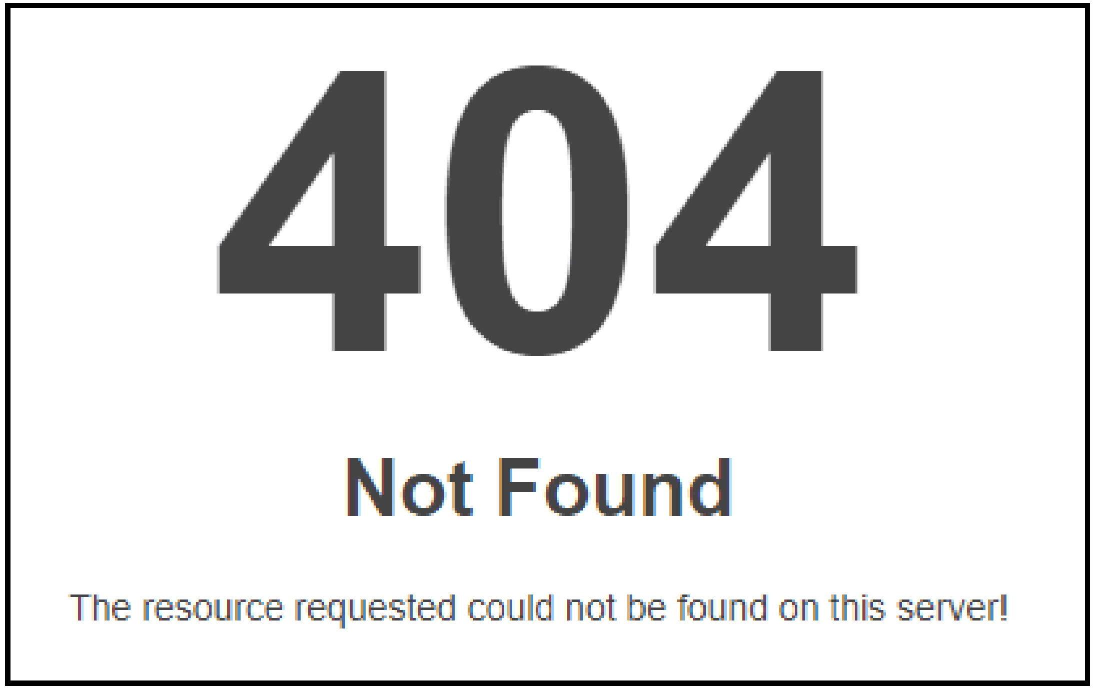
Image: SI



Google

403. That’s an error.

We're sorry, but you do not have access to this document. That’s all we know.



404

Not Found

The resource requested could not be found on this server!

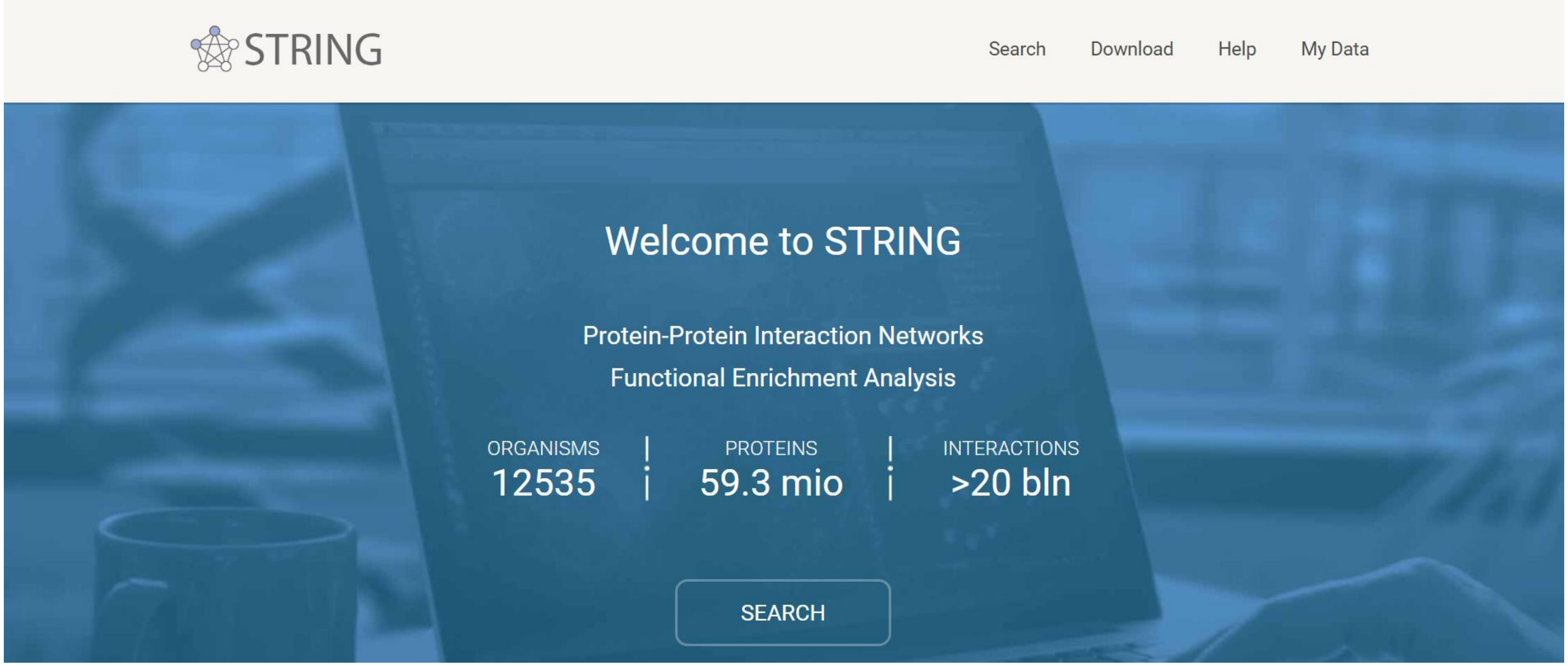
Discussion

What are your main goals & interest in APIs?

GUI (Graphical User Interface)

- Different from an API ---> A GUI is a visual interface that allows users to interact with applications using elements like windows, buttons, and menus.
- Designed to be intuitive and user-friendly, enabling real-time data manipulation and interaction
- Interactive and platform specific
- Requires minimal programming skill
- Often used to make initial assessments, like a broad search or distinct information.

GUI - brief demonstration



STRING

Search Download Help My Data

Welcome to STRING

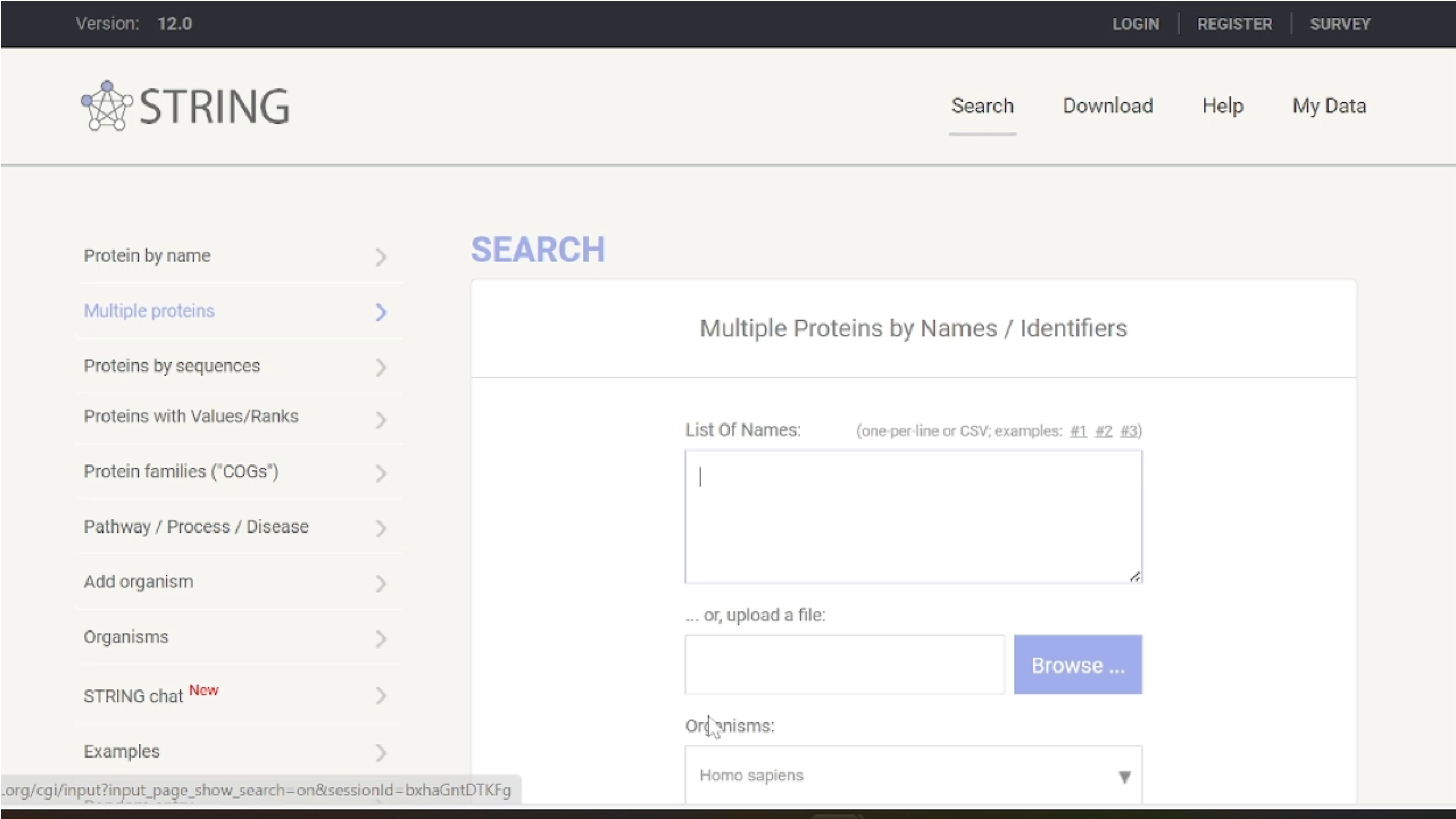
Protein-Protein Interaction Networks

Functional Enrichment Analysis

ORGANISMS 12535 | PROTEINS 59.3 mio | INTERACTIONS >20 bln

SEARCH

https://string-db.org/cgi/input?sessionId=bxhaGntDTKFg&input_page_show_search=off



SEARCH

Multiple Proteins by Names / Identifiers

List Of Names: (one-per-line or CSV; examples: #1 #2 #3)

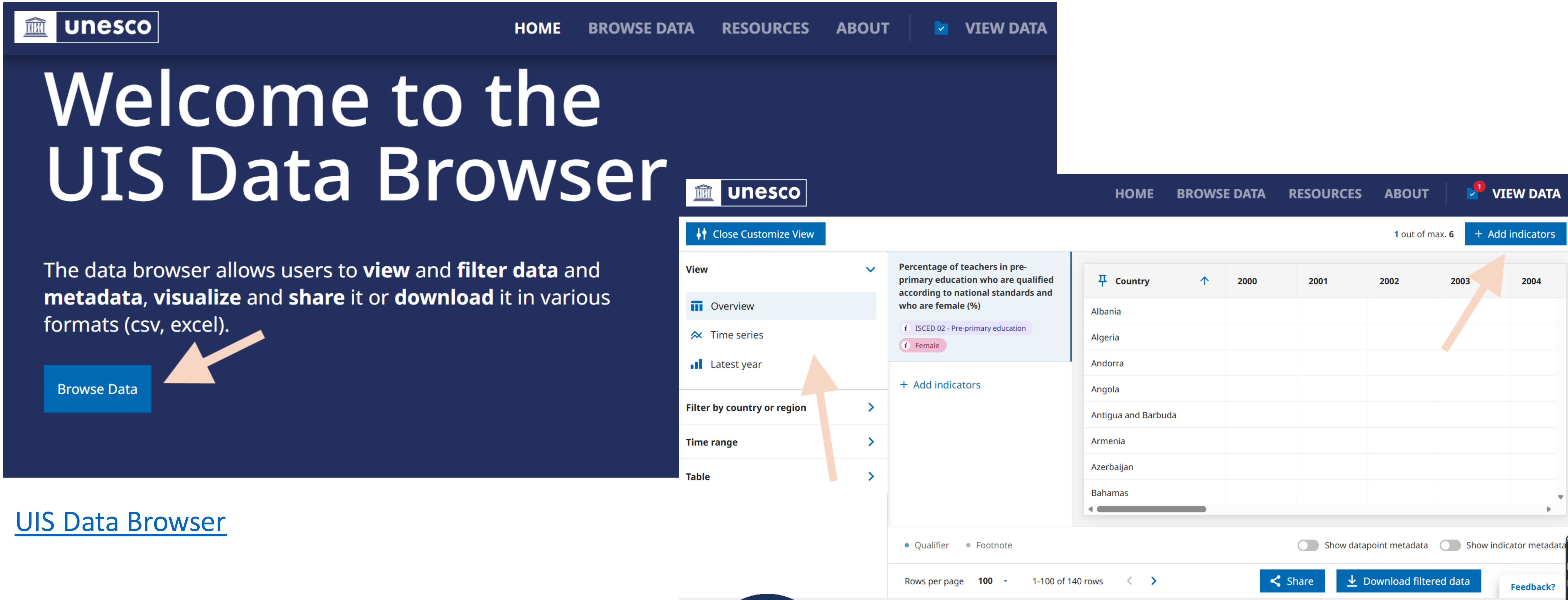
... or, upload a file:

Organisms:

Homo sapiens ▼

.org/cgi/input?input_page_show_search=on&sessionId=bxhaGntDTKFg

GUI



Welcome to the UIS Data Browser

The data browser allows users to **view** and **filter data** and **metadata**, **visualize** and **share** it or **download** it in various formats (csv, excel).

Percentage of teachers in pre-primary education who are qualified according to national standards and who are female (%)

ISCED 02 - Pre-primary education

Female

Country	2000	2001	2002	2003	2004
Albania					
Algeria					
Andorra					
Angola					
Antigua and Barbuda					
Armenia					
Azerbaijan					
Bahamas					

Rows per page 100

1-100 of 140 rows

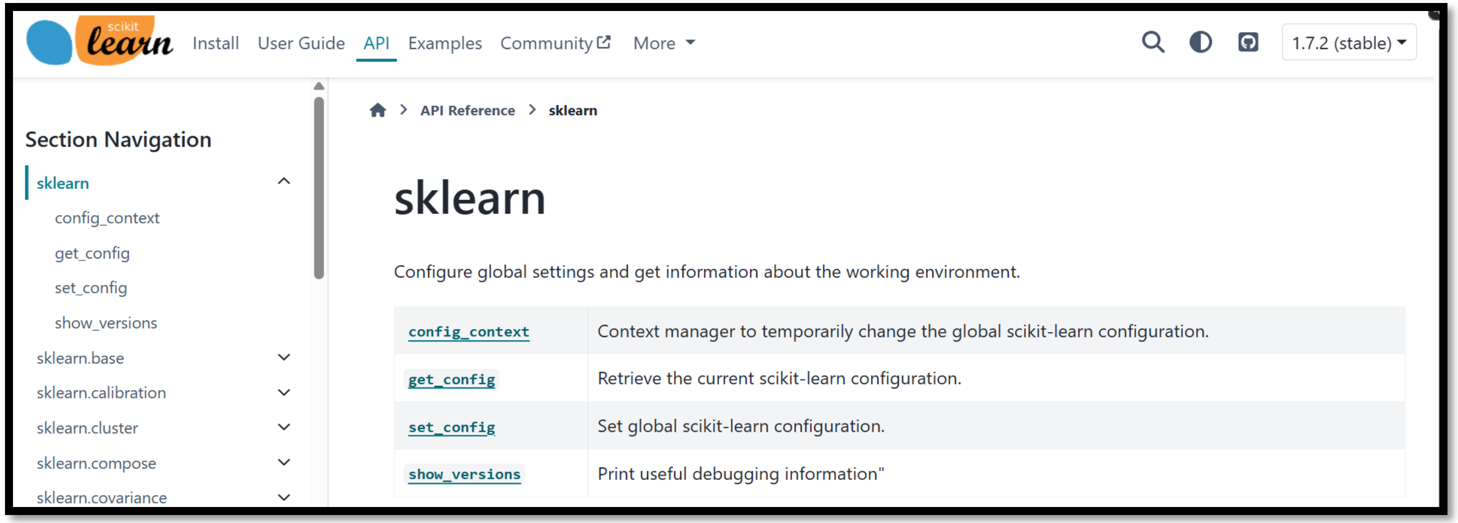
[UIS Data Browser](#)

Software/library examples

Scikit learn is a well-known library for machine learning users

- Composite API
- Can be accessed using Python programming
- Contains validated algorithms for users to apply with ease
- Encourages reproducibility

https://scikit-learn.org/stable/api/index.html



sklearn

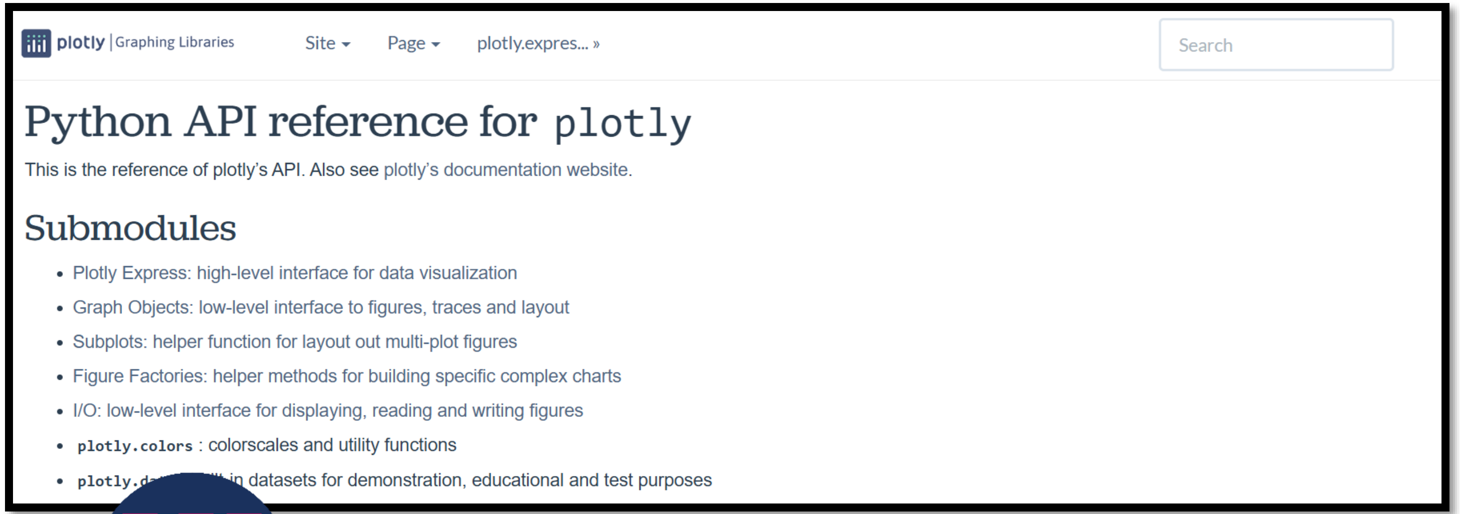
Configure global settings and get information about the working environment.

`config_context`	Context manager to temporarily change the global scikit-learn configuration.
`get_config`	Retrieve the current scikit-learn configuration.
`set_config`	Set global scikit-learn configuration.
`show_versions`	Print useful debugging information"

Plotly is a graphing library –> for visualisation (web-based) and interactive

- Plotly has well laid out API for making flexible objects from your visualisations
- We use the library features to describe what we want to view and implementation is quick
- Outputs in web formats so users can have interactive features.

https://plotly.com/python-api-reference/



Python API reference for `plotly`

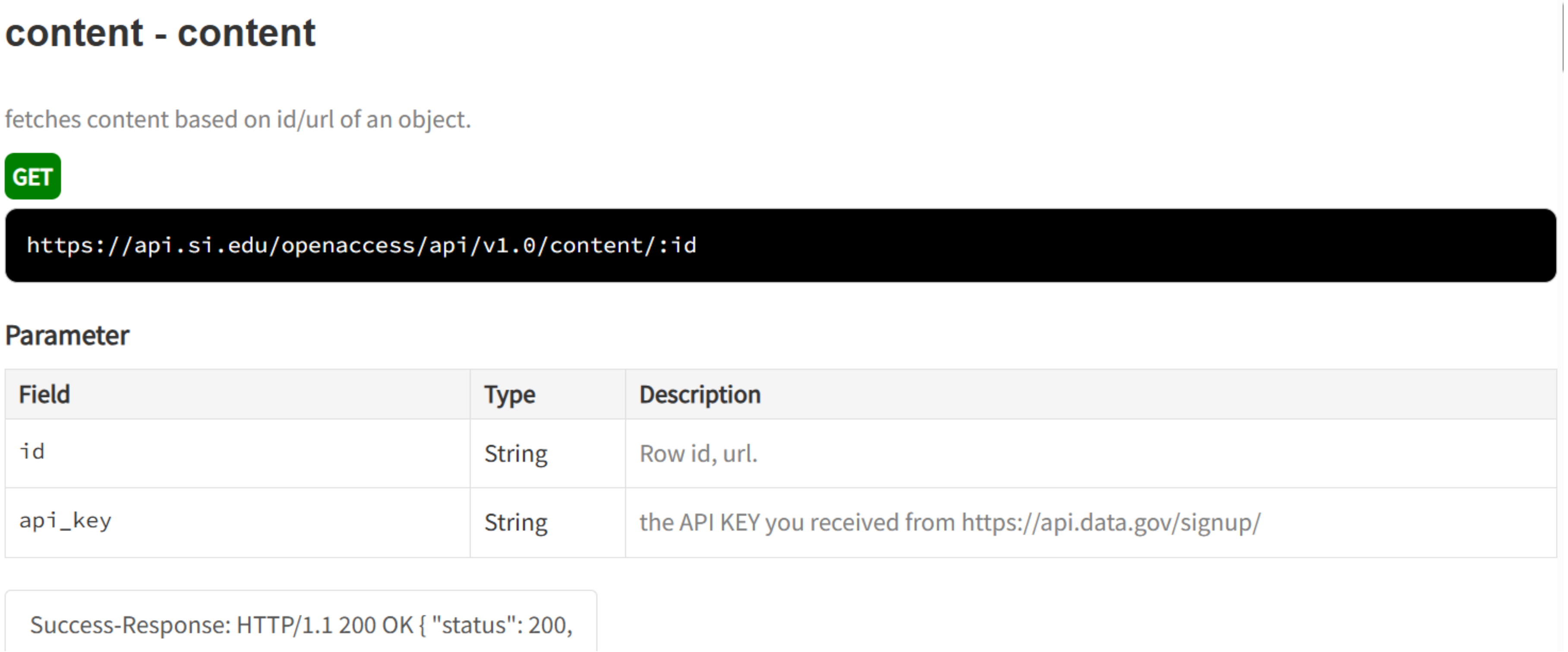
This is the reference of plotly's API. Also see plotly's documentation website.

Submodules

- Plotly Express: high-level interface for data visualization
- Graph Objects: low-level interface to figures, traces and layout
- Subplots: helper function for layout out multi-plot figures
- Figure Factories: helper methods for building specific complex charts
- I/O: low-level interface for displaying, reading and writing figures
- `plotly.colors` : colorscales and utility functions
- `plotly.da` in datasets for demonstration, educational and test purposes

Understanding documentation for APIs

- The user manual for an application programming interface.
- Each API has a **controlled vocabulary** but differs between them.
- Instructions are clear and updated as well as details about versions.
- Specific R and Python projects also have documentation for APIs.



content - content

fetches content based on id/url of an object.

GET

```
https://api.si.edu/openaccess/api/v1.0/content/:id
```

Parameter

Field	Type	Description
id	String	Row id, url.
api_key	String	the API KEY you received from https://api.data.gov/signup/

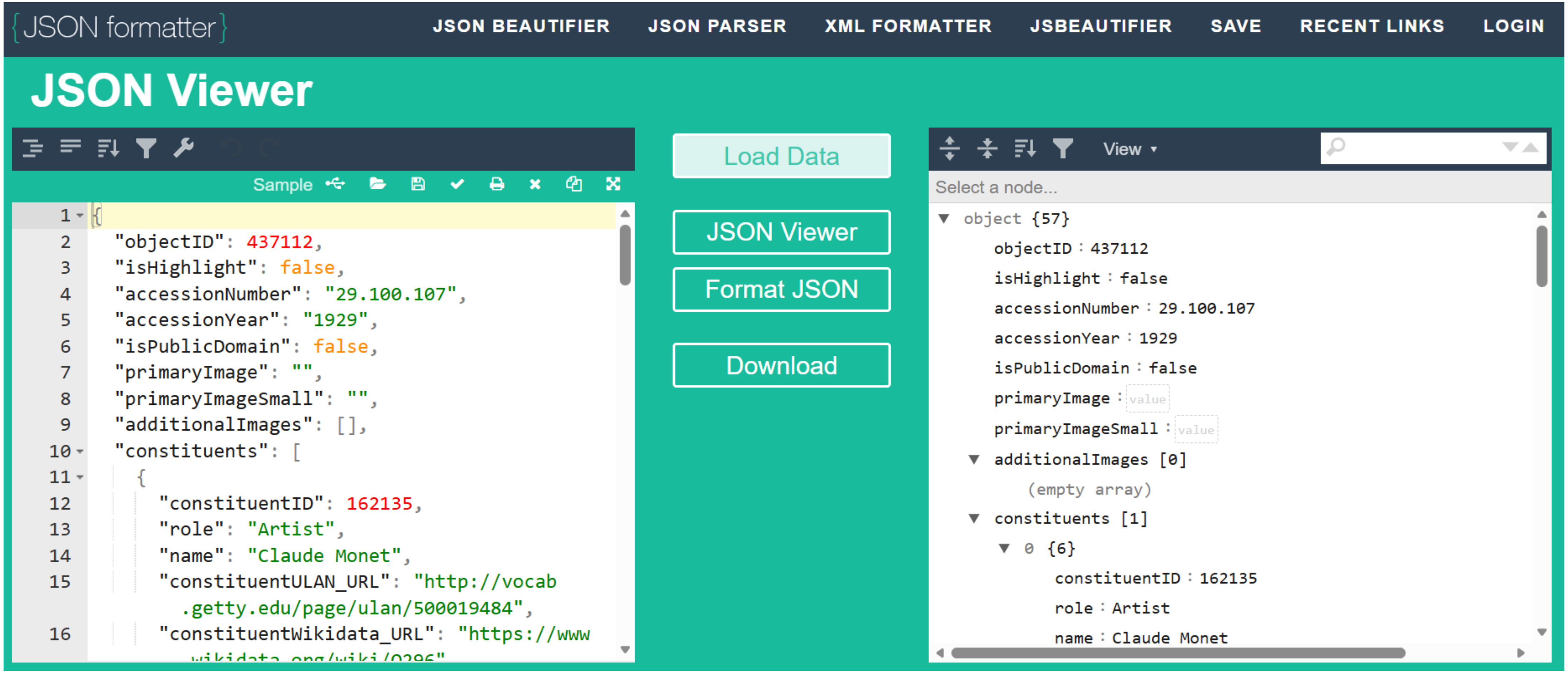
Success-Response: HTTP/1.1 200 OK { "status": 200,

Example from Smithsonian: Documentation

Common file outputs & types

JSON	**J**ava **S**cript **O**bject **N**otation
XML	E**x**tensible **M**arkup **L**anguage
RDF	**R**esource **D**escription **F**ramework
CSV	**C**omma **S**eparated **V**alues

File outputs & types cont….

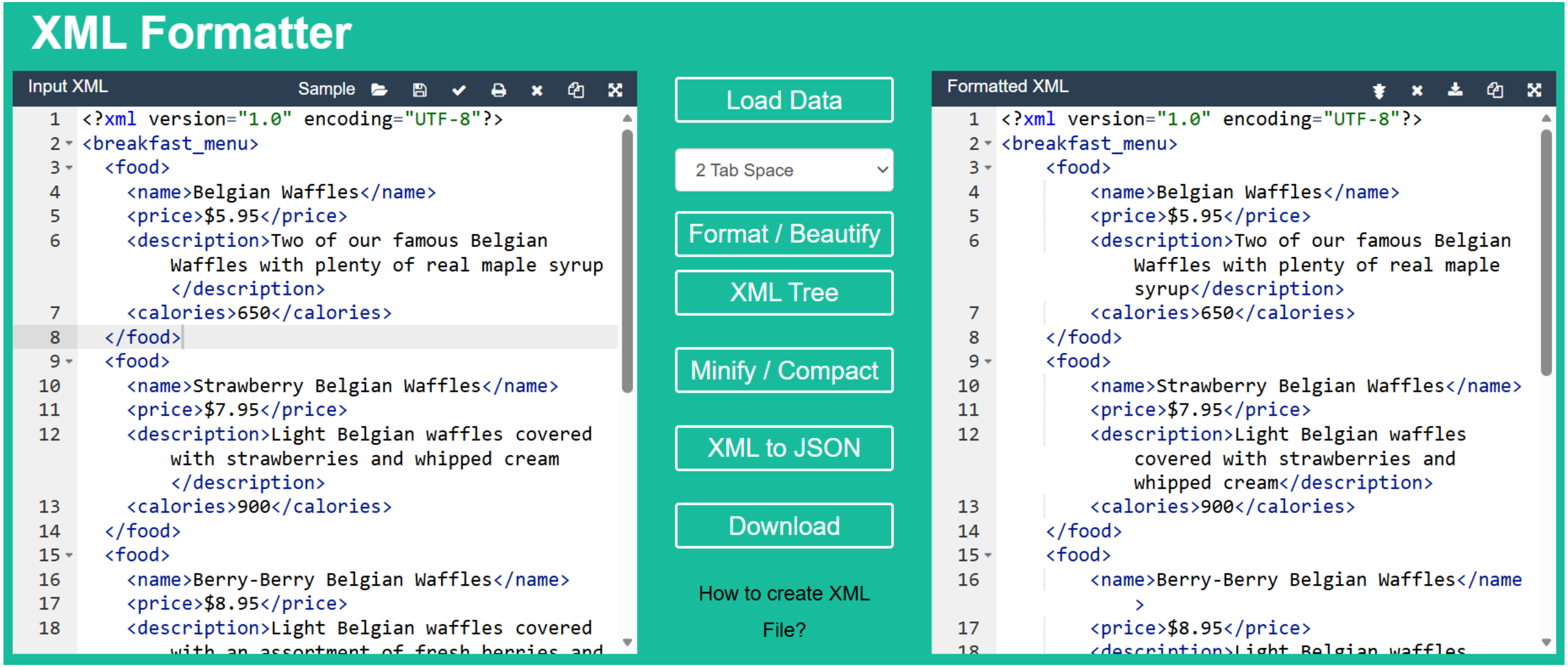


```
{
  "objectID": 437112,
  "isHighlight": false,
  "accessionNumber": "29.100.107",
  "accessionYear": "1929",
  "isPublicDomain": false,
  "primaryImage": "",
  "primaryImageSmall": "",
  "additionalImages": [],
  "constituents": [
    {
      "constituentID": 162135,
      "role": "Artist",
      "name": "Claude Monet",
      "constituentULAN_URL": "http://vocab
        .getty.edu/page/ulan/500019484",
      "constituentWikidata_URL": "https://www
        .wikidata.org/wiki/Q296",
```

```
▼ object {57}
    objectID : 437112
    isHighlight : false
    accessionNumber : 29.100.107
    accessionYear : 1929
    isPublicDomain : false
    primaryImage : value
    primaryImageSmall : value
  ▼ additionalImages [0]
        (empty array)
  ▼ constituents [1]
    ▼ 0 {6}
          constituentID : 162135
          role : Artist
          name : Claude Monet
```

https://jsonformatter.org/

File outputs & types cont....



Input XML

```
<?xml version="1.0" encoding="UTF-8"?>
<breakfast_menu>
  <food>
    <name>Belgian Waffles</name>
    <price>$5.95</price>
    <description>Two of our famous Belgian
        Waffles with plenty of real maple syrup
        </description>
    <calories>650</calories>
  </food>
  <food>
    <name>Strawberry Belgian Waffles</name>
    <price>$7.95</price>
    <description>Light Belgian waffles covered
        with strawberries and whipped cream
        </description>
    <calories>900</calories>
  </food>
  <food>
    <name>Berry-Berry Belgian Waffles</name>
    <price>$8.95</price>
    <description>Light Belgian waffles covered
        with an assortment of fresh berries and
```

Formatted XML

```
<?xml version="1.0" encoding="UTF-8"?>
<breakfast_menu>
    <food>
        <name>Belgian Waffles</name>
        <price>$5.95</price>
        <description>Two of our famous Belgian
            Waffles with plenty of real maple
            syrup</description>
        <calories>650</calories>
    </food>
    <food>
        <name>Strawberry Belgian Waffles</name>
        <price>$7.95</price>
        <description>Light Belgian waffles
            covered with strawberries and
            whipped cream</description>
        <calories>900</calories>
    </food>
    <food>
        <name>Berry-Berry Belgian Waffles</name
            >
        <price>$8.95</price>
        <description>Light Belgian waffles
```

10-MINUTE BREAK

What are keys & authentication?

- Authentication mechanisms are used to verify the identity of users and applications when accessing APIs.
- An API key is a unique identifier used to authenticate requests to an API, ensuring that the caller has the necessary permissions.
- Authentication is the broader process of verifying that a user or application is who it claims to be, often involving credentials like API keys, tokens, or user passwords to protect access and data
- API Keys are embedded in the scripts researchers use to make calls to any API. They require the researcher to request as access key from the institution or API they are attempting to access and have this prior to running their script.

Google Developer Console

Required: Google Account

Google has many APIs that access different features. We will be creating a project to access information on Places – Points marked on Google maps.

This API Requires a Billing Account – but features can be accessed for free.

I will be generating keys for the account to use in any script – you may use this in the next session or in your own time to generate data.
https://console.cloud.google.com/apis/dashboard

Q & A

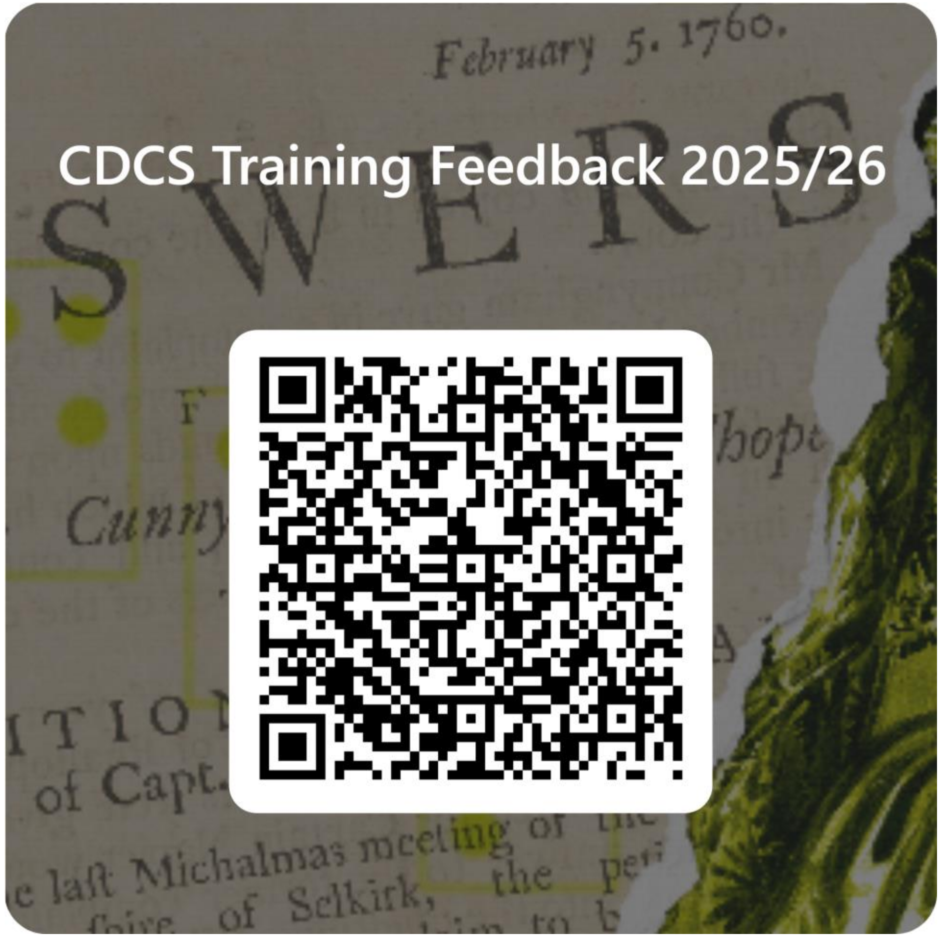
Next session overview

- **To do:** *'Bring'* an API you are interested in and associated documentation to go through next session for the exercise

- During the exercise we encourage you to work in pairs and go through the documentation with tutors stopping by to discuss

- Identify any features useful to you, as well as any bottlenecks you notice.

- **Next week** we will also run through a group practical working with a web API (REST) for the Met Museum.



CDCS Training Feedback 2025/26

API USE FOR RESEARCH

SESSION 2

Housekeeping and itinerary

Session 2 - Day 2 03/12/25

APIs in practice

- [14:00-14:15]: Introduction to session and recap
- [14:15-14:40]: Exercise - working through API documentation
- [14:40-14:50]: Feedback to group on exercise

10-minute break & setup

- [15:00-15-45 +- 5 mins]: Practical with code book (accessing an API)
- [15:45-15:50]: Discussion

Last 10 minutes

- In closing: signpost to resources
- Record attendance & feedback

Recap talk & plans

Exercise 1

1. Which API are you interested in?
2. What service does your chosen API offer?
3. Group together and go through some of key features listed in the documentation for API access.
4. Is it open or requires authentication?
5. Discuss and lay out the core features you would need

Quick feedback from the group exercise

1. Any bottlenecks
2. Is the API suitable for your use case?
3. Any questions?

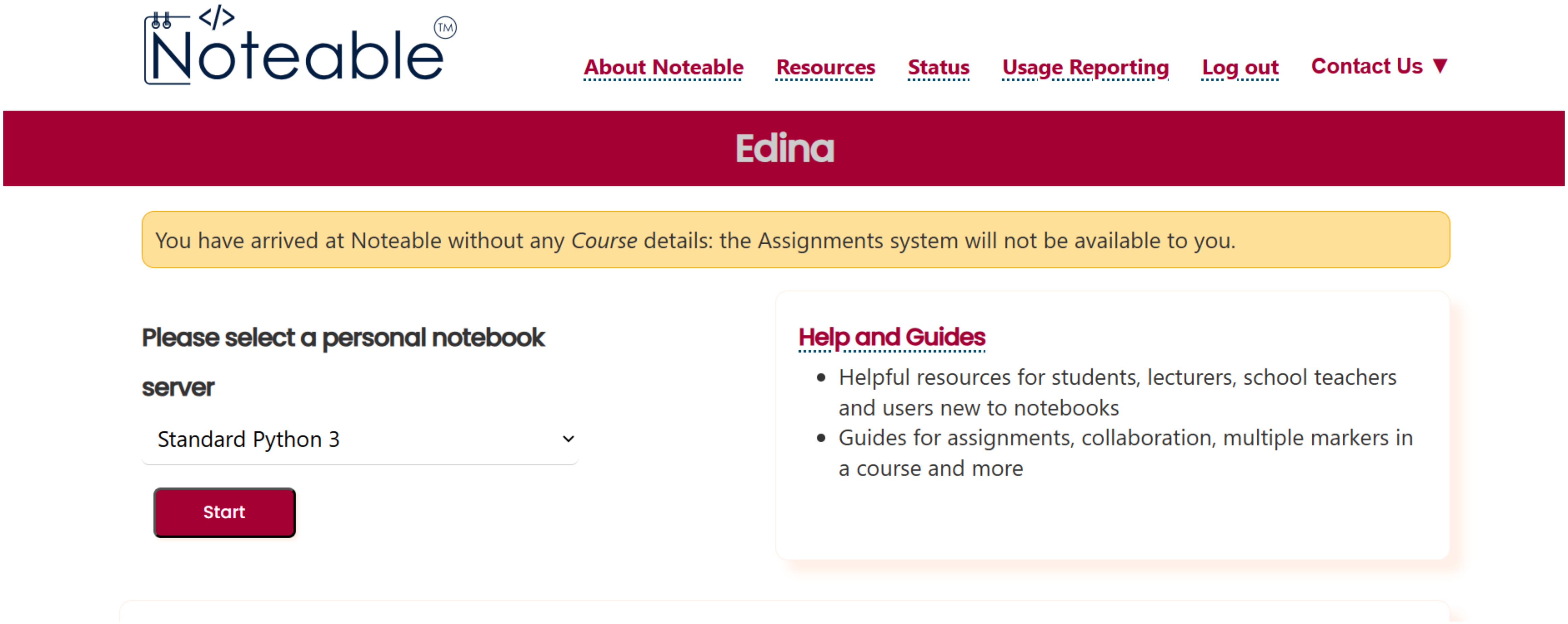
10 – MINUTE BREAK
(practical setup)

Practical exercise

Programmatically accessing a web API to collect data

- Groups (4)
- Setup (4 notebooks open)
- Guided walk through with Somya (notebook on screen)

GITHUB repository



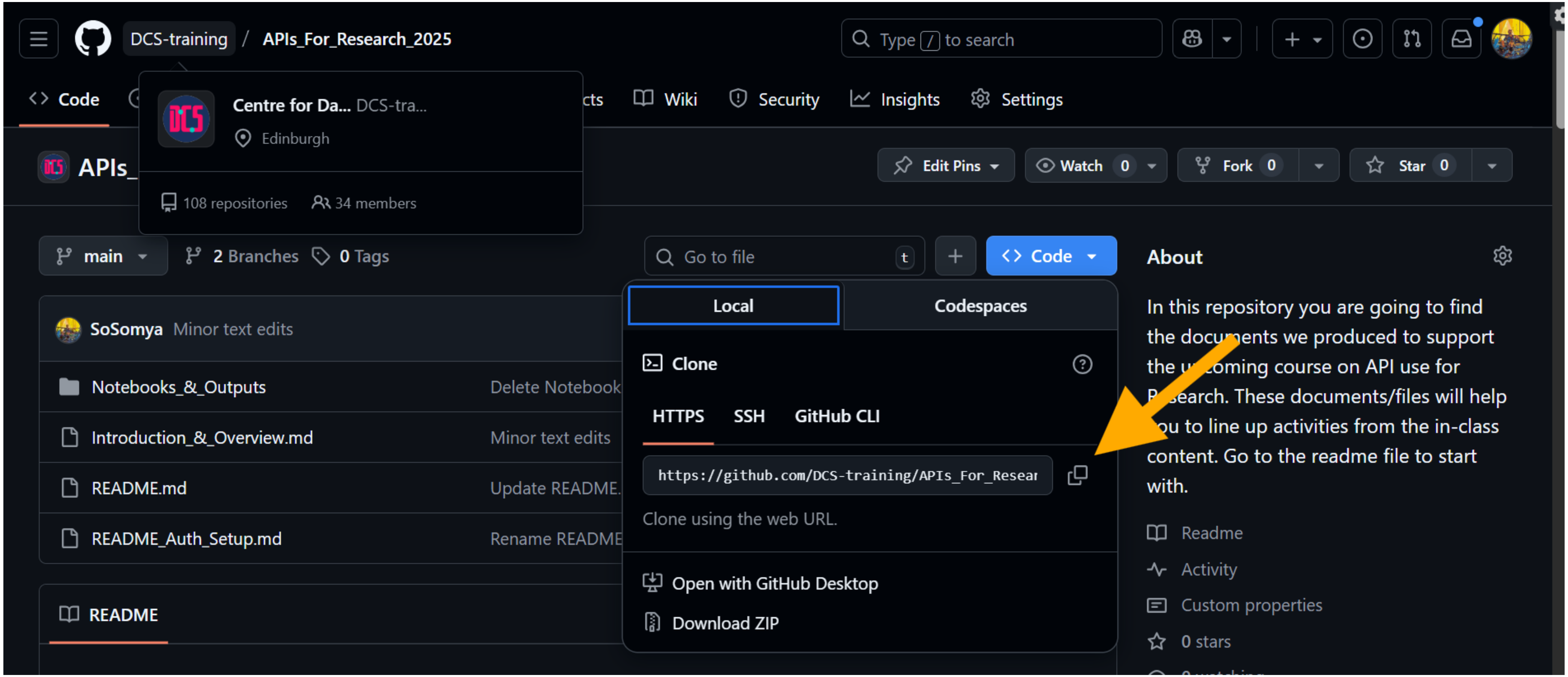
You have arrived at Noteable without any *Course* details: the Assignments system will not be available to you.

Please select a personal notebook server

Standard Python 3

Help and Guides

- Helpful resources for students, lecturers, school teachers and users new to notebooks
- Guides for assignments, collaboration, multiple markers in a course and more



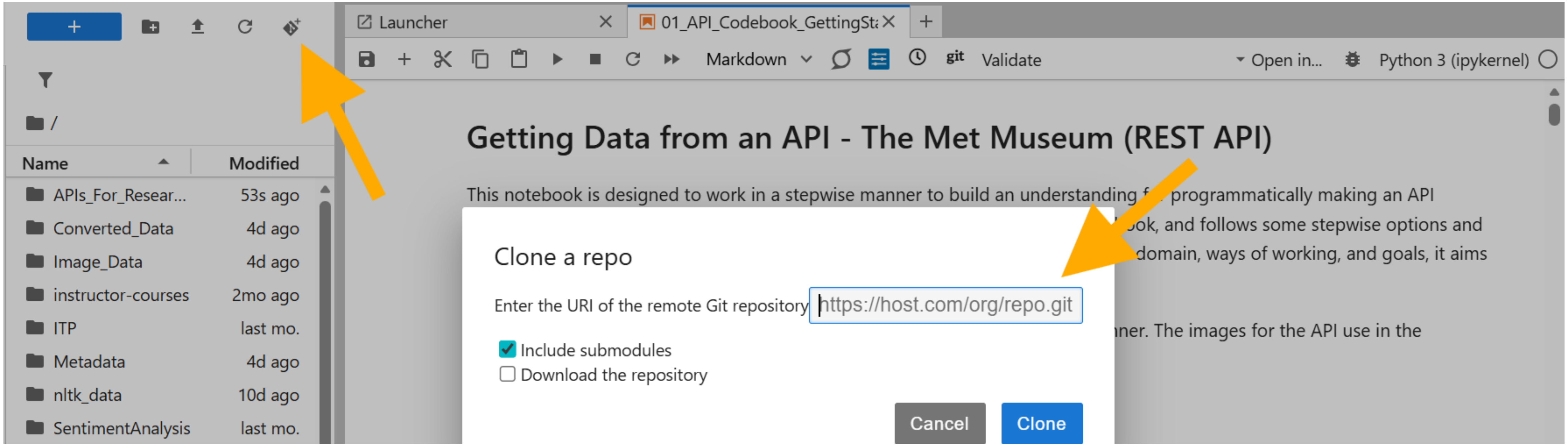
SoSomya Minor text edits

Notebooks_&_Outputs	Delete Notebook
Introduction_&_Overview.md	Minor text edits
README.md	Update README.
README_Auth_Setup.md	Rename README

```
https://github.com/DCS-training/APIs_For_Resear
```

Clone using the web URL.

In this repository you are going to find the documents we produced to support the upcoming course on API use for Research. These documents/files will help you to line up activities from the in-class content. Go to the readme file to start with.



Name	Modified
APIs_For_Resear...	53s ago
Converted_Data	4d ago
Image_Data	4d ago
instructor-courses	2mo ago
ITP	last mo.
Metadata	4d ago
nltk_data	10d ago
SentimentAnalysis	last mo.

Getting Data from an API - The Met Museum (REST API)

This notebook is designed to work in a stepwise manner to build an understanding f programmatically making an API ook, and follows some stepwise options and domain, ways of working, and goals, it aims

ner. The images for the API use in the

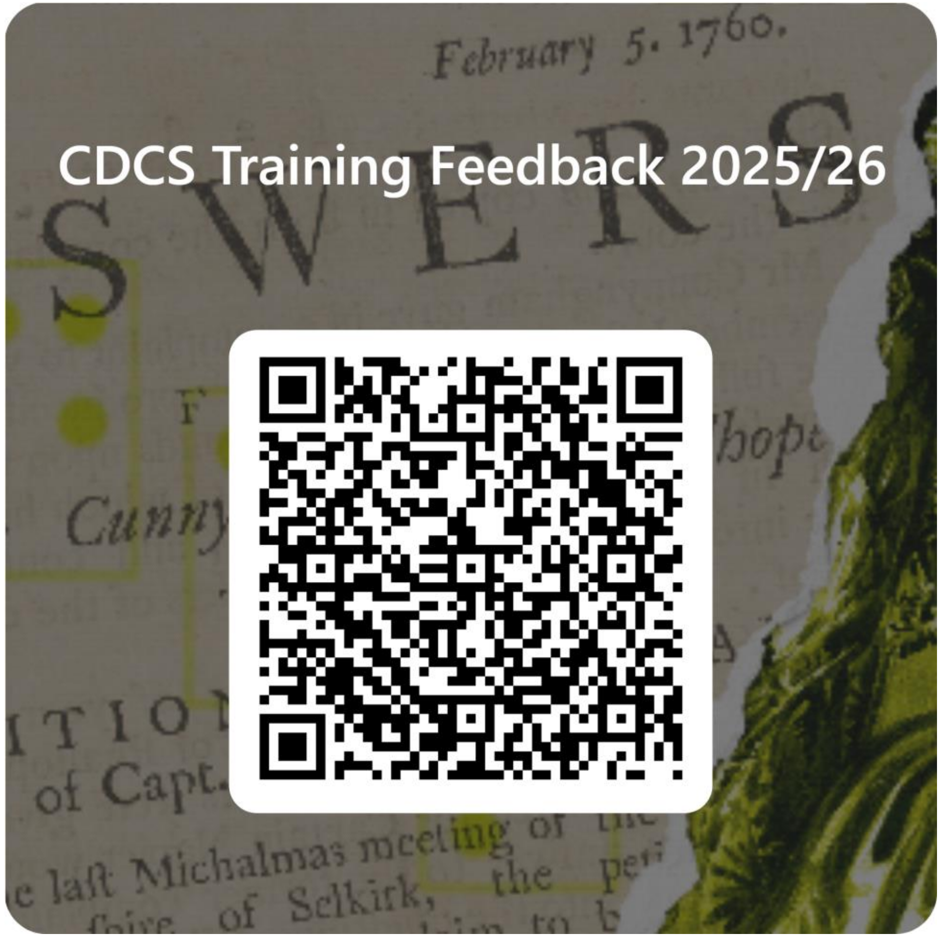
Clone a repo

Enter the URI of the remote Git repository https://host.com/org/repo.git

☑ Include submodules

☐ Download the repository

Q & A



CDCS Training Feedback 2025/26