Introduction to the course and final project specifications

Ivan Heibi
ivan.heibi2@unibo.it – https://orcid.org/0000-0001-5366-5194 – https://ivanhb.it

Computational Management of Data – Part II (A.Y. 2025/2026)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

Course organisation

Computational Management of Data is a 12 ECTS course organised in two parts:

- Part I (October-December, 6 ECTS): dedicated to explore the basics of **computational thinking and programming** - taught by Silvio Peroni
- Part II (February-March, 6 ECTS): dedicated to deep in topics related to **data management and databases** - taught by Ivan Heibi (me!)

Interacting

- **GitHub** repository (https://github.com/comp-data/2025-2026) of the course for a series of activities, such as exercises and raising issues
- We use a the **Discord** discussion group for communicating with each other (you should already have it!)
    - the private channel dedicated to the course is the same one: *comp-data-25-26*

Books and material

- Think and Compute: a Primer for Digital Humanists (official book of the course) + accompanying book How To Code in Python
  The course book parts of this module are:
  - **Part 5: Managing Data in Python**
  - **Part 6: Databases**
- All the material (including slides) is available in the GitHub repository of the course at https://github.com/comp-data/2025-2026/
- The slides are provided as Google Slides, links will be provided in the Github repository

* Have you found a mistake in the official book or into the slides? Please add a GitHub issue

Goal of Part II

- **Part I Goal:** how to use a language to communicate with and instruct an information-processing agent – using Python (as programming language)

- **Part II Goal:** to build on the concepts introduced in Part I by focusing on such aspects through a practical approach to data management

Rules

- Links to the free text book are provided, along with additional material – all can be found on the GitHub repository of the course
- If you cannot attend, do not attend (even if attendance is recommended) – On Virtuale you will find links to the recorded lessons (on Panopto)
- At least six exam sessions per academic year and will take place on 26 May 2026, 25 June 2026, 23 July 2026, 10 September 2026, 5 November 2026, and 29 January 2027.
- Lessons (usually) start 10 minutes late, include a 10-minute break, and finish 10 minutes early

Part II organisation

- ***Theoretical*** (15 hours) – 8 lectures of two hours each (today included) – where I provide a theoretical introduction about the specific topic computer not necessary but you can bring it with you of course
- ***Hands-on sessions*** (15 hours) – 7 lectures of two hours each – where I conduct a practical session based on the materials presented in Parts 5 and 6 of the book (https://thinkcompute.github.io), as well as on existing tools that enable hands-on experimentation with the topics introduced in the theoretical lectures – a computer is required.
- **Tutor sessions (to be agreed with the tutor)** – where Arcangelo Massari recalls some of the topics introduced in the course according to your specific needs

Calendar

Mondays and Wednesdays 9:00-11:00, Fridays 12:00-14:00, theoretical lectures and hands-on sessions

9/2/26	Introduction to the course and final project specifications
11/2/26	What is a datum and how it can be represented computationally
13/2/26	Data formats and methods for storing data in Python
16/2/26	Introduction to data modelling
18/2/26	Implementation of data models via Python classes
20/2/26	Processing and querying the data
23/2/26	Introduction to Pandas
25/2/26	Describing and visualising data
27/2/26	Descriptive statistics and graphs about data using Pandas

Calendar

2/3/26	Database Management Systems
4/3/26	Configuring and populating a relational database
6/3/26	SQL, a query language for relational databases
11/3/26	Configuring and populating a graph database
13/3/26	SPARQL, a query language for RDF databases
18/3/26	Interacting with databases using Pandas

JupyterLab for hands-on sessions

JupyterLab (https://jupyter.org/) is a web-based interactive development environment for notebooks, code, and data

You need to have Python installed to running Jupyter. To install JupyterLab, you can use the command pip as follows:

```
pip install jupyterlab
```

It will be used for all the hands-on sessions, thus be sure to have it available on your computer, and the website includes a very good documentation, while a lot of tutorials are available online

Exam

Part I

- A written examination (duration: one hour and an half) about the Part I content (max. score: 32)
- A non-mandatory workshop (the day after the last lecture of Part I, i.e. 18 December 2025, 13:00-16:00), where the students are asked to organise themself in groups of 3-4 people (max. score: 3)

Part II

- **The implementation of a group project about the Part II content, where students are mandatorily asked to organise themself in groups of 3-4 people (max. score: 32)**

Exam – implementation of a project

- The exam consists of the **implementation of a project –** yes! you have to write a software, and the specifications will be introduced today
- an **oral colloquium** on the project implemented, for assessing the contribution of each student
- It will be assigned maximum 16 points for the correctness of the project – I will run a series of tests aiming at assessing all the code developed
- The points above are assigned to all the students that have worked to the project
- Other 16 points (maximum) will be assigned to each individual student as result of the oral colloquium

Evaluation of the course

The last lectures of the course, you will be asked to fill-up a questionnaire on the organisation of the course and related stuff - it is anonymous, of course

Please, do it carefully and honestly, since it is one of the most important inputs I will have to understand what can be improved in the next year course

it is crucial to have your feedback in order to understand how to improve it for the next year

The Project

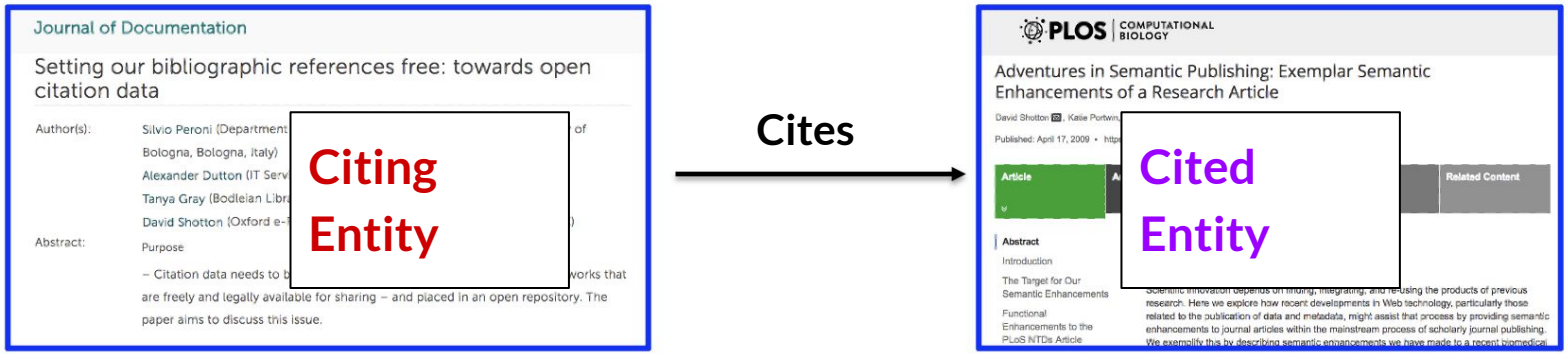
Goal of the project

Develop a software that enables one:

- to process data stored in different formats and to upload them into two distinct databases (i.e. *Relational* and *Graph)*
- to query these databases simultaneously according to predefined operations

Information about the project can be found in the GitHub repository of the course (https://github.com/comp-data/2025-2026) (and will be updated in due course)

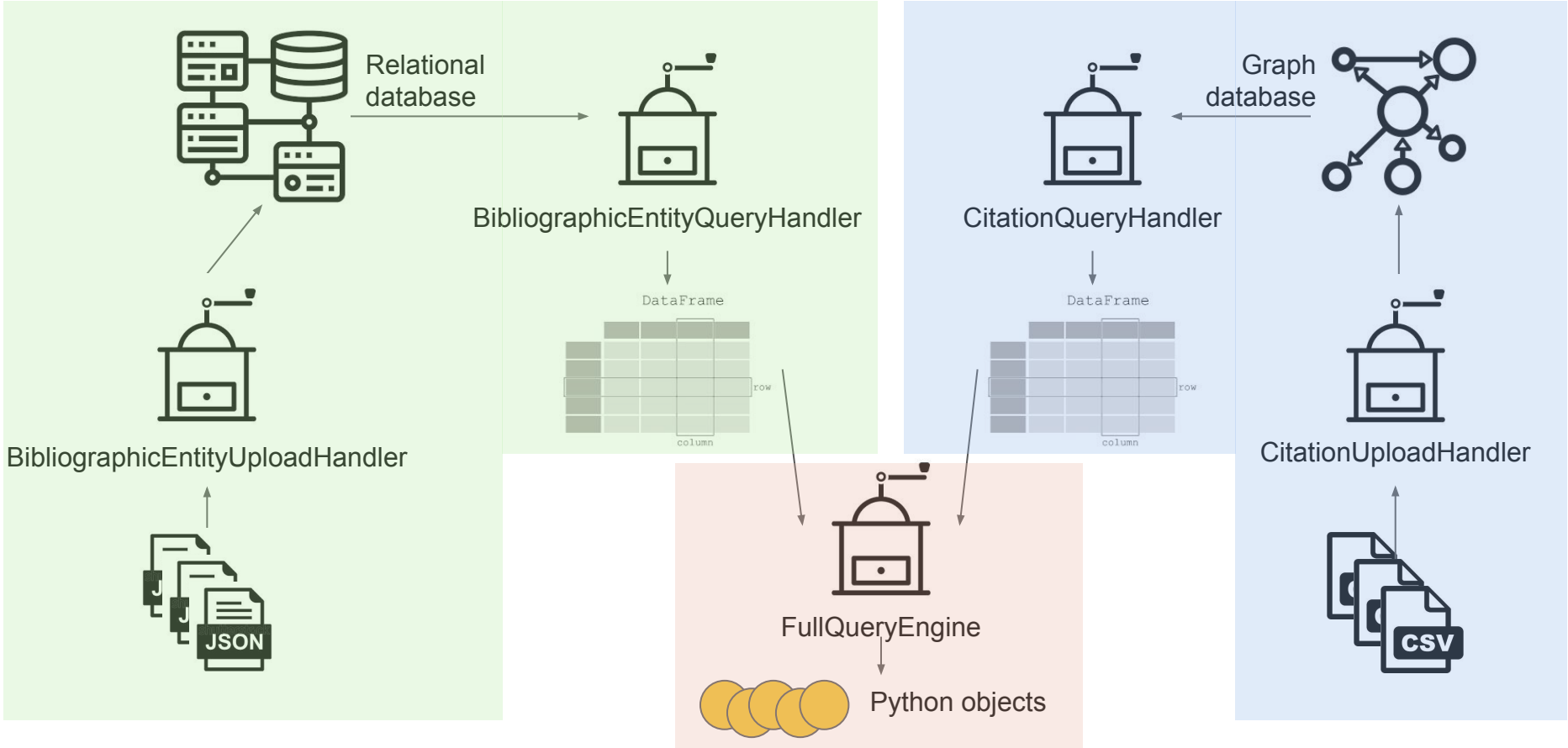
What is it about? … Citations



Citation attributes:

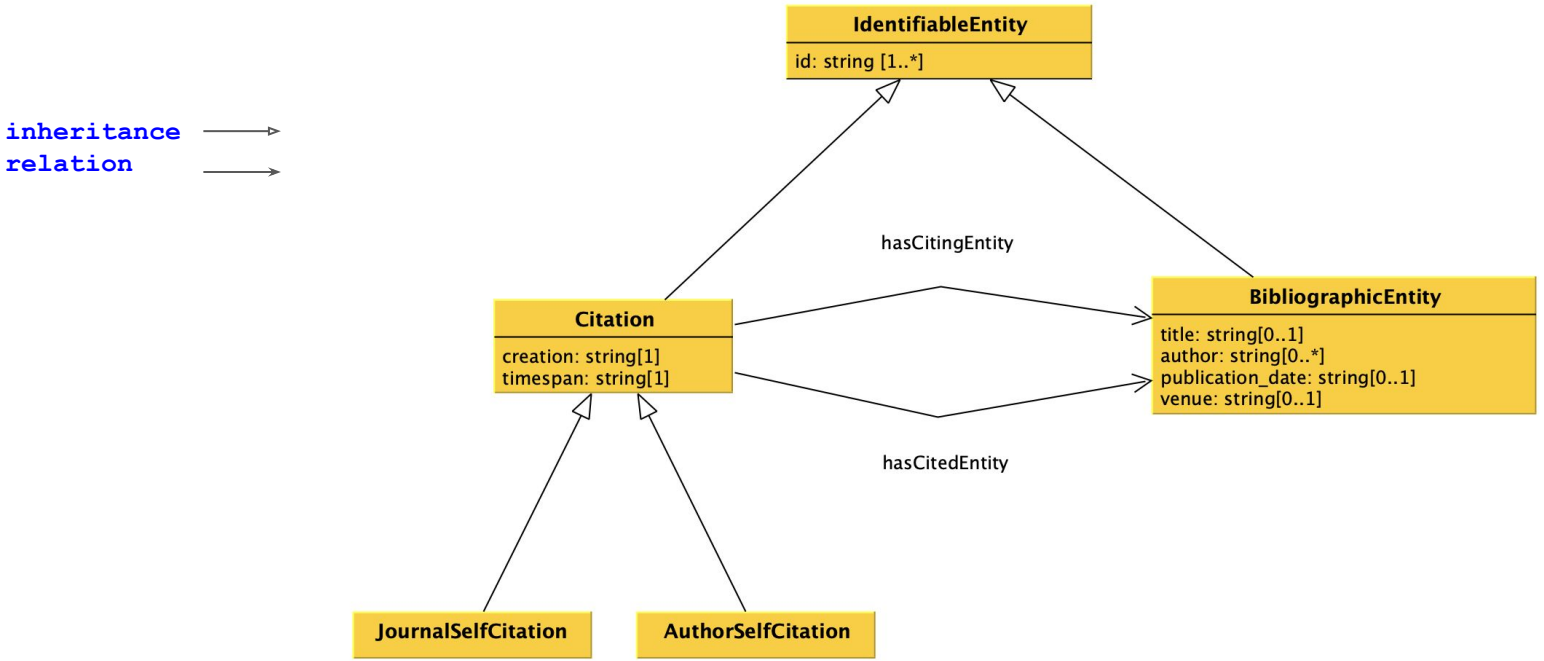
- **Creation**
- **Timespan**
- **Is Author Self Citation?**
- **Is Journal Self Citation?**

Project Workflow

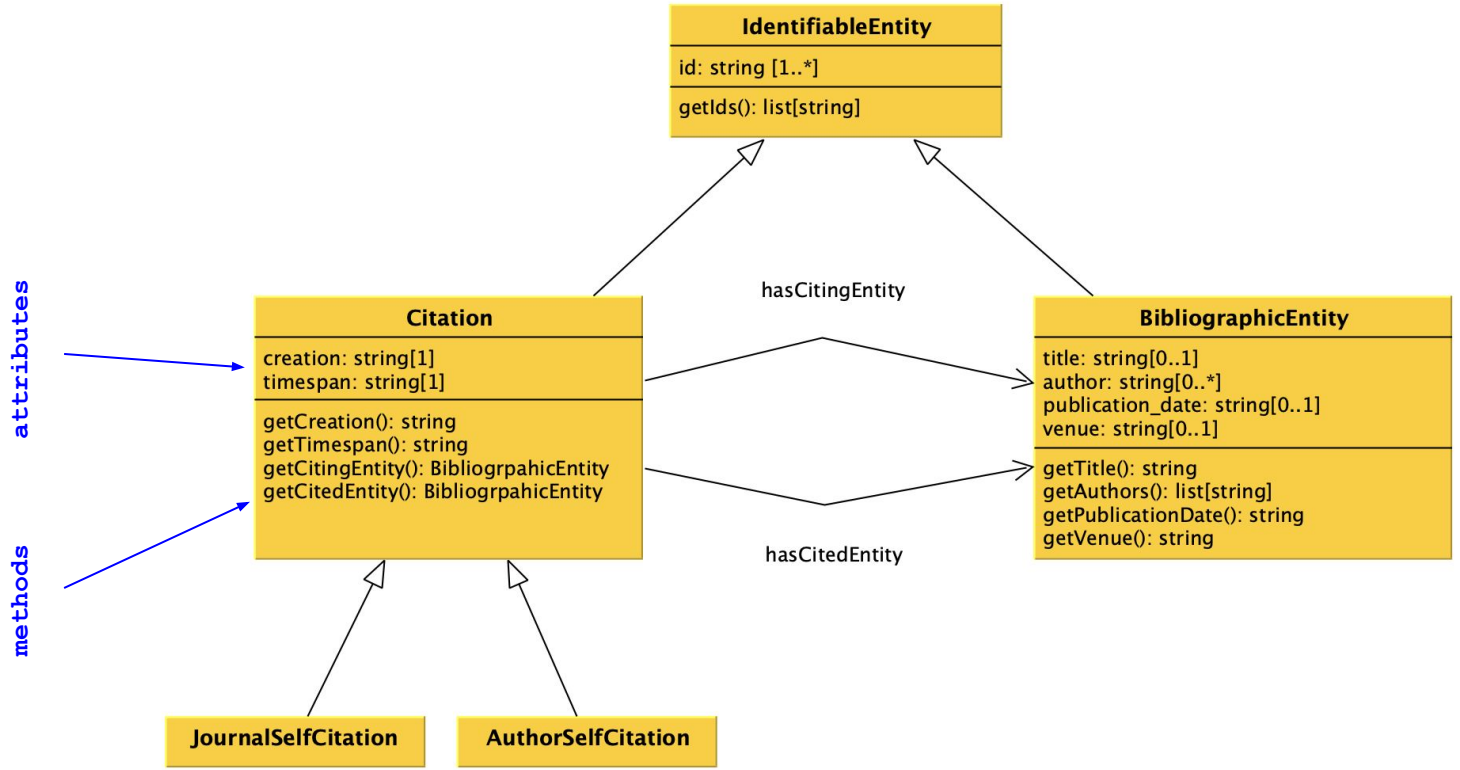


Data Model

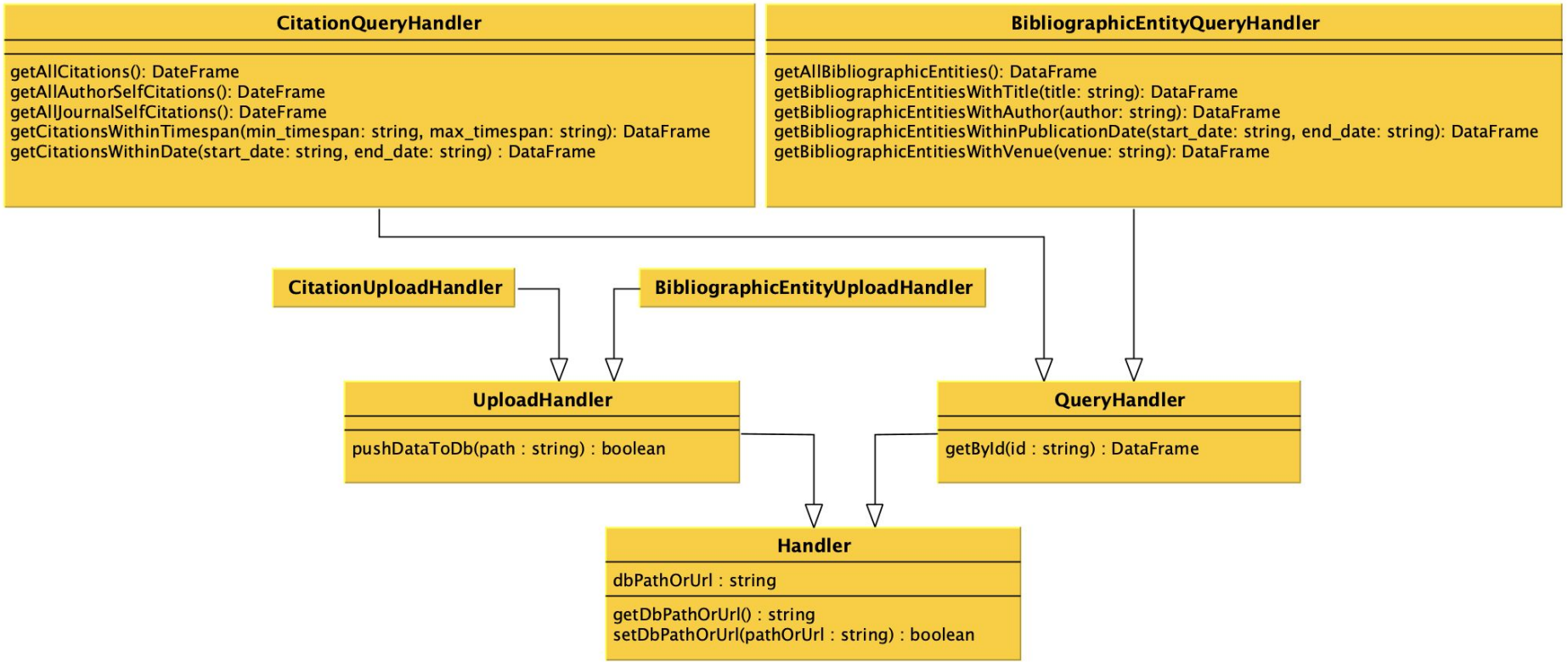
The data model for the entities under consideration is loosely inspired by metadata describing citations and bibliographic entities within the OpenCitations Infrastructure



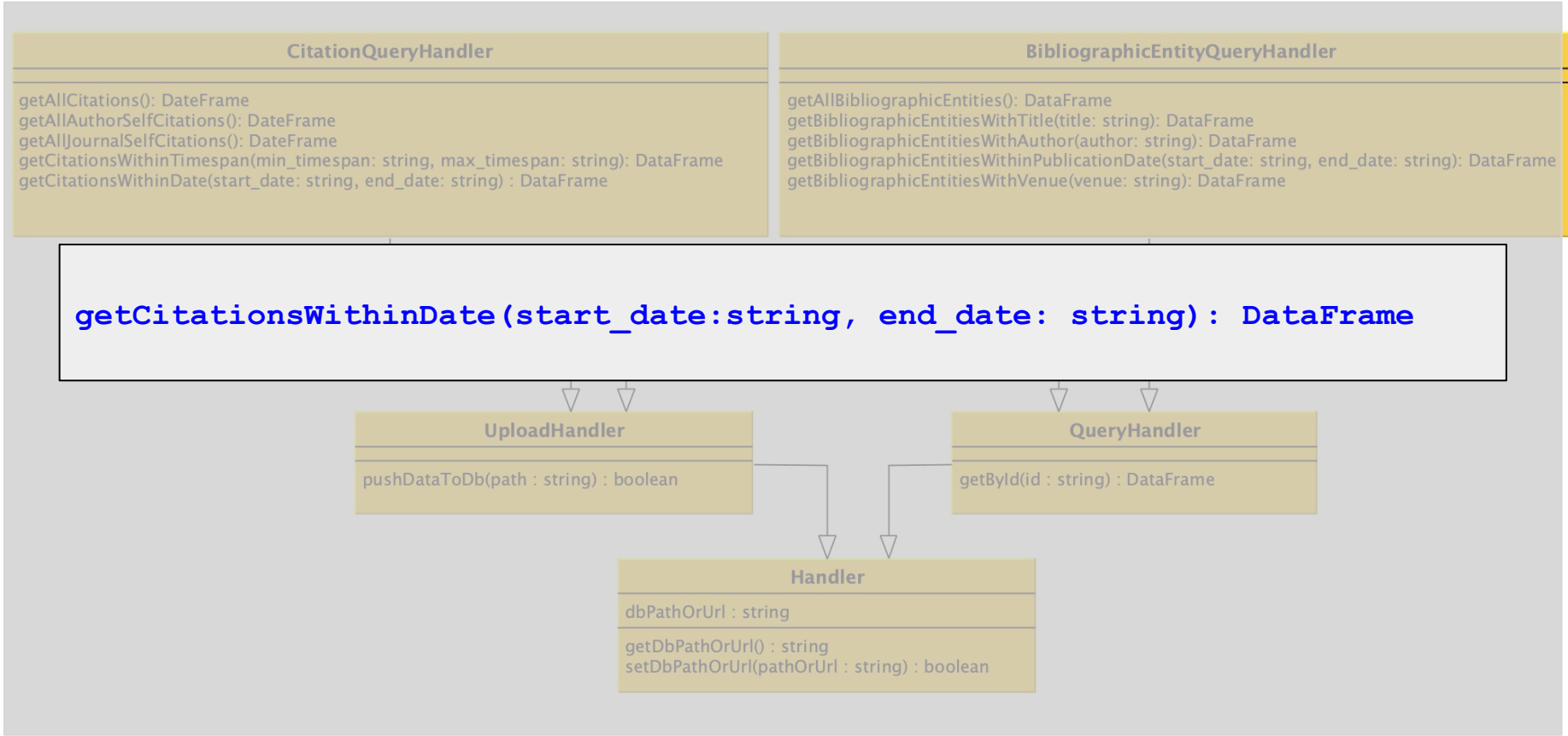
Software: UML diagram (1)



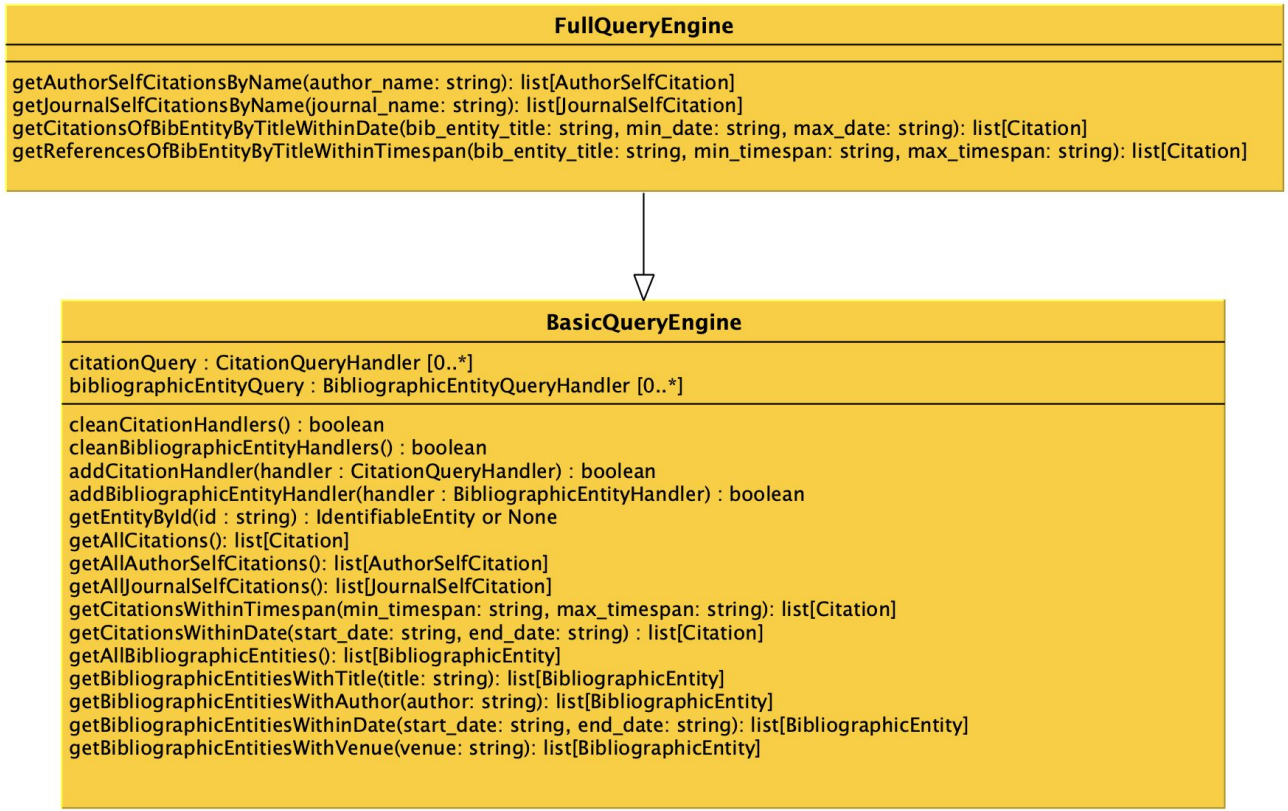
Software: UML diagram (2)



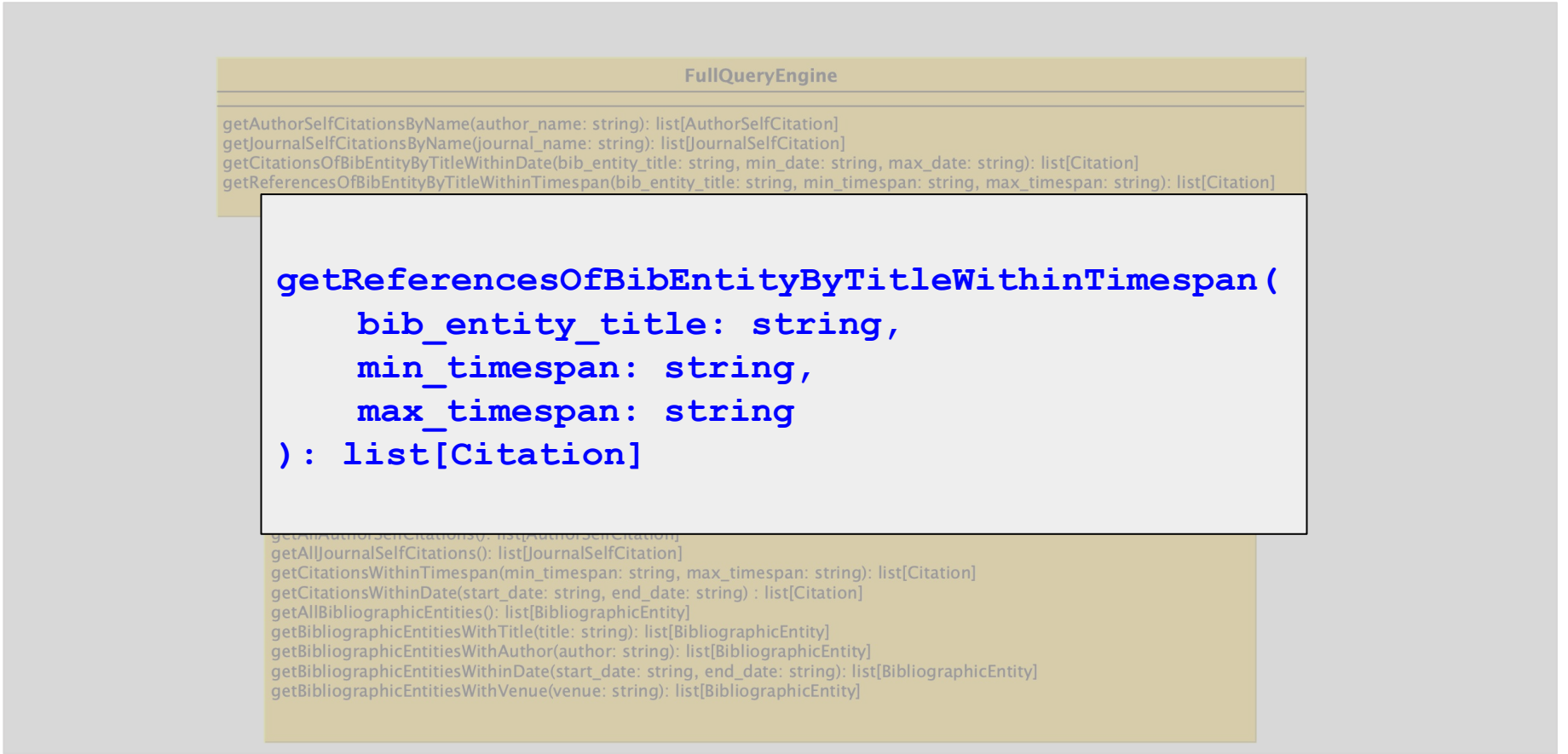
Software: UML diagram (2)



Software: UML diagram (3)



Software: UML diagram (3)



Goal of the project

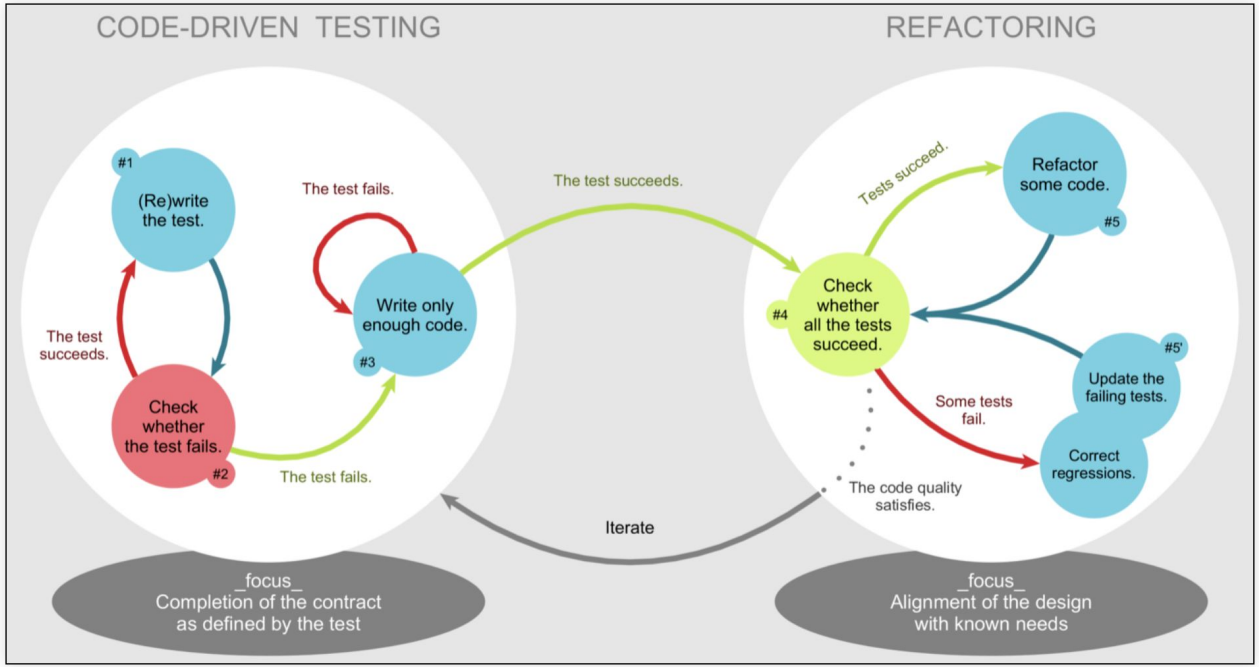
You have to provide all Python files implementing your project, by sharing them in some way (e.g. via OneDrive)

You have to send all the files three days before the exam session you want to take

Before submitting the project, you must be sure that your code passes a particular basic test (that will be provided) which aims at checking if the code is runnable and compliant with the specification

The test-driven development (TDD)

You have already seen the test-driven development (TDD) – I strongly suggest you to systematically adopt this development technique for developing your code



Python tests

There is at least a Python library available which has been entirely developed to facilitate the creation of tests, i.e. unittest

Online, you can find several documents describing how to create tests in Python, such as one at the Real Python website (https://realpython.com/python-testing/)

END

Ivan Heibi
ivan.heibi2@unibo.it – https://orcid.org/0000-0001-5366-5194 – https://ivanhb.it

Computational Management of Data – Part II (A.Y. 2025/2026)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna