People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of JIJEL

Practical Implementation of FOSS in Bioinformatics: A Focus on Biopython and GitHub

The students:

-Rekrouk Samah

-Hamidouche Chiraz

-Rekrouk Djihane

The teacher:

-Dr Bensalem Amel

2026-2027

Part 1 - Theoretical Study of the Tool: Biopython

►Introduction:

In the field of modern bioinformatics, the processing and analysis of biological data require efficient and flexible computational tools. Several libraries have been developed to facilitate these tasks, among which **Biopython** occupies an important position. This study aims to present Biopython, its main functionalities, technical aspects, as well as its strengths and limitations.

1. General Presentation of Biopython:

Biopython is a long-standing, mature open-source project developed by an international collaboration of volunteer developers. Founded in **1999**, it was created to provide a comprehensive set of libraries for the Python programming language to address diverse problems in bioinformatics and computational molecular biology. The project is hosted and supported by the Open Bioinformatics Foundation (OBF) and is compatible with all major operating systems **(Cock et al., 2009)**.

2. Main Functionalities:

Biopython acts as a massive collection of modules that researchers can use in their own scripts or software. Its core functionalities include sequence manipulation through the Seq object, support for various biological file formats such as FASTA and GenBank, interaction with online databases like NCBI Entrez and KEGG, integration of tools such as BLAST and ClustalW, structural biology analysis, motif analysis, advanced statistical methods, and genomic data visualization **(Cock et al., 2009)**.

3. Technical Aspects:

Biopython is based on the Python programming language, known for its easy-to-learn syntax and object-oriented programming capabilities. The Seq object supports biological operations and is extended by SeqRecord and SeqFeature classes. Although Python is a high-level language, Biopython can interface with optimized code and integrates with NumPy

for numerical computations. It also allows data storage through BioSQL **(Cock et al., 2009)**.

4. Strengths:

The strengths of **Biopython** lie in its extensive functional coverage, interoperability with other bioinformatics tools, comprehensive documentation, active community support, workflow integration, and free cross-platform availability **(Cock et al., 2009)**.

5. Limitations and Weaknesses:

Biopython requires Python programming skills and does not provide a graphical user interface for non-programmers. For highly demanding computational tasks, additional libraries or lower-level languages are needed. Due to the vast scope of the library, users must invest time in exploring the documentation to fully utilize its potential **(Cock et al., 2009)**.

6. Conclusion:

Biopython has established itself as an essential tool for the development of bioinformatics software and routine biological data analysis. Its ability to integrate external tools, query global databases, and manipulate complex biological structures makes it a cornerstone of modern computational biology.

Part 2 – Practical Study: Exploration of Zenodo.

1. Presentation of Zenodo.

►General Overview:

Zenodo is a global open-access repository designed to enable researchers, scientists, and scholars to share and preserve their research outputs. It was launched in **2013** through a collaboration between **CERN** (the European Organization for Nuclear Research) and the **OpenAIRE** project, funded by the European Commission **(Zenodo, n.d.)**. Unlike many repositories, **Zenodo** is "catch-all," meaning it is open to all fields of science and all types of digital artifacts.

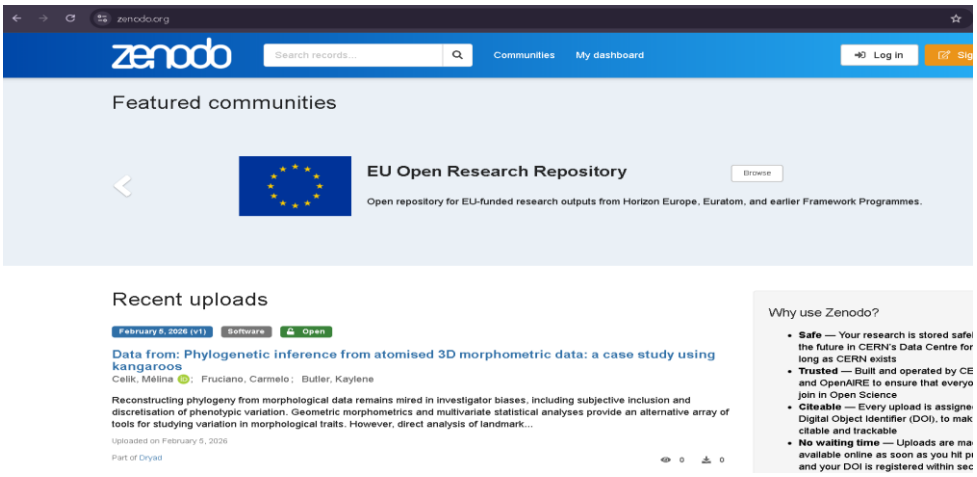


Figure : zenodo platform.

►Platform Objectives:

The primary mission of **Zenodo** is to support the transition toward **Open Science** by providing:

- **Persistent Identification:** Assigning a Digital Object Identifier (DOI) to every upload to make the work citable and trackable **(CERN, 2023)**.
- **Accessibility:** Ensuring research data is free to access for anyone, anywhere, immediately upon publication.
- **Sustainability:** Leveraging CERN's high-performance IT infrastructure to guarantee the long-term storage of data for at least 20 years **(European Commission, 2021)**.

►**Types of Hosted Content:**

Zenodo accepts a wide variety of research-related digital objects, including:

- **Datasets:** Raw or processed data from experiments.
- **Publications:** Peer-reviewed articles, preprints, and conference papers.
- **Software:** Code and scripts (often integrated with GitHub).
- **Multimedia:** Images, videos, and posters related to scientific dissemination.

►**Importance for Open Science and NLS (Natural Life Sciences)**

In the field of **Natural Life Sciences (NLS)**, **Zenodo** plays a critical role in the "FAIR" data principles (Findable, Accessible, Interoperable, and Reusable). For instance, sharing genomic sequences or cellular imaging data on **Zenodo** allows for:

1. **Transparency:** Allowing other researchers to verify findings.
2. **Collaboration:** Enabling scientists to build upon existing datasets without repeating expensive experiments **(OpenAIRE, 2022)**.

2. Description of the steps carried out.

1. Visit the **Zenodo** platform (https://zenodo.org/)

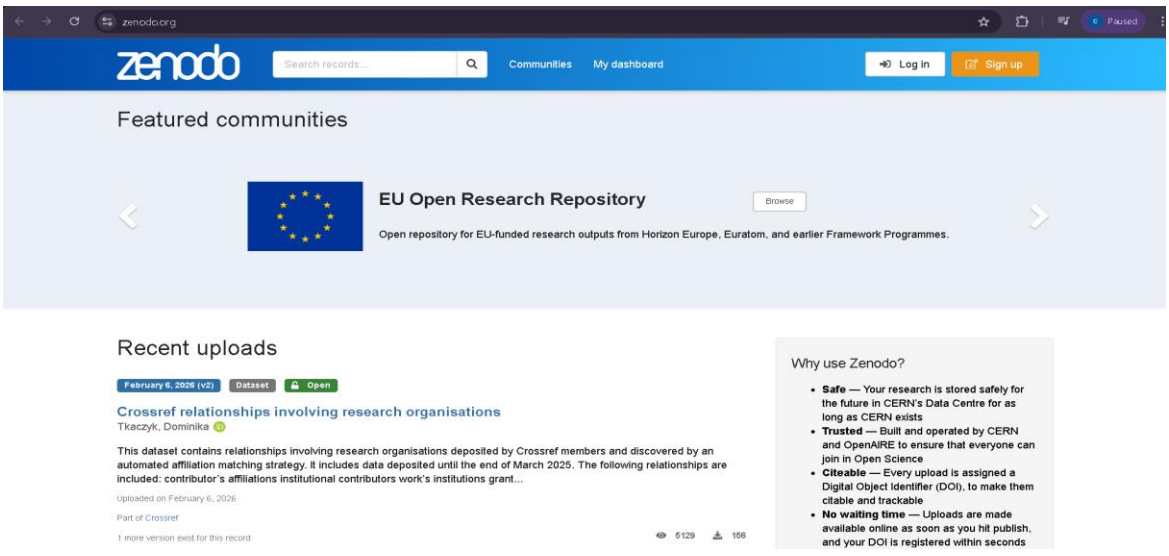


Figure 1 : Zenodo Platform.

2.Search for a dataset using a query contain keyword: **genome**

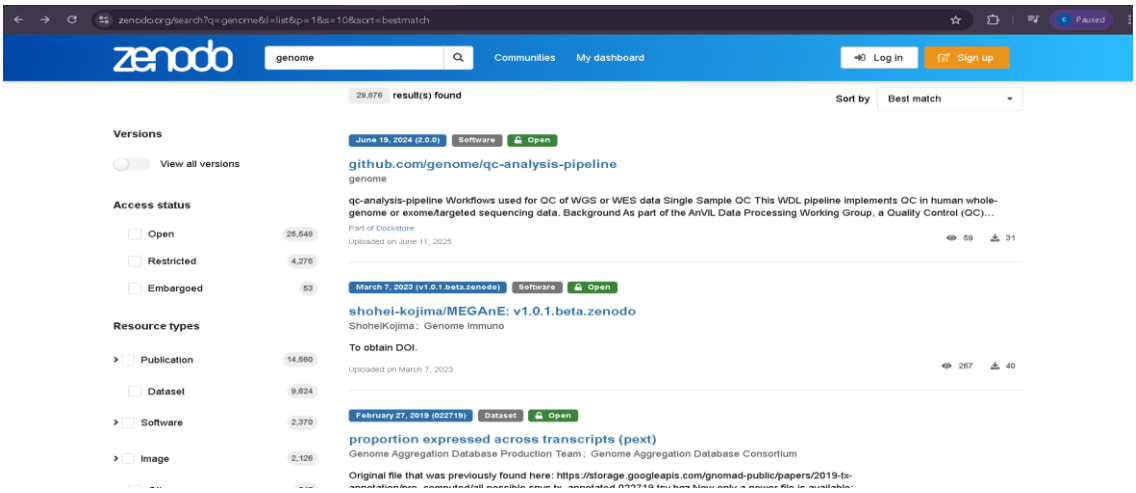


Figure 2: Search results for the query 'Genome' on **Zenodo**.

3. Select a relevant dataset.

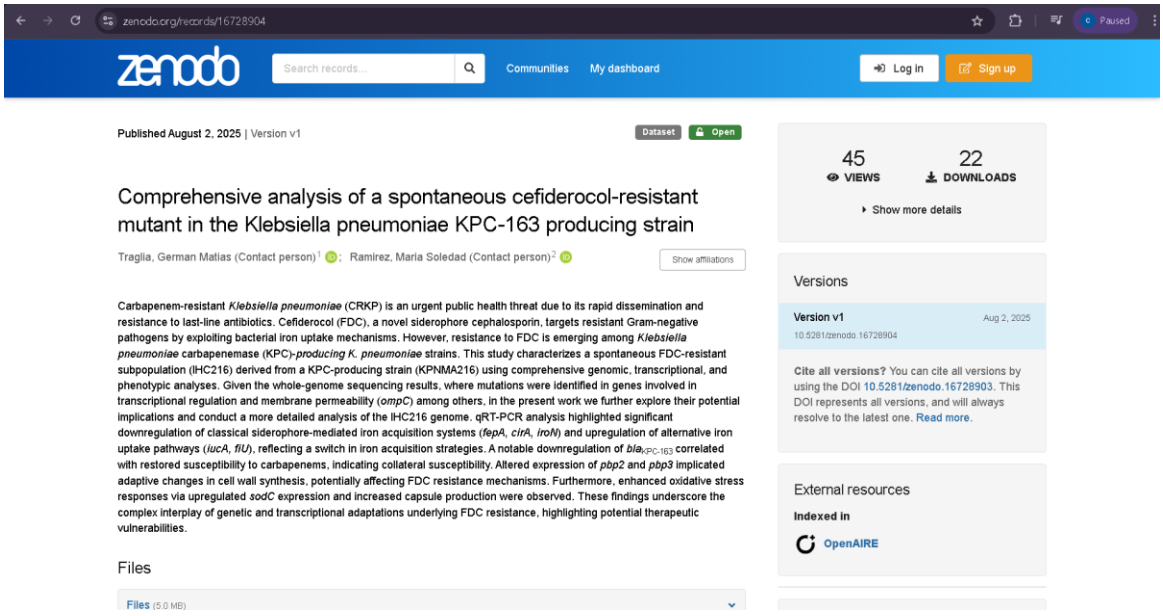


Figure 3: Detailed view of the selected dataset. The record displays the title, authors, and the unique DOI assigned to the research on antibiotic resistance in *K. pneumoniae*.

4. Download the selected dataset.

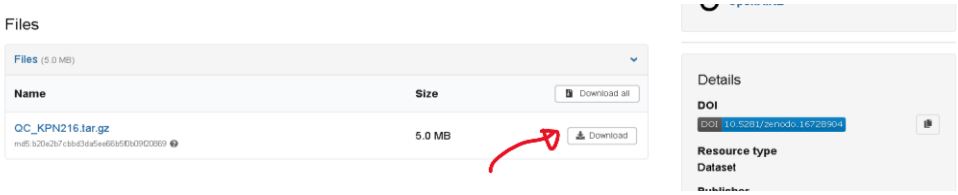
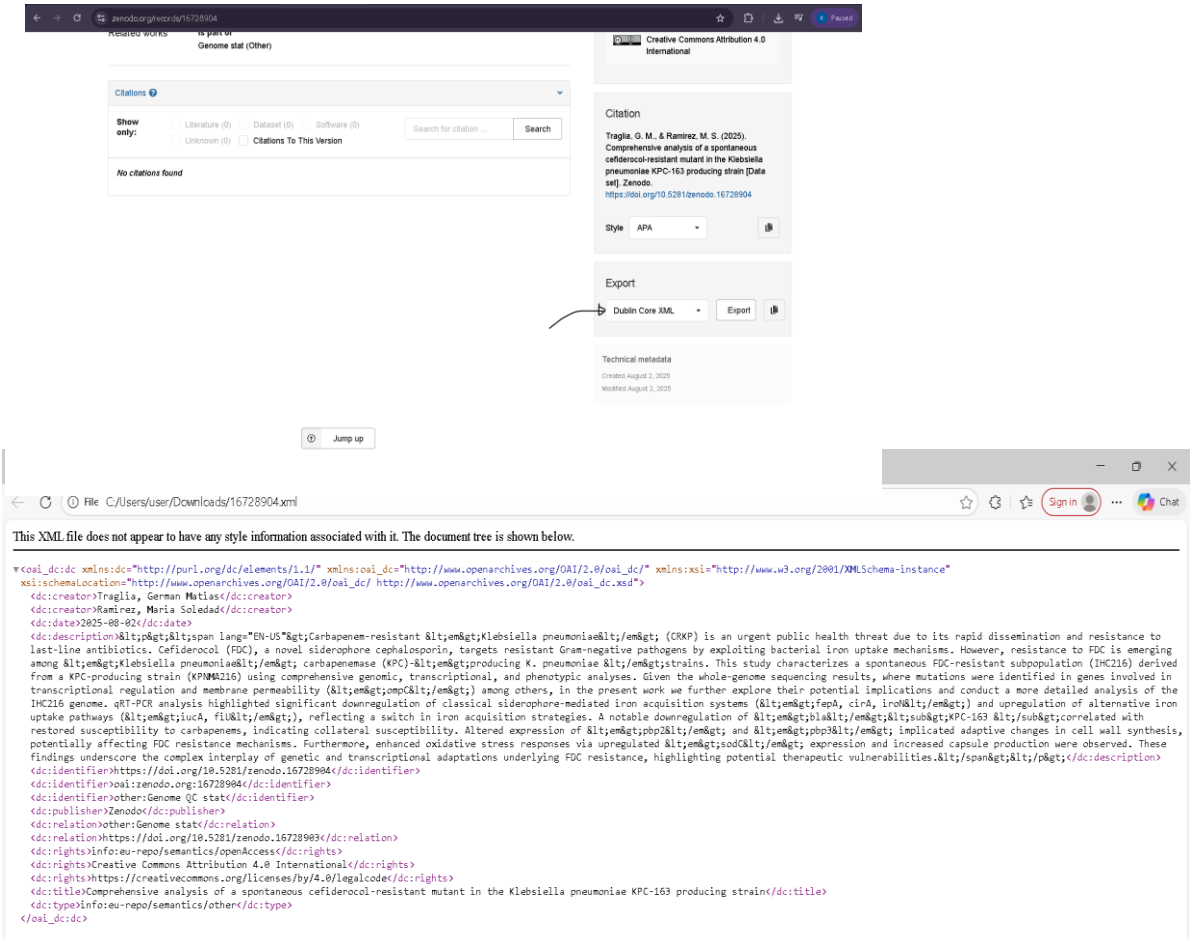


Figure 4: Downloading the research files for local analysis. The platform provides direct access to open-source data according to Open Science principles.

5. Retrieve the dataset metadata using the **"Dublin Core"** Standard.




```
<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>Traglia, German Matias</dc:creator>
  <dc:creator>Ramirez, Maria SoLedad</dc:creator>
  <dc:date>2025-08-02</dc:date>
  <dc:description>&lt;p&gt;&lt;span lang="EN-US"&gt;Carbapenem-resistant &lt;em&gt;Klebsiella pneumoniae&lt;/em&gt; (CRKP) is an urgent public health threat due to its rapid dissemination and resistance to
  last-line antibiotics. Cefiderocol (FDC), a novel siderophore cephalosporin, targets resistant Gram-negative pathogens by exploiting bacterial iron uptake mechanisms. However, resistance to FDC is emerging
  among &lt;em&gt;Klebsiella pneumoniae&lt;/em&gt; carbapenemase (KPC)-&lt;em&gt;producing K. pneumoniae &lt;/em&gt;strains. This study characterizes a spontaneous FDC-resistant subpopulation (IHC216) derived
  from a KPC-producing strain (KPNMA216) using comprehensive genomic, transcriptional, and phenotypic analyses. Given the whole-genome sequencing results, where mutations were identified in genes involved in
  transcriptional regulation and membrane permeability (&lt;em&gt;ompC&lt;/em&gt;) among others, in the present work we further explore their potential implications and conduct a more detailed analysis of the
  IHC216 genome. qRT-PCR analysis highlighted significant downregulation of classical siderophore-mediated iron acquisition systems (&lt;em&gt;fepA, cirA, iroN&lt;/em&gt;) and upregulation of alternative iron
  uptake pathways (&lt;em&gt;iucA, fiU&lt;/em&gt;), reflecting a switch in iron acquisition strategies. A notable downregulation of &lt;em&gt;bla&lt;/em&gt;&lt;sub&gt;KPC-163 &lt;/sub&gt;correlated with
  restored susceptibility to carbapenems, indicating collateral susceptibility. Altered expression of &lt;em&gt;pbp2&lt;/em&gt; and &lt;em&gt;pbp3&lt;/em&gt; implicated adaptive changes in cell wall synthesis,
  potentially affecting FDC resistance mechanisms. Furthermore, enhanced oxidative stress responses via upregulated &lt;em&gt;sodC&lt;/em&gt; expression and increased capsule production were observed. These
  findings underscore the complex interplay of genetic and transcriptional adaptations underlying FDC resistance, highlighting potential therapeutic vulnerabilities.&lt;/span&gt;&lt;/p&gt;</dc:description>
  <dc:identifier>https://doi.org/10.5281/zenodo.16728904</dc:identifier>
  <dc:identifier>oai:zenodo.org:16728904</dc:identifier>
  <dc:identifier>other:Genome QC stat</dc:identifier>
  <dc:publisher>Zenodo</dc:publisher>
  <dc:relation>other:Genome stat</dc:relation>
  <dc:relation>https://doi.org/10.5281/zenodo.16728903</dc:relation>
  <dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
  <dc:rights>Creative Commons Attribution 4.0 International</dc:rights>
  <dc:rights>https://creativecommons.org/licenses/by/4.0/legalcode</dc:rights>
  <dc:title>Comprehensive analysis of a spontaneous cefiderocol-resistant mutant in the Klebsiella pneumoniae KPC-163 producing strain</dc:title>
  <dc:type>info:eu-repo/semantics/other</dc:type>
</oai_dc:dc>
```


Figure 5: Exporting the dataset metadata in Dublin Core (XML format). This step demonstrates the interoperability of the platform and the standardized description of the research components.

6. Extract and present as much information as possible.

Dublin Core Era	Information
Title	● Comprehensive analysis of a spontaneous cefiderocol-resistant mutant in the Klebsiella pneumoniae KPC-163 producing strain
Creator	1. Traglia, German Matias 2. Ramirez, Maria Soledad
Date	● 2025-08-02
Langue	● EN-US
Description	● Carbapenem-resistant *Klebsiella pneumoniae* (CRKP) is an urgent public health threat due to its rapid dissemination and resistance to last-line antibiotics. Cefiderocol (FDC), a novel siderophore cephalosporin, targets resistant Gram-negative pathogens by exploiting bacterial iron uptake mechanisms. However, resistance to FDC is emerging among *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* strains. This study characterizes a spontaneous FDC-resistant subpopulation (IHC216) derived from a KPC-producing strain (KPNMA216) using comprehensive genomic, transcriptional, and phenotypic analyses. Given the whole-genome sequencing results, where mutations were identified in genes involved in transcriptional regulation and membrane permeability (*ompC*) among others, in the present work we further explore their potential implications and conduct a more detailed analysis of the IHC216 genome. qRT-PCR analysis highlighted significant downregulation of classical siderophore-mediated iron acquisition systems (*fepA, cirA, iroN*) and upregulation of alternative iron uptake pathways (*iucA, fiu*), reflecting a switch in iron acquisition strategies. A notable downregulation of *bla*KPC-163 correlated with restored susceptibility to carbapenems, indicating collateral susceptibility. Altered expression of *pbp2* and *pbp3* implicated adaptive changes in cell wall synthesis, potentially affecting FDC resistance mechanisms. Furthermore, enhanced oxidative stress responses via upregulated *sodC* expression and increased capsule production were observed. These findings underscore the complex interplay of genetic and

	transcriptional adaptations underlying FDC resistance, highlighting potential therapeutic vulnerabilities.
Identifier	● https://doi.org/10.5281/zenodo.16728904 ● oai:zenodo.org:16728904 ● other:Genome QC stat
Publisher	● Zenodo
Relation	● other:Genome stat ● https://doi.org/10.5281/zenodo.16728903
Rights	● info:eu-repo/semantics/openAccess ● Creative Commons Attribution 4.0 International ● https://creativecommons.org/licenses/by/4.0/legalcode
Type	● info:eu-repo/semantics/other

Part 3 – Bonus: GitHub Repository

In this part, a GitHub account was created in order to share and store the final report of the practical study.

GitHub is a collaborative platform that allows researchers and students to manage versions of their projects and make their work publicly accessible.

First, a GitHub account was created on the official platform (https://github.com).

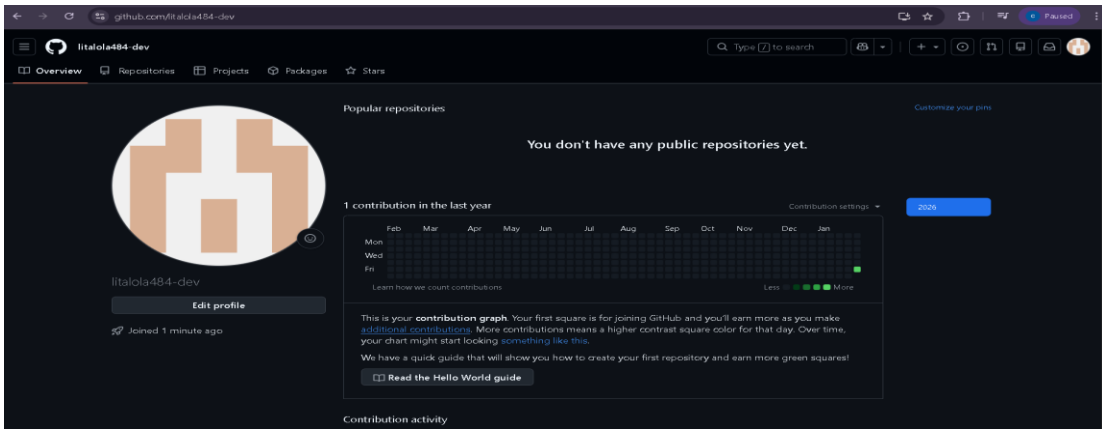


Figure 1: Creation of GitHub account.

Then, a new public repository entitled "**Practical-Work-Free-and-Open-Source-Software**" was created. This repository was dedicated to storing the final report of Part I and Part II.

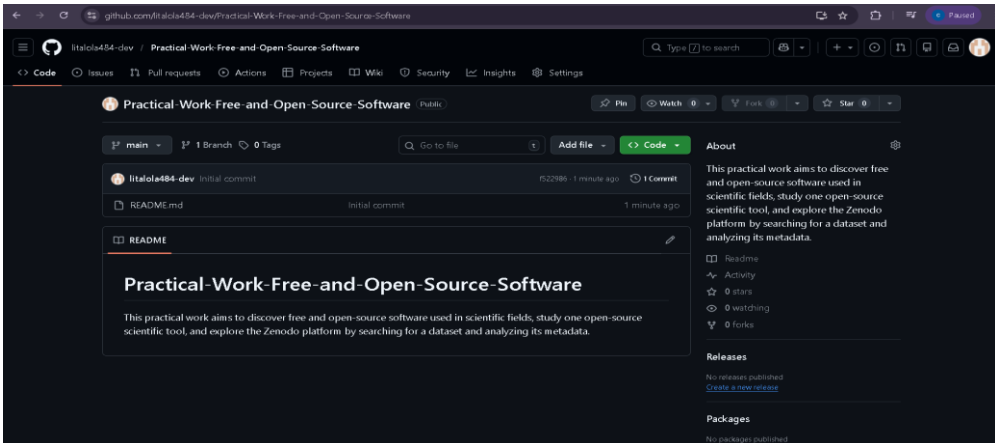


Figure 2: Creation of the repository.

Finally, the report was uploaded in PDF format to the created repository. This step allows easy access to the work, ensures transparency, and facilitates sharing with teachers and other users.

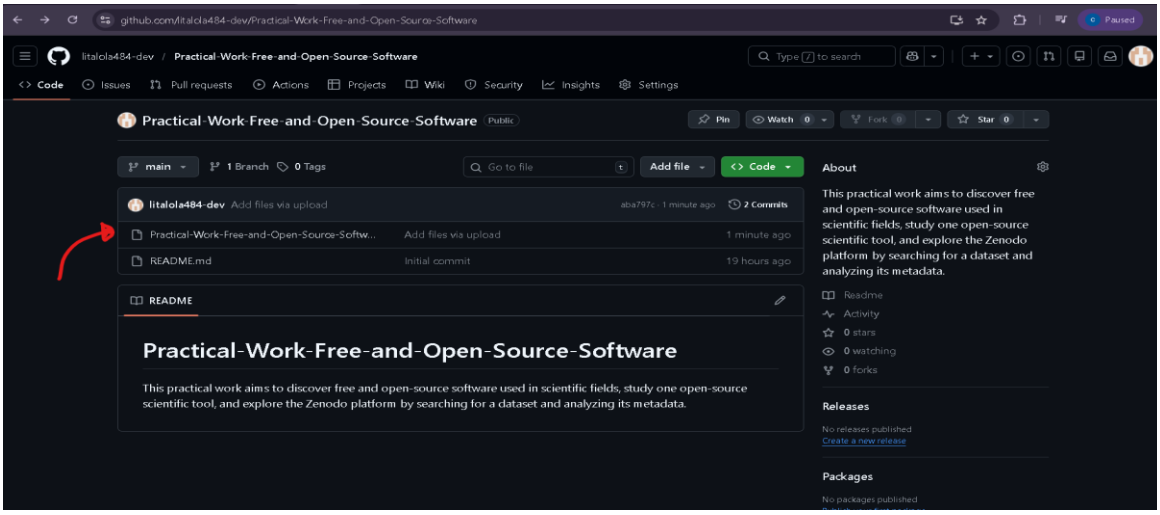


Figure 3: Successful upload of the final report (PDF) combining Part I and Part II to the GitHub repository.

●Final Submission Link

The complete project, including all documentation and the final report, is hosted and publicly accessible via the following GitHub repository link:

URL:

https://github.com/litalola484-dev/Logicial/upload/main

This platform is widely used in scientific research for open science practices and reproducibility.

References :

● **CERN. 2023**. ***Zenodo*** *– Research data repository*. European Organization for Nuclear Research.

● **Cock**, P.J.A., **Antao**, T., **Chang**, J.T., **Chapman**, B.A., **Dalke**, A., **Friedberg**, I., **Hamelryck**, T., **Kauff**, F., **Wilczynski**, B., and **de Hoon**, M.J.L. **2009**. ***Biopython****: freely available Python tools for computational molecular biology and bioinformatics*. Bioinformatics, 25(11): 1422–1423.

● **European Commission. 2021.** *Open science and data sustainability*. European Union.

●**GitHub:** https://github.com

● **OpenAIRE. 2022**. *FAIR data principles and open science*. OpenAIRE Initiative.

● **Traglia**, G.M. and **Ramirez**, M.S. 2025. *Comprehensive analysis of a spontaneous cefiderocol-resistant mutant in the Klebsiella pneumoniae KPC-163 producing strain*. **Zenodo**. DOI: 10.5281/zenodo.16728904.

● **Zenodo. n.d**. *Zenodo: Research. Shared*. European Organization for Nuclear Research (**CERN**).