



Université Mohamed Sedik Ben Yahia

Faculté des sciences de la nature et de la vie

Département de biologie moléculaire et cellulaires

Spécialité: Master 1 Biochimie

Devoirs:

TP: Logiciels Libres et Open Source

Réalisé par:

- Hezil Fatima Ezzahra
- Deffas Salsabil
- Marrouche hanadi

Proposé par:

- Bensalem amel

Année universitaire:2025_2026

Sommair:

Introduction:	2
PARTIE 1 : Étude théorique de l'outil Biopython.	3
Présentation générale de Biopython:.....	3
Fonctionnalités principales:.....	3
Aspects techniques:.....	3
Points forts:.....	3
Limites et points faibles:.....	4
PARTIE 2 : Étude pratique – Exploration de la plateforme Zenodo.	4
1. Présentation de la plateforme Zenodo:.....	4
2. Recherche effectuée sur Zenodo:.....	5
3. Dataset sélectionné:.....	5
4. Téléchargement du dataset:.....	7
5. Métadonnées du dataset (norme Dublin Core):.....	8
6. Exploitation des informations:.....	9
Conclusion:	10
Références:	11

Introduction:

Le développement rapide des technologies numériques et des outils informatiques a profondément transformé les méthodes de recherche en sciences de la nature et de la vie. La production massive de données biologiques, notamment en génomique, en biologie cellulaire et en bioinformatique, nécessite aujourd’hui l’utilisation de logiciels performants, fiables et accessibles [1]. Dans ce contexte, les logiciels libres et open source occupent une place centrale, en offrant des solutions flexibles, reproductibles et largement diffusées au sein de la communauté scientifique [1][3].

Ce travail s’inscrit dans le cadre du module *Logiciels Libres et Open Source* et vise à explorer des outils couramment utilisés en sciences de la vie. Il est structuré en deux parties complémentaires. La première partie est consacrée à une étude théorique de **Biopython**, une bibliothèque open source développée en langage Python, largement utilisée en bioinformatique pour l’analyse et la manipulation des données biologiques [2][3][4]. Cette partie permet de présenter ses fonctionnalités, ses aspects techniques ainsi que ses avantages et ses limites. La seconde partie porte sur une étude pratique de la plateforme **Zenodo**, un dépôt en libre accès dédié à la science ouverte. Cette étape a pour objectif d’explorer la plateforme, de sélectionner un dataset pertinent en lien avec les sciences de la vie, de le télécharger et d’extraire les métadonnées associées en utilisant une norme reconnue, en l’occurrence **Dublin Core**. L’accent est mis sur l’organisation, la structuration et la présentation des informations, conformément aux principes de la science ouverte.

À travers ce travail, l’objectif est de mieux comprendre l’importance des logiciels libres et des plateformes open access dans la recherche scientifique, ainsi que leur rôle dans le partage, la valorisation et la réutilisation des données en sciences de la nature et de la vie [1][4].

PARTIE 1 : Étude théorique de l'outil Biopython

Présentation générale de Biopython:

Biopython est une bibliothèque open source développée en langage Python, dédiée à la bioinformatique et aux sciences de la vie. Elle fournit un ensemble cohérent d'outils permettant la manipulation, l'analyse et l'exploitation des données biologiques telles que les séquences d'ADN, d'ARN et de protéines [1][2]. Avec l'essor des technologies de séquençage à haut débit, la quantité de données biologiques produites a considérablement augmenté, rendant nécessaire l'utilisation d'outils informatiques performants. Biopython répond à ce besoin en proposant une interface simple et flexible, adaptée aussi bien aux biologistes qu'aux bioinformaticiens [1].

Biopython fait partie d'une famille de bibliothèques bioinformatiques telles que BioPerl ou BioJava, mais se distingue par son intégration étroite avec l'écosystème Python, très utilisé en analyse de données scientifiques [3].

Fonctionnalités principales:

Biopython offre un large éventail de fonctionnalités couvrant les principales tâches en bioinformatique [2][4] :

Manipulation des séquences biologiques : lecture, écriture et modification de séquences dans différents formats standards (FASTA, FASTQ, GenBank, EMBL).

Analyse des séquences : calcul de la composition en nucléotides ou acides aminés, transcription, traduction et recherche de motifs biologiques.

Accès aux bases de données biologiques : interrogation directe de bases publiques telles que NCBI ou UniProt, ainsi que l'utilisation d'outils comme BLAST [2].

Alignement et phylogénie : comparaison de séquences et construction d'arbres phylogénétiques.

Analyse des structures protéiques : manipulation des fichiers PDB et analyse structurale [4].

Ces fonctionnalités permettent de réaliser des analyses biologiques complexes sans avoir à développer des algorithmes à partir de zéro.

Aspects techniques:

Biopython est entièrement développée en Python, un langage interprété, multiplateforme et largement utilisé dans la recherche scientifique. Elle est distribuée sous une licence open source, permettant son utilisation libre à des fins académiques et scientifiques [2].

La bibliothèque est compatible avec les principaux systèmes d'exploitation (Linux, Windows et macOS) et peut être installée facilement à l'aide de gestionnaires de paquets comme pip. Elle s'intègre également avec d'autres bibliothèques scientifiques Python telles que NumPy ou Matplotlib, ce qui renforce ses capacités analytiques [2][4].

Points forts:

Parmi les principaux avantages de Biopython, on peut citer [1][4] :

- Logiciel gratuit et open source
- Syntaxe claire et accessible grâce à Python
- Documentation officielle riche et détaillée

- Large communauté scientifique active
- Accès simplifié aux bases de données biologiques publiques
- Ces points forts font de Biopython un outil de référence en bioinformatique.

Limites et points faibles:

Malgré ses nombreux atouts, Biopython présente certaines limites [1][3] :

- Nécessite une connaissance de base du langage Python
 - Performances parfois limitées pour le traitement de très grands jeux de données
 - Certaines fonctionnalités avancées dépendent d'outils externes ou de services en ligne
- Cependant, ces limitations n'affectent pas son utilité dans la majorité des contextes pédagogiques et de recherche.

PARTIE 2 : Étude pratique – Exploration de la plateforme Zenodo

1. Présentation de la plateforme Zenodo:

Zenodo est une plateforme de dépôt en libre accès dédiée à la science ouverte. Elle a été développée par le CERN dans le cadre de l'initiative européenne OpenAIRE afin de permettre aux chercheurs de partager, conserver et diffuser leurs productions scientifiques.

Zenodo héberge différents types de contenus scientifiques, notamment des datasets, des articles, des logiciels, des posters, ainsi que d'autres résultats de recherche.

L'un des principaux atouts de Zenodo est l'attribution automatique d'un DOI (Digital Object Identifier) à chaque dépôt, garantissant une identification permanente et facilitant la citation académique. Cette plateforme joue ainsi un rôle important dans la promotion de la science ouverte, en assurant l'accessibilité, la visibilité et la réutilisation des données scientifiques.

Lien officiel de la plateforme :

<https://zenodo.org>

The screenshot shows the Zenodo website interface. At the top, there's a blue header bar with the Zenodo logo, a search bar, and navigation links for 'Communities' and 'My dashboard'. On the right side of the header are 'Log in' and 'Sign up' buttons. Below the header, there's a section titled 'Featured communities' which includes a thumbnail for 'AAS Journals PUBLISHING' and a 'Browse' button. The main content area has a heading 'Recent uploads'. It lists two items: 'HF-ETIOLOGY: Program Team Update Meeting (February 2026)' and 'HF-ETIOLOGY: Program Team Update Meeting (January 2026)'. Each item has a small thumbnail, a timestamp, a 'Presentation' label, and an 'Open' button. To the right of the uploads, there's a sidebar with the title 'Why use Zenodo?'. It contains a bulleted list of reasons: 'Safe' (research stored safely), 'Trusted' (built by CERN and OpenAIRE), 'Citable' (uploads assigned DOI), 'No waiting time' (available online immediately), 'Open or closed' (share data with access controls), 'Versioning' (easily update datasets), and 'GitHub integration'.

Figure 1: page d'accueil de la plateforme Zenodo, montrant les options principales et le champ de recherche

2. Recherche effectuée sur Zenodo:

Requête utilisée

Conformément aux consignes, une recherche a été effectuée sur Zenodo en utilisant un mot-clé lié aux sciences de la vie : “cell”

Cette requête permet d'identifier des jeux de données en relation avec l'étude des cellules, domaine central en biologie et en bioinformatique

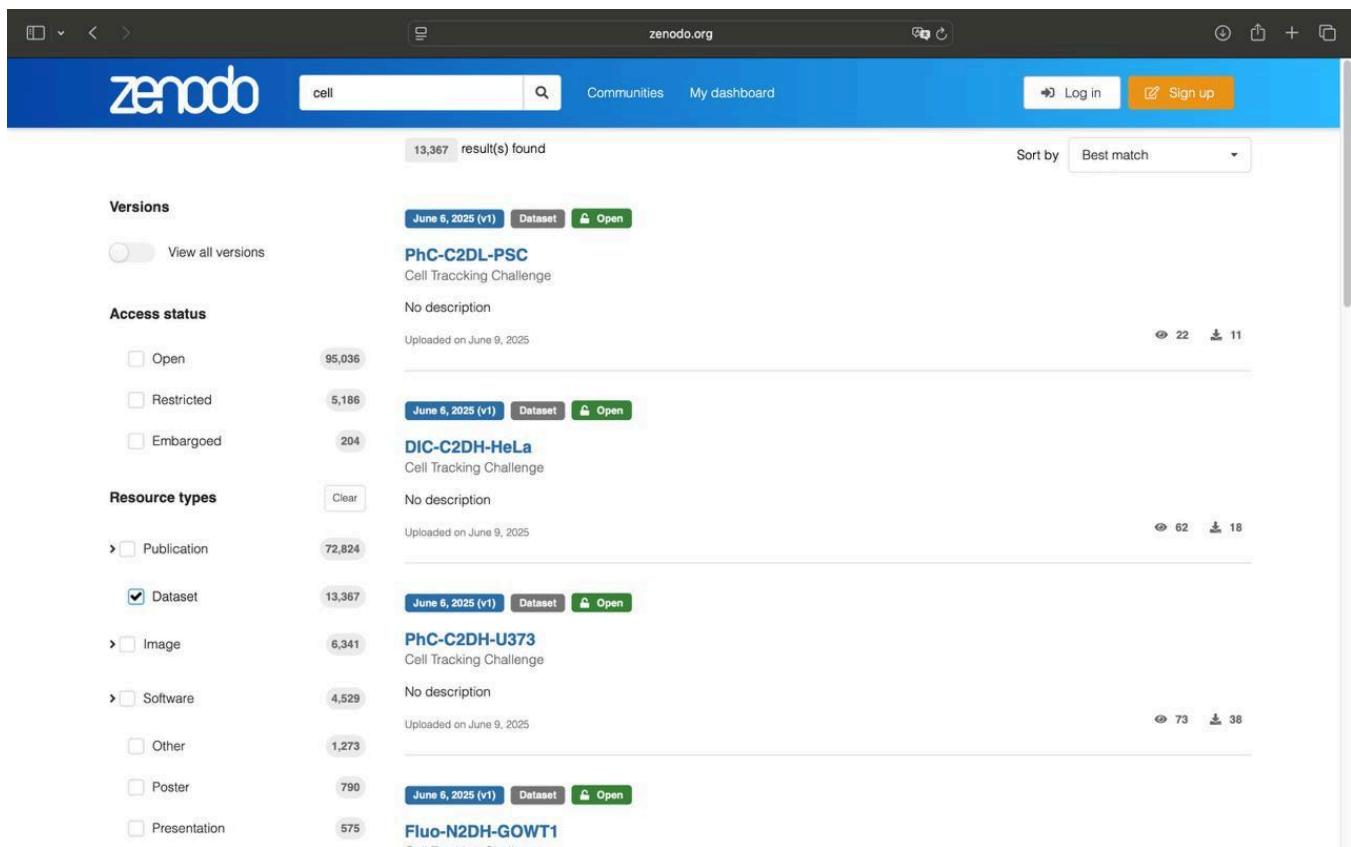


Figure 2: résultats de la recherche pour le mot-clé “cell” sur Zenodo

- **Critères de sélection du dataset**

Le dataset retenu a été sélectionné selon les critères suivants :

- Hébergement sur la plateforme Zenodo.
- Présence d'un DOI.
- Lien direct avec le thème cellulaire
- Disponibilité des métadonnées selon une norme reconnue
- Données en accès libre (Open Access).

3. Dataset sélectionné:

- **Titre du dataset : PhC-C2DL-PSC**
- **Lien officiel (DOI): <https://doi.org/10.5281/zenodo.1560823>**

Ce lien permet de vérifier directement l'authenticité du dataset, ses fichiers et ses métadonnées sur Zenodo.

- **Description générale:** Ce dataset fait partie du Cell Tracking Challenge, une initiative scientifique dédiée à l'analyse et au suivi automatique des cellules à partir d'images biologiques. Il est utilisé comme jeu de données de référence pour l'évaluation des méthodes de segmentation et de suivi cellulaire, ce qui le rend pertinent pour les domaines de la biologie cellulaire, de l'imagerie et de la bioinformatique.

The screenshot shows the Zenodo search results for the query "cell". The results are filtered by "Dataset" and sorted by "Best match". A red circle highlights the first result, "PhC-C2DL-PSC", which is described as a "Cell Tracking Challenge" dataset. An arrow points from the left towards this highlighted entry. Below the search results, a detailed view of the "PhC-C2DL-PSC" dataset is shown, including its files, versions, external resources, and details.

Zenodo Search Results:

- PhC-C2DL-PSC** (Cell Tracking Challenge) - Published June 6, 2025 (v1) | Dataset | Open
- DIC-C2DH-HeLa** (Cell Tracking Challenge) - No description | Uploaded on June 9, 2025
- PhC-C2DH-U373** (Cell Tracking Challenge) - No description | Uploaded on June 9, 2025
- Fluo-N2DH-GOWT1** (Cell Tracking Challenge) - No description | Uploaded on June 9, 2025

Dataset Details:

PhC-C2DL-PSC (Cell Tracking Challenge)

Files:

- PhC-C2DL-PSC.zip
- PhC-C2DL-PSC.zip

The previewer is not showing all the files. The files listed are:

- 01
- t000.tif (121.2 kB)
- t001.tif (121.2 kB)
- t002.tif (126.1 kB)
- t003.tif (131.1 kB)
- t004.tif (136.6 kB)
- t005.tif (141.6 kB)
- t006.tif (143.5 kB)
- t007.tif (145.0 kB)
- t008.tif (146.6 kB)

Statistics: 22 Views, 11 Downloads

Versions: Version v1 (Jun 6, 2025) | DOI: 10.5281/zenodo.1560823

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.15608231. This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

External resources: Indexed in OpenAIRE

Details: DOI: 10.5281/zenodo.15608232

Figure 3 : Page descriptive du dataset « PhC-C2DL-PSC » sur Zenodo.

4. Téléchargement du dataset:

The screenshot shows the Zenodo dataset page for "PhC-C2DL-PSC". The top navigation bar includes links for "zenodo.org", "Communities", "My dashboard", "Log in", and "Sign up". Below the header, it displays "Published June 6, 2025 | Version v1".

Dataset Summary: 22 VIEWS, 11 DOWNLOADS. A link to "Show more details" is available.

Files: A list of files including "PhC-C2DL-PSC.zip" (145.2 MB) and "Files (145.2 MB)". A "Download all" button is present.

Versions: Version v1 (Jun 6, 2025, DOI: 10.5281/zenodo.1560823). A note explains that all versions can be cited using this DOI.

Citations: No citations found.

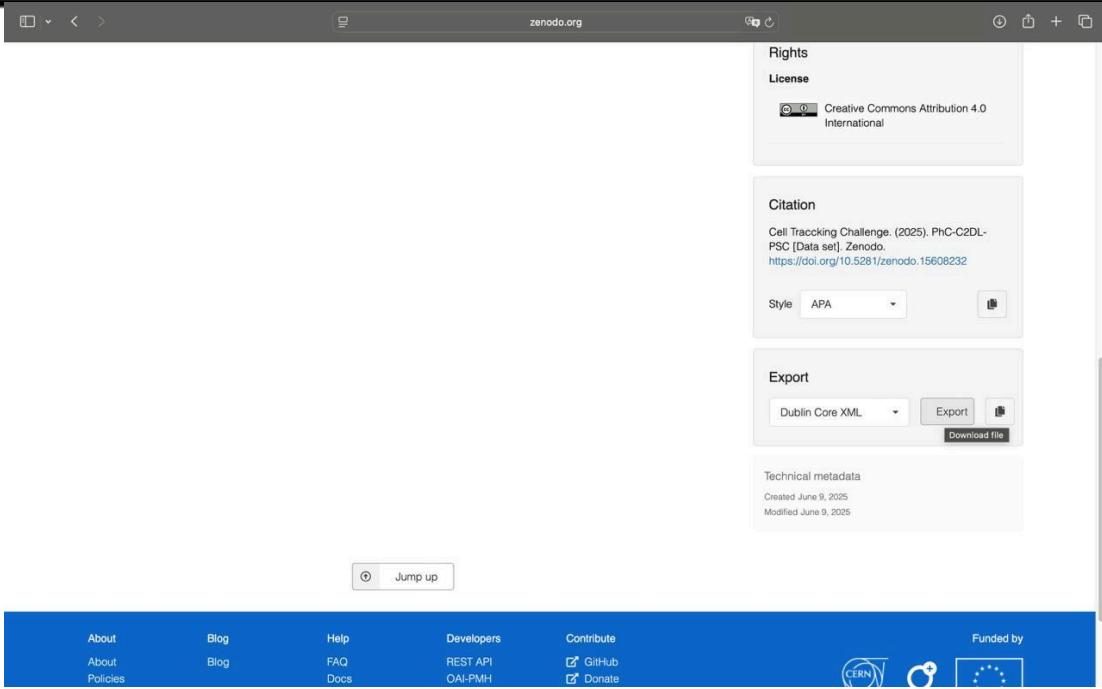
External resources: Indexed in OpenAIRE.

Details: DOI: 10.5281/zenodo.15608232.

The screenshot shows the Zenodo JSON metadata export interface. It lists various formats for exporting the dataset's metadata, including JSON-LD, CSL, DataCite JSON, DataCite XML, Dublin Core XML, MARCXML, BibTeX, GeoJSON, DCAT, Codemeta, Citation File Format, Data Package JSON, and JSON. An "Export" button is located at the bottom right of the list.

Technical metadata: Created June 9, 2025, Modified June 9, 2025.

Footer: Links for "About", "Blog", "Help", "Developers", "Contribute", "About Policies", "Blog", "FAQ", "Docs", "REST API", "OAI-PMH", "GitHub", "Donate", and "Funded by" (with logos for CERN, CERN Open Data, and the European XFEL).



Figures 4 : Téléchargement de dataset a partir de **Zenodo** avec la forme «**Dublin Core**»

5. Métadonnées du dataset (norme Dublin Core):

Les métadonnées du dataset ont été extraites selon la norme Dublin Core, comme exigé. Elles sont présentées ci-dessous sous forme de tableau:

Élément Dublin Core	information
Title	PhC-C2DL-PSC
Creator	Cell Tracking Challenge
Date	2025-06-06
Publisher	Zenodo
Identifier (DOI)	https://doi.org/10.5281/zenodo.15608232
Identifier (OAI)	oai:zenodo.org:15608232
Type	Other
Rights	Open Access
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
Relation	https://doi.org/10.5281/zenodo.15608231

Figure 5 : Métadonnées du dataset selon la norme **Dublin Core**.

Cette organisation structurée facilite la compréhension, l'indexation et la réutilisation du dataset.

6. Exploitation des informations:

Les métadonnées du dataset “PhC-C2DL-PSC” montrent que les données sont liées à un challenge de suivi cellulaire (“Cell Tracking Challenge”), publiées en juin 2025 et disponibles en accès libre au format ZIP. Ces informations permettent de comprendre le contexte scientifique du dataset et facilitent sa réutilisation pour des analyses ou des applications en imagerie cellulaire.

Conclusion:

Ce travail a mis en évidence le rôle essentiel des logiciels libres et open source ainsi que des plateformes de science ouverte dans les sciences de la nature et de la vie. L'étude théorique de Biopython a permis de souligner l'importance de cette bibliothèque en bioinformatique, notamment pour l'analyse et la manipulation des données biologiques, grâce à sa flexibilité, sa richesse fonctionnelle et son accessibilité.

Par ailleurs, l'étude pratique de la plateforme Zenodo a permis de comprendre l'intérêt des dépôts en libre accès pour le partage, la diffusion et la réutilisation des données scientifiques. La sélection d'un dataset et l'extraction de ses métadonnées selon la norme Dublin Core ont illustré l'importance de la structuration et de l'organisation des informations dans un contexte de science ouverte.

Dans l'ensemble, ce travail montre que l'association des outils logiciels open source et des plateformes open access constitue un levier fondamental pour améliorer la reproductibilité, la collaboration et la valorisation des données en recherche scientifique.

Références:

[1] CERN. (n.d.). Zenodo.

<https://zenodo.org>

[2] Cell Tracking Challenge. (2025). PhC-C2DL-PSC [Dataset]. Zenodo.

<https://doi.org/10.5281/zenodo.15608232>

[3] OpenAIRE. (n.d.). Zenodo: Research. Shared.

<https://www.openaire.eu/zenodo>

[4] Dublin Core Metadata Initiative. (n.d.). Dublin Core Metadata Element Set, Version 1.1.

<https://www.dublincore.org/specifications/dublin-core/dces/>