



Faculté des sciences de la nature et de la vie

Département : Biochimie moléculaires et cellulaires

Spécialité : Master 1 - Biochimie G 1

Module : Logiciels libres et open source

Rapport sur Logiciels Libres et Open Source

Réalisé par :

❖ *Boughaba Loubna*

❖ *Bechekef Maria*

Proposé par :

❖ *Mme. Bensalem*

Sommaire

Partie 1 : Étude théorique de EMBOSS.....	1
Présentation générale de	1
Fonctionnalités principales	1
Aspects techniques	2
Points forts	2
Limites et points faibles	2
Conclusion	3
Partie 2- Étude pratique : exploration de Zenodo	4
Présentation de Zenodo.....	4
Objectifs de la plateforme.....	4
Types de contenus4.	
Intérêt de Zenodo pour la science ouverte et la recherche en science de la nature et de la vie	5
Description Des étapes réalisées.....	5
Métadonnées du dataset.....	11
Fiche d'identification du Dataset.....	11
Analyse du tableau.....	12
Conclusion	12
Références	13

Partie 1 : Étude théorique de EMBOSS :



1. Présentation générale de :

EMBOSS (European Molecular Biology Open Software Suite), lancée en 1998 par le Sanger Institute et l'Université d'Oxford, est un ensemble gratuit et stable de logiciels pour la manipulation de séquences **ADN/ARN/protéines**, développé pour la communauté européenne de biologie moléculaire. Elle se distingue par son architecture modulaire, ses bibliothèques **AJAX** pour le développement, et sa compatibilité avec de multiples formats (**EMBL**, **GenBank**, **FASTA**), permettant une récupération automatique de données en ligne sans saisie manuelle. Mise à jour régulièrement (dernière version stable autour de 6.6.0+), elle est idéale pour les environnements **Linux/Unix**, avec des interfaces graphiques optionnelles et une documentation exhaustive.(1) (2)

➤ Exemples d'outils dans EMBOSS :

- needle → alignement global.
- water → alignement local.
- transeq → traduction ADN en protéine.
- restrict → recherche de sites de restriction.

EMBOSS Needle
Pairwise Sequence Alignment (PSA)

Job Dispatcher Help & Privacy Your Jobs Input form Feedback

Welcome to the Job Dispatcher website! If you need assistance or have feedback, please [contact](#) us.

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

Input sequence ⓘ Sequence type
☒ Protein ☐ DNA

Paste your first sequence here - or use the example sequence

```
GAAGGGGAAGAGTGGGT TGGTGGCCCTGGCCCTGCGTGAAACCCCGGCTGACCTTGGCC  
TTTGCTCACTGGGCTTGTGGCTCTGTACCCCGCGTGTGTCCATCTCGGGTCTGAGGAACC  
TTGCCCCAGTGCCCGCCAGGCTTCGGGTGTCTGTCCCCAGTGACCTCTCAACTGTGTC  
TGTTCTTTTCTCTAACCCCAATCTGTCTCTGAGGCTCTTTGAGGTTGCTTTAAGTAAG  
TATGGTTACTGGGTTGCAACAGATTGACACATCTGAGCAGTTTAAACCCAAAGCCAGGGG  
TATCCTCTTCGACGGAGGACACTTGTGGAGAATCGACCCCTTGTGGCTGCTATGTTAT  
AATCTCTACTTCCAGATGTCTAGAGGCAAGCTGTCATACCTTTGGCCCCACAGCAATC  
ATAGTAACAGTAGGGGTAGTTTTCAGGATGGTTATAATTAGAATGATTATGAGTCTCCAG
```

Figure N° 1 : capture d'écran de la surface de la plateforme EMBOSS Needle

2. Fonctionnalités principales :

- Analyse de séquences nucléotidiques et protéiques : longueur, composition (%GC), traduction
- Alignement de séquences : pairwise (needle, water) basé sur les algorithmes de Needleman-Wunsch et Smith-Waterman.
- Recherche de motifs et régions spécifiques.
- Analyse protéique : poids moléculaire, sites de clivage, domaines, hélice (pepwheel).
- Manipulation de formats : conversion entre différents formats biologiques (FASTA, GenBank...).(3)(4)

3. Aspects techniques :

- Développé principalement en “langage C”, garantissant performance et flexibilité.
- Compatible avec Linux/Unix, macOS et Windows, assurant portabilité sur toutes les plateformes majeures.
- Architecture modulaire : chaque outil fonctionne de manière indépendante via la ligne de commande.
- Intégration avec les bases de données biologiques (NCBI, EMBL, etc.), permettant l'accès direct aux séquences.
- Utilise des algorithmes classiques d'alignement (global, local, multiple) et des méthodes statistiques pour l'analyse des séquences.(3)(4)(5)

4. Points forts :

- Gratuit et open source: Fonctionne sous licence GPL, permettant l'utilisation et le développement d'outils sans frais.
- Support étendu des formats biologiques : Compatible avec FASTA, GenBank, GCG, EMBL, avec récupération directe des données depuis Internet
- Compatibilité multiplateforme : Fonctionne sur UNIX, Windows et MacOS, avec une gestion efficace de la mémoire.
- Plus de 200 outils analytiques: Alignements de séquences, recherche de motifs, analyse des protéines, prédiction de gènes, repliement de l'ARN, et sites de restriction enzymatique.
- Bibliothèques logicielles puissantes : Bibliothèques C (NUCLEUS et AJAX) permettant de développer de nouveaux outils et d'intégrer EMBOSS avec d'autres logiciels.
- Interface conviviale: Ligne de commande avec vérification automatique des entrées, sans configuration complexe ni recompilation nécessaire.
- Maintenance et support communautaire : Mises à jour régulières et ressources abondantes fournies par la communauté scientifique.
- Analyse complète et fiable : Convient à toutes les séquences biologiques (ADN, ARN, protéines) avec précision et grande flexibilité.(5)(6)(7)

5. Limites et points faibles :

- Absence d'évaluation scientifique formelle : EMBOSS est fiable, mais certains outils ou analyses n'ont pas de validation peer-review intégrée ; la qualité dépend de l'utilisateur.
- Métadonnées et documentation limitées : Les informations descriptives sont parfois générales, ce qui peut compliquer la recherche précise de fonctionnalités ou de formats spécifiques.
- Limites pour les très grandes données: Certaines analyses sur des génomes ou ensembles de protéines très volumineux peuvent être lentes ou nécessiter plus de ressources.
- Fonctionnalités collaboratives limitées : EMBOSS est conçu pour l'analyse et l'archivage, pas pour le travail collaboratif ou l'édition scientifique en temps réel.
- Performance variable selon l'utilisation : Une forte charge ou l'usage simultané de nombreux outils peut ralentir certaines opérations ou générer des restrictions système/API.(6)(7)

6. Conclusion :

EMBOSS constitue une suite bioinformatique robuste et académique, adaptée à l'analyse fondamentale des séquences biologiques. Malgré certaines limitations liées à son ancienneté et à son interface, il reste un outil fiable et largement utilisé dans l'enseignement et la recherche en biologie moléculaire.

Partie 2- Étude pratique : exploration de Zenodo



1. Présentation de Zenodo :

a. Objectifs de la plateforme :

Zenodo est un entrepôt pluridisciplinaire créé **en 2013** par le **CERN** dans le cadre du projet européen **OpenAIRE**. Il a été conçu comme un dépôt « généraliste » destiné à accueillir les résultats de la recherche scientifique, notamment lorsqu'il n'existe pas de dépôt institutionnel ou disciplinaire adapté. (8)(9)(10)

L'objectif principal de la plateforme est d'assurer la conservation et le partage des différents produits issus de la recherche, tels que les publications, les données ou les logiciels, dans l'ensemble des domaines scientifiques, conformément aux principes de la science ouverte.

Plus précisément, Zenodo a pour objectifs de :

- Permettre le dépôt, l'organisation et la publication de tous les artefacts numériques associés à un projet de recherche (données, codes sources, présentations, rapports, etc.).
- Garantir la citabilité et la traçabilité des travaux scientifiques grâce à l'attribution d'un identifiant DOI (Digital Object Identifier) pour chaque enregistrement publié.
- Réduire au maximum les barrières techniques au partage des résultats, en offrant une grande flexibilité concernant les formats, la taille des fichiers, les licences et les types d'accès (ouvert ou restreint).
- Soutenir les chercheurs dans le respect des exigences des organismes financeurs, en particulier celles liées aux politiques d'Open Access et d'Open Data de la Commission européenne.

b. Types de contenus :

Zenodo accepte des fichiers dans tous les domaines de la science et dans un large éventail de formats (texte, feuilles de calcul, images, audio, vidéo, code, archives de données, métadonnées, etc.).(10)(11)(12)

On peut y déposer par exemple :

- Publications scientifiques : articles, preprints, chapitres d'ouvrage, actes de congrès, thèses, rapports.
- Jeux de données (datasets) de toutes tailles, y compris des données brutes ou des données associées à un article.
- Logiciels et codes sources (par ex. Paquets Python, Julia, R...), avec gestion des versions.
- Présentations, posters, supports de cours et vidéos de conférences.
- Matériel complémentaire de projets, communautés ou revues (documents de méthodologie, matériels pédagogiques, etc.).
- La typologie des ressources est standardisée et basée sur le vocabulaire DataCite, ce qui facilite l'indexation et le moissonnage par les moteurs de recherche.

c. Intérêt de Zenodo pour la science ouverte et la recherche en science de la nature et de la vie :

Zenodo occupe une place importante dans le cadre de la science ouverte, car il met à disposition un dépôt gratuit, durable et interopérable pour l'ensemble des productions issues de la recherche scientifique. La plateforme permet le partage des résultats avec des métadonnées réutilisables sous licence CC0, ainsi qu'un accès via des API ouvertes, ce qui facilite leur diffusion et leur réutilisation.

Dans le domaine des sciences de la nature et de la vie, Zenodo présente plusieurs avantages. Tout d'abord, il permet la diffusion en accès ouvert des données expérimentales, des scripts d'analyse et des protocoles, ce qui contribue à améliorer la reproductibilité et la transparence des travaux scientifiques.

Ensuite, la plateforme offre une meilleure visibilité aux travaux déposés grâce à leur indexation dans différentes bases de données et moteurs de recherche (tels que BASE ou OpenAIRE). L'attribution d'un DOI à chaque dépôt améliore également la citabilité des travaux et peut avoir un impact positif sur leur reconnaissance scientifique.

Zenodo permet aussi aux chercheurs de déposer des données sensibles ou non destinées à être publiées immédiatement, en utilisant des accès restreints ou des périodes d'embargo. Cette fonctionnalité est particulièrement utile lorsque les projets sont encore en cours ou lorsque certaines données doivent être protégées temporairement.

Enfin, la création de communautés thématiques (en biologie, écologie ou sciences de la Terre, par exemple) facilite le regroupement, la découverte et la réutilisation des jeux de données et des logiciels par d'autres équipes de recherche. La diversité des contenus disponibles sur la plateforme est visible dès la page d'accueil, où l'on trouve aussi bien des données de géophysique, des outils de modélisation numérique que des ressources issues d'autres disciplines scientifique. (12) (13)(14)

2. Description Des étapes réalisées :

1- Ouvrez Google chrome , puis saisissez « Zenodo » dans la barre de recherche .

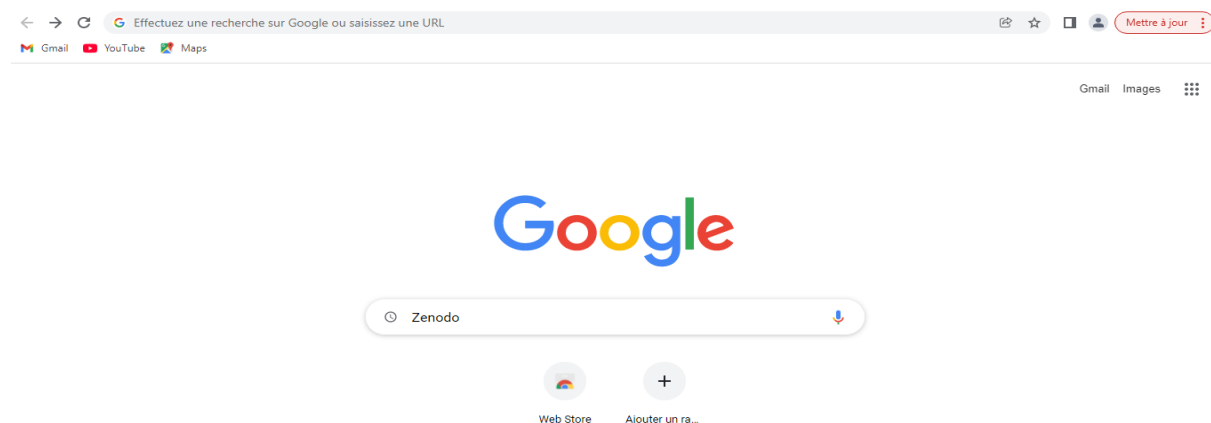


Figure N°2 :capture d'écran montrant l'étape de recherche de la plateforme Zenodo sur le navigateur Google chrome

2- Sélectionnez le lien affiché dans les résultats afin d'accéder au site Zenodo .

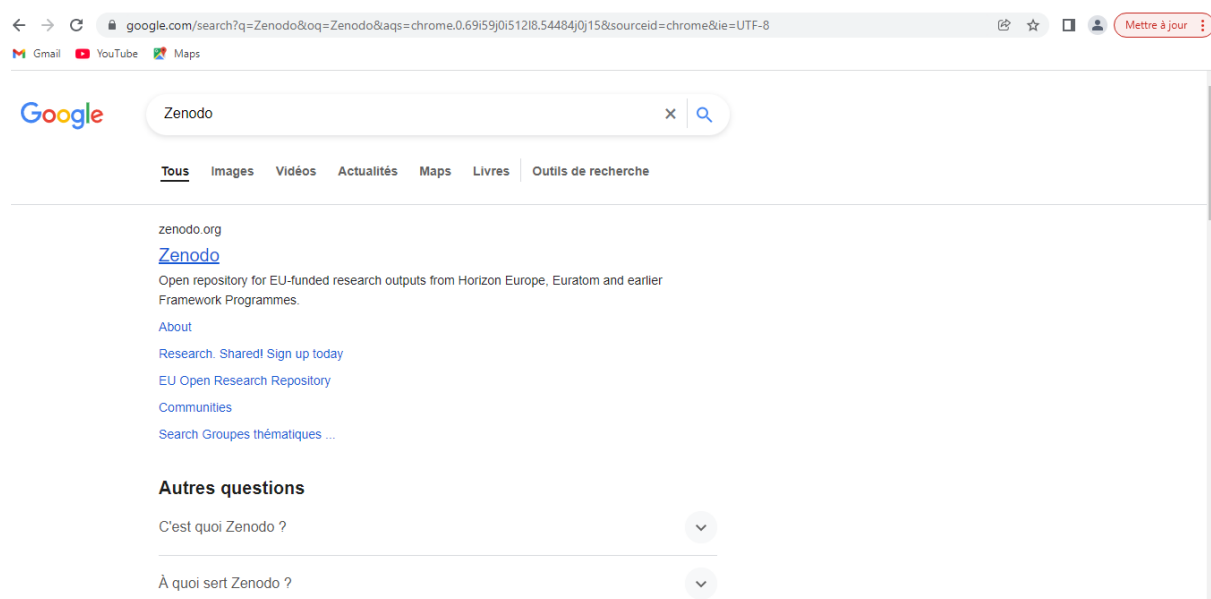


Figure N°3 :capture d'écran montrant les résultats de recherche Google et le lien d'accès à Zenodo .

3- La page d'accueil de Zenodo apparaît .

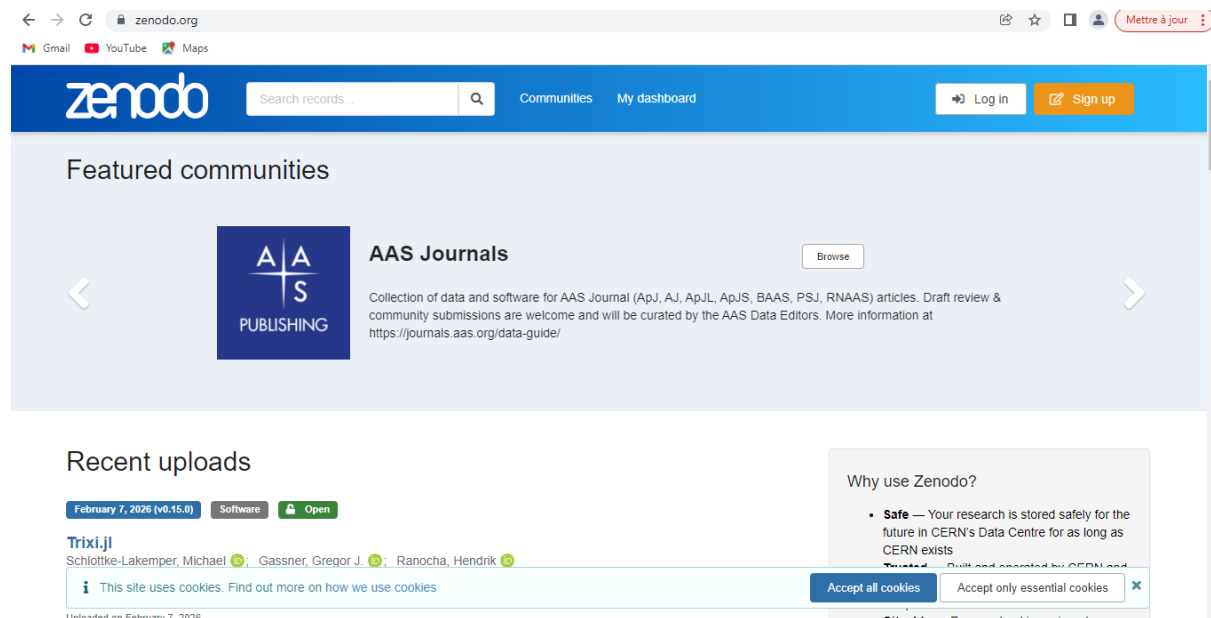


Figure N°4 :capture d'écran montrant la vue générale de la plateforme Zenodo.

4- Dans la barre de recherche, saisissez le mot « Tissue » et cliquez sur l'icône de recherche .

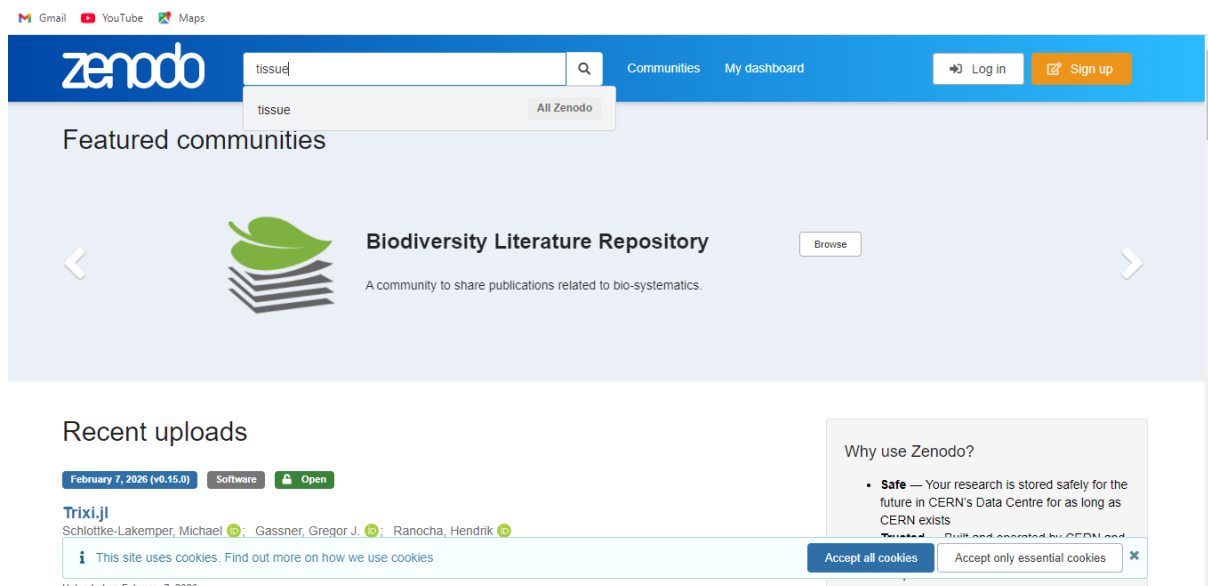


Figure N°5 :capture d'écran montrant l'utilisation du mot-clé « Tissue » dans la barre de recherche .

5- Après , sur la gauche ,cliquez sur « Resource types » et sélectionnez « Dataset » .

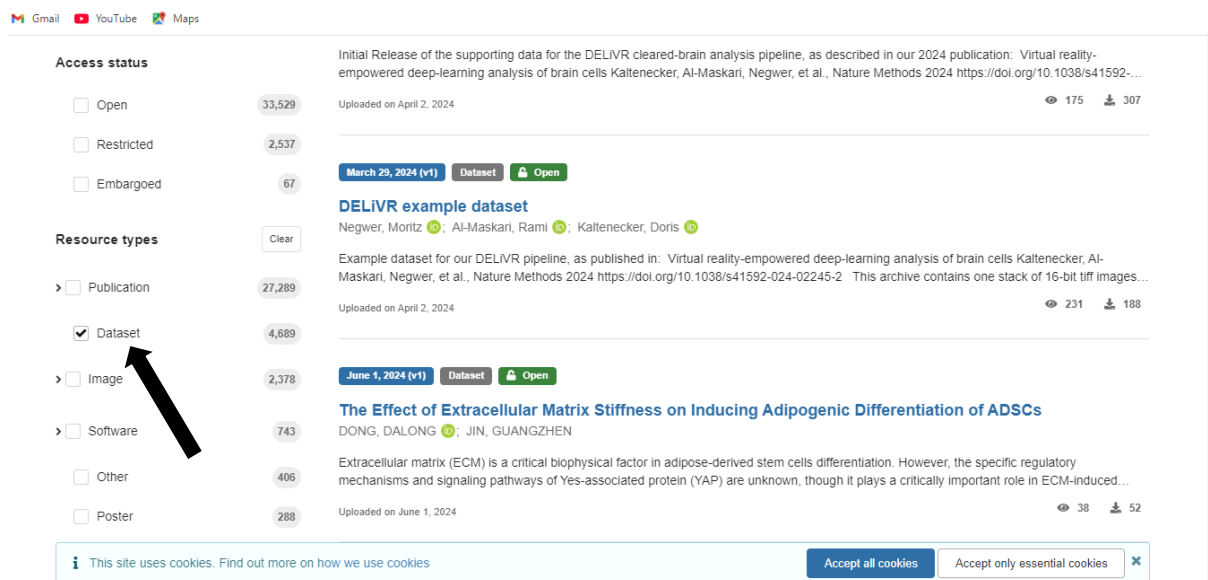


Figure N°6 :capture d'écran montrant l'activation du filtre « Dataset » dans la section « Resource types »pour affiner les résultats.

6- sélectionnez un titre quelconque, nous allons choisir celui-ci comme exemple, puis cliquez dessus pour l'ouvrir .

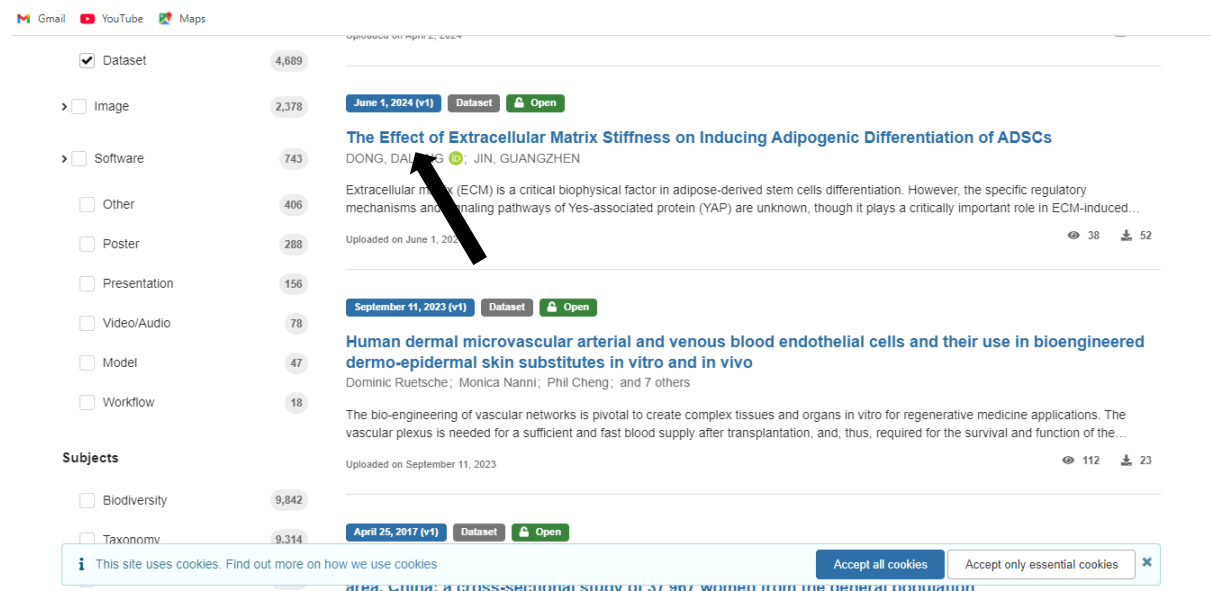


Figure N°7 :capture d'écran montrant la sélection du titre du dataset retenu pour l'étude .

7-Après avoir ouvert le titre ,regardez le nom du dataset , la date de publication , l'auteur et d'autres informations utiles .comprenez ce que le contient le dataset et b qui créé .

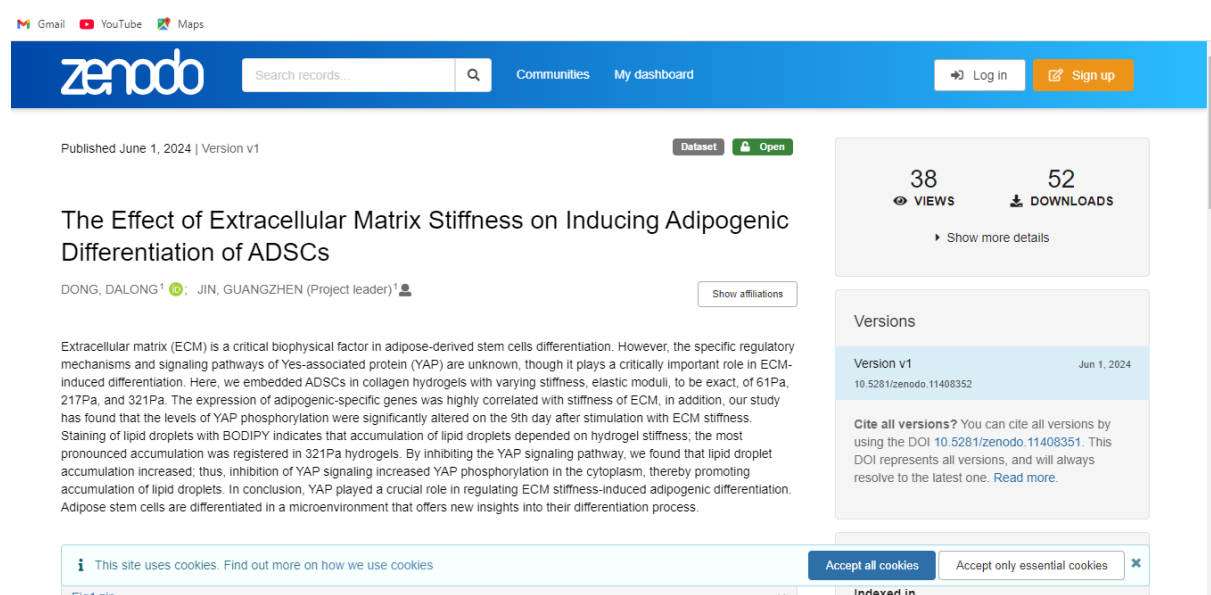


Figure N°8 :capture d'écran illustrant l'affichage des métadonnées et des détails du dataset sur Zenodo .

8- Téléchargez le fichier .

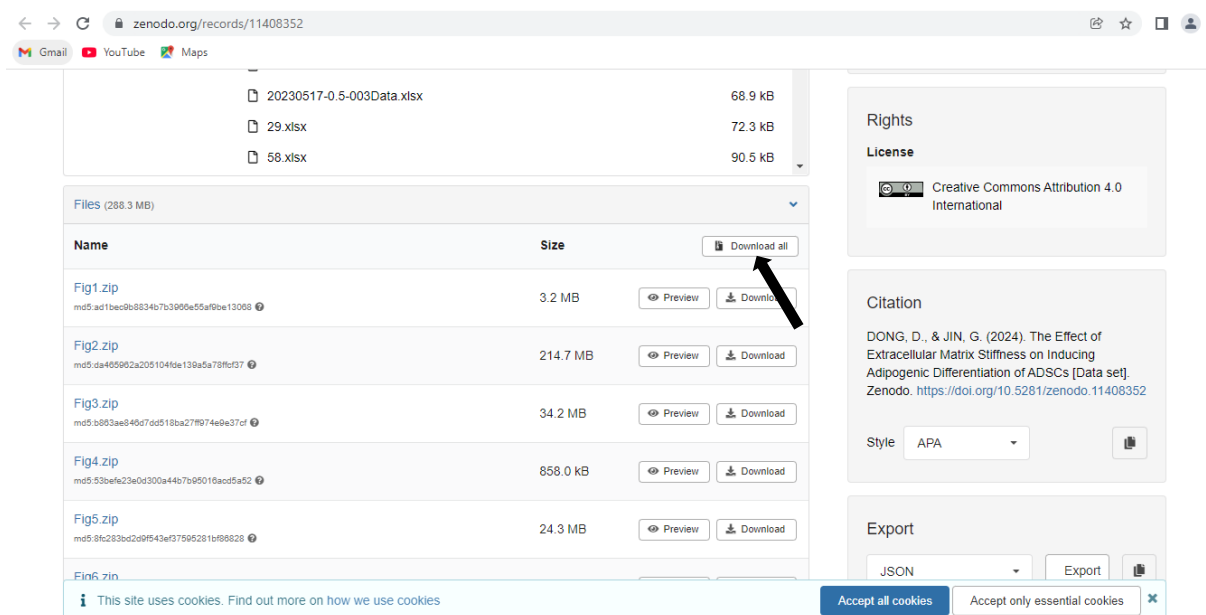


Figure N°9 :capture d'écran illustrant l'étape de téléchargement des fichiers du dataset .

9- Le téléchargement est terminé ; le fichier est maintenant enregistré sur l'ordinateur .

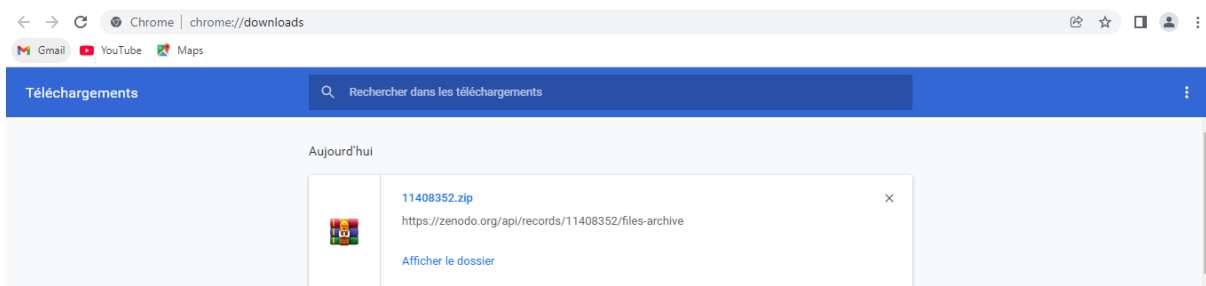


Figure N°10 :capture d'écran illustrant la confirmation du téléchargement du fichier compressé (ZIP) .

10- Retournez à la page ,allez sur le côté droit et trouvez « Export » .

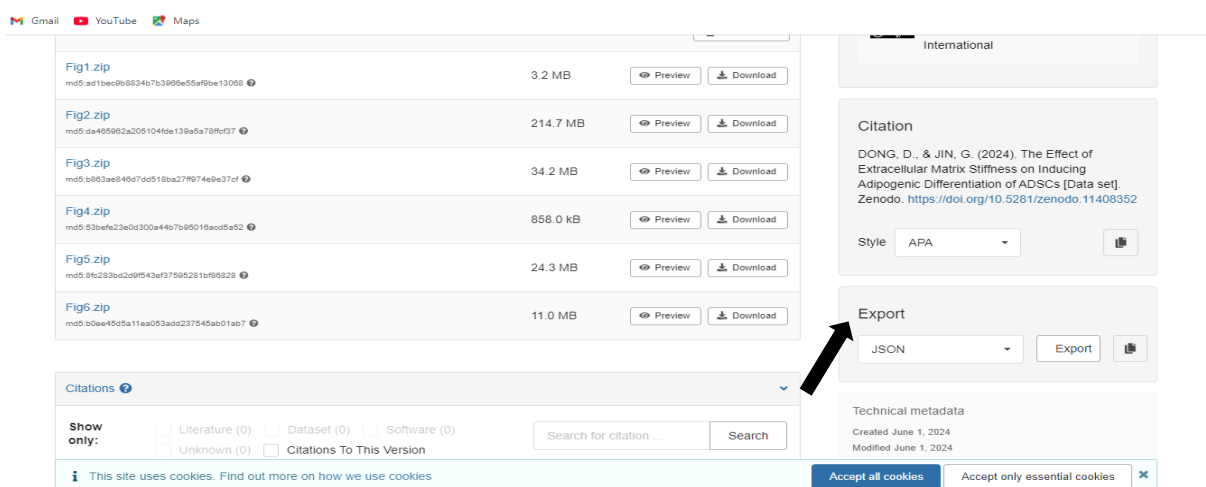


Figure N°11 :capture d'écran montrant l'accès à l'option « Export » pour la citation .

11- Cliquez sur le format « JSON » puis choisissez « Dublin Core XML » dans la liste déroulante .

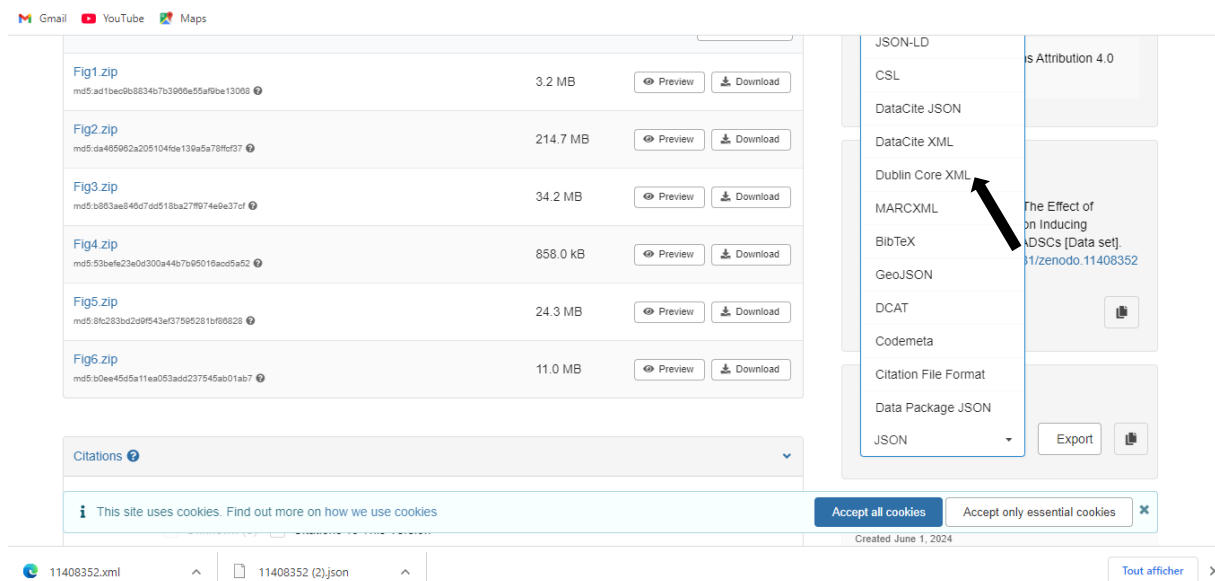


Figure N°12 : capture d'écran montrant la sélection du format « Dublin Core XML » pour l'exportation .

12- Cliquez sur le bouton « Export » pour télécharger les métadonnées sous le format choisi .

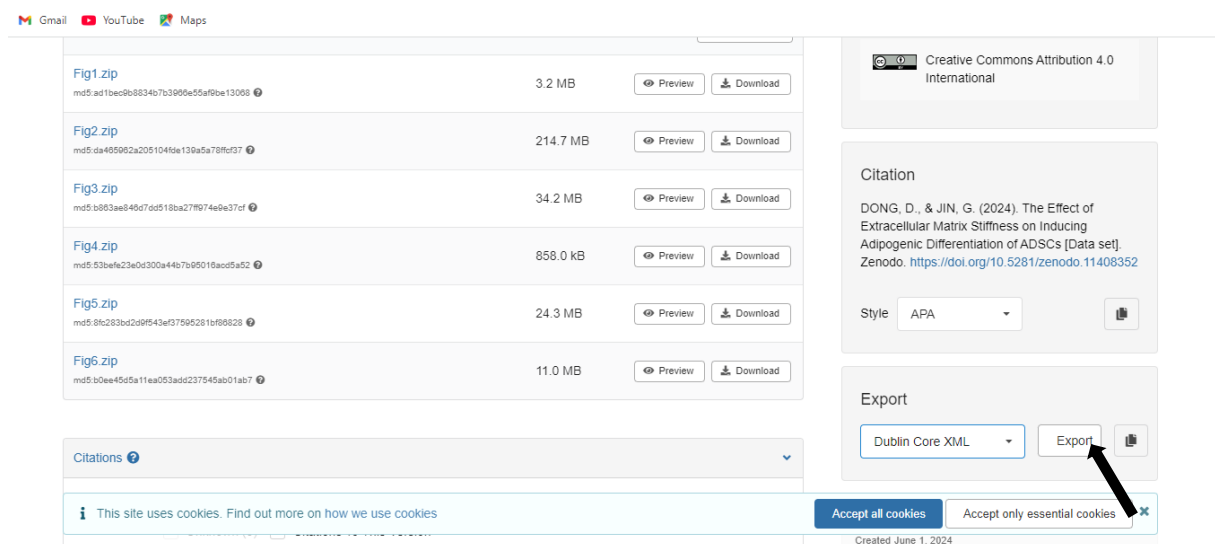


Figure N°13 : capture d'écran illustrant le clic sur le bouton « Export » pour générer le fichier de citation .

3. Métadonnées du dataset :

- Pour l'extraction des métadonnées, nous avons utilisé le standard international Dublin Core, reconnu pour sa simplicité et son efficacité dans la description des ressources numériques. La capture d'écran ci-dessous montre l'interface de récupération des données à partir de Zenodo .



Figure N°14 :capture d'écran montrant l'interface d'extraction des métadonnées sous le standard « Dublin Core ».

Ces informations sont synthétisées et organisées dans le tableau suivant :

❖ Fiche d'identification du Dataset

Élément Dublin Core	Valeur (Données du Dataset)
DOI	10.5281/zenodo.11408352
Titre du dataset title	The Effect of Extracellular Matrix Stiffness on Inducing Adipogenic Differentiation of ADSCs
Créateur(s)	DONG, D., & JIN, G.
Date de publication	June 1,2024
Plateforme d'hébergement	Zenodo
Type de ressource	Datasets
Description	Cette étude examine l'effet de la rigidité de la matrice extracellulaire sur la différenciation adipogénique des cellules souches dérivées du tissu adipeux .
Objectif du dataset	Fournir des données exploitables pour la recherche scientifique sur l'ingénierie tissulaire et la biologie cellulaire.
Licence	Creative Commons Attribution 4.0 International
Accès	Open Access (Accès Libre)
Mots-clés	ADSCs, ECM Stiffness, Adipogenic Differentiation, Regenerative Medicine.
Format des fichiers	Application/vnd.openxmlformats-office document.spreadsheetml.sheet (XLSX, 22.4 KB)
Langue	Anglais
Version	Version v1 10.5281/zenodo.11408352 Jun 1, 2024
Lien officiel	https://doi.org/10.5281/zenodo.11408352/

Les métadonnées ont été extraites à partir du DOI officiel (10.5281/zenodo.11408352) le [8/2/2026]

❖ **Analyse du tableau :**

D'après les informations que nous avons regroupées dans ce tableau, nous pouvons voir que ce dataset est très bien documenté. L'utilisation du standard Dublin Core nous a permis de structurer clairement les données importantes, comme le nom des auteurs, la date de publication et le lien DOI pour garantir la source. Nous avons également noté que le fichier est disponible en format Excel , ce qui facilite l'accès aux données pour les analyser plus tard

❖ **Conclusion :**

Pour conclure, ce travail de recherche sur les métadonnées nous a permis de comprendre comment sont organisées les données scientifiques sur une plateforme comme Zenodo. En remplissant ces champs, nous rendons l'information plus accessible et plus facile à partager. C'est pour nous une étape essentielle pour assurer la visibilité et la fiabilité des résultats de cette étude.

Références

- (1). [emboss.open-bio] (<http://emboss.open-bio.org/html/use/pr02s01.html>)
- (2). rssf.i2bc.paris-saclay] (<http://rssf.i2bc.paris-saclay.fr/gautheret/cours/EMBOSS.pdf>) – [emboss.open-bio] (<http://emboss.open-bio.org/html/use/pr02s01.html>)
- (3). Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS : The European Molecular Biology Open Software Suite. Trends in Genetics, 16(6), 276–277.
- (4). Site officiel : <http://emboss.open-bio.org> (introduction, téléchargement, guide).
- (5). Documentation utilisateur : <http://emboss.sourceforge.net/docs/> (instructions détaillées).
- (6). EBI exemple : https://www.ebi.ac.uk/jdispatcher/psa/emboss_needle (outil Needle).
- (7). Interface web : <https://www.bioinformatics.nl/cgi-bin/emboss> (EMBOSS Explorer).
- (8). <https://home.cern/fr/news/news/computing/open-science-against-covid-19-how-zenodo-and-openaire-support-scientists> [home] (<https://home.cern/fr/news/news/computing/open-science-against-covid-19-how-zenodo-and-openaire-support-scientists>)
- (9). <https://openscience.pasteur.fr/2022/12/07/zenodo/> [openscience.pasteur] (<https://openscience.pasteur.fr/2022/12/07/zenodo/>)
- (10). <https://about.zenodo.org/> [about.zenodo] (<https://about.zenodo.org/>)
- (11). <https://coop-ist.cirad.fr/trouver-l-information/les-ressources/zenodo> [coop-ist.cirad] (<https://coop-ist.cirad.fr/trouver-l-information/les-ressources/zenodo>)
- (12). <https://www.univ-smb.fr/bu/science-ouverte/zenodo/> [univ-smb] (<https://www.univ-smb.fr/bu/science-ouverte/zenodo/>)
- (13). Guide officiel OpenAIRE : <https://www.openaire.eu/zenodo-guide> [openaire] (<https://www.openaire.eu/zenodo-guide>)
- (14). Page officielle Zenodo : <https://about.zenodo.org> [about.zenodo] (<https://about.zenodo.org>)