Université Mohamed Sedik Benyahyia Jijel

Faculté de science de la nature et de la vie

Département de biologie moléculaire et cellulaire

****

TP :logiciel libre et open source

Etude théorique et pratique et exploration de zenodo

Préparé par : _Youla Dounia

_ ladra riham

Réalisé par : A_Bensalem

Année universitaire :2025/2026

Plan de travail :

- **Partie 1** :Étude théorique
- Présentation générale de l'outil(EMBOSS)
- Fonctionnalités principales
- Aspects techniques
- Points forts Limites et points faibles
- Conclusion
- Référence
- **Partie 2** : Étude pratique Exploration de Zenodo :
- Recherche,
- sélection et téléchargement d'un dataset pertinent.
- Analyse des métadonnées : Présentation structurée des métadonnées selon la norme choisie (Darwin Core ou Dublin Core). Qualité globale du rapport (2 points) Clarté, structuration, qualité de la forme et des références.
- Bonus :Création d'un compte GitHub, dépôt

Partie 1 :étude théorique de l'outil EMBOSS

1 /présentation général de l'outil :

EMBOSS (European Molecular Biology Open Software Suite) est une suite de logiciels libres et open source dédiée à l'analyse des données biologiques, notamment les séquences d'ADN, d'ARN et de protéines. Elle a été développée pour fournir aux chercheurs et aux étudiants un ensemble d'outils fiables pour la bioinformatique et la biologie moléculaire. Développé à l'origine par l'EMBL-EBI, EMBOSS met à disposition plusieurs centaines d'outils en ligne de commande permettant d'effectuer des analyses courantes en biologie moléculaire, aussi bien en recherche qu'en enseignement ;EMBOSS est largement utilisée dans les laboratoires de recherche et les institutions académiques, et fonctionne principalement via une interface en ligne de commande.

2/Principales fonctionnalités :

EMBOSS comprend plus de 100 programmes en ligne de commande couvrant de nombreux aspects de la biologie moléculaire. Parmi ses fonctionnalités principales

- **Analyse de séquences** : traduction des séquences nucléotidiques en protéines, recherche de motifs, complément inverse de l'ADN, calcul de la composition.
- **recherche dans les bases de données** : alignements de paires de séquences, recherche de motifs et interrogation de bases de données locales ou distantes.
- **Analyse des protéines** : prédiction de la structure secondaire, sites de clivage, hydrophobicité, calcul du poids moléculaire.
- Conversion de formats de séquences : conversion entre différents formats (FASTA, GenBank, EMBL…).
- **Outils de visualisation** : génération de graphiques, diagrammes et rapports pour représenter les caractéristiques des séquences.

3/Aspects techniques

- **Plateforme** : multiplateforme (Linux, Windows, macOS).
- **Langage de programmation** : principalement en C, avec certains scripts en Perl et Python.
- **Interface** : outils en ligne de commande, avec interfaces graphiques optionnelles comme JEMBOSS.
- **Intégration** : compatible avec les bases de données biologiques (NCBI, UniProt) et autres logiciels bioinformatiques.
- **Licence** : open-source sous GNU General Public License (GPL), permettant l'utilisation, la modification et la redistribution gratuites.

4/. Points forts :

Libre et open-source : gratuit et modifiable.

Large éventail d'outils : solutions pour de nombreuses analyses en biologie moléculaire.

Intégration facile : compatible avec d'autres logiciels et bases de données.

Multiplateforme : fonctionne sur différents systèmes d'exploitation.

Documentation **complète** : manuels, tutoriels et communauté active.

5. /Limites et faiblesses :

Interface en ligne de commande : peut être difficile pour les débutants sans connaissances en informatique.

GUI limitée : les interfaces graphiques sont moins développées que dans certains logiciels commerciaux.

Maintenance : certains outils anciens ne sont plus activement mis à jour.

Performance : pour des analyses à grande échelle, des scripts supplémentaires ou outils externes peuvent être nécessaires.

6/ Conclusion :

EMBOSS est une suite logicielle puissante, polyvalente et open-source pour la biologie moléculaire et la bioinformatique. Son large éventail d'outils et sa flexibilité en font un outil précieux pour la recherche et l'enseignement. Bien que l'interface en ligne de commande puisse représenter une difficulté pour les débutants, la disponibilité de tutoriels et le soutien de la communauté rendent EMBOSS accessible et fiable pour l'analyse de séquences et les flux de travail bioinformatique.

Références:

1.Rice P., Longden I., Bleasby A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics, 16(6), 276–277.

2.EMBOSS official site: https://emboss.open-bio.org

3.Wikipedia: EMBOSS – https://en.wikipedia.org/wiki/EMBOSS

PARTIE2 -Étude pratique :

Exploration de zenodo :

1/présentation de plateforme zenodo :

Zenodo est une plateforme de dépôt numérique en libre accès destinée à la diffusion et à la conservation des productions scientifiques. Elle a été développée pour soutenir le mouvement de la science ouverte (Open Science) en offrant aux chercheurs un espace fiable pour partager leurs données et leurs travaux.

Objectifs de la plateforme :

- Assurer l'accès libre aux données scientifiques
- Permettre l'archivage à long terme des ressources de recherche
- Attribuer un identifiant permanent (DOI) à chaque dépôt
- Favoriser la réutilisation et la reproductibilité des recherches

Types de contenus hébergés :

Zenodo héberge plusieurs types de ressources scientifiques :

- Jeux de données (datasets)
- Logiciels et codes sources
- Publications scientifiques
- Présentations et posters
- Supports pédagogiques

Importance de Zenodo pour la science ouverte et la recherche en NLS :

Zenodo joue un rôle fondamental dans la diffusion des connaissances scientifiques. Il permet aux chercheurs en sciences naturelles et sciences du langage (NLS) de partager leurs données, d'améliorer la transparence des recherches et de renforcer la collaboration scientifique à l'échelle internationale.

2/ description des étapes réalisées :

2 .1 Accès a la platforme zenodo :

La plateforme est accessible via l'adresse suivante :

https://zenodo.org

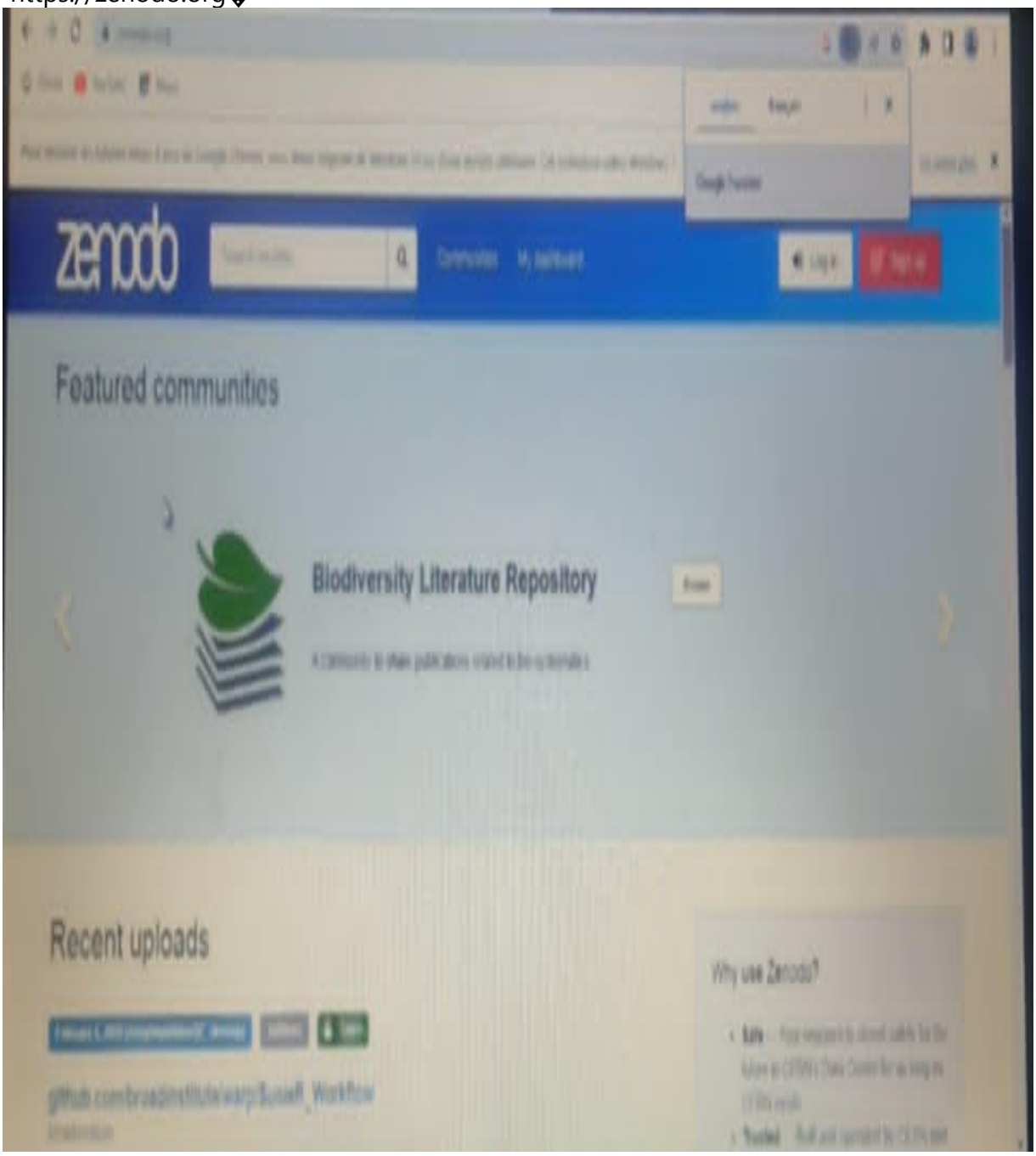


Figure 1 :page d'accueil de la plateforme zenodo

2.2 recherche effectuée d'un dataset

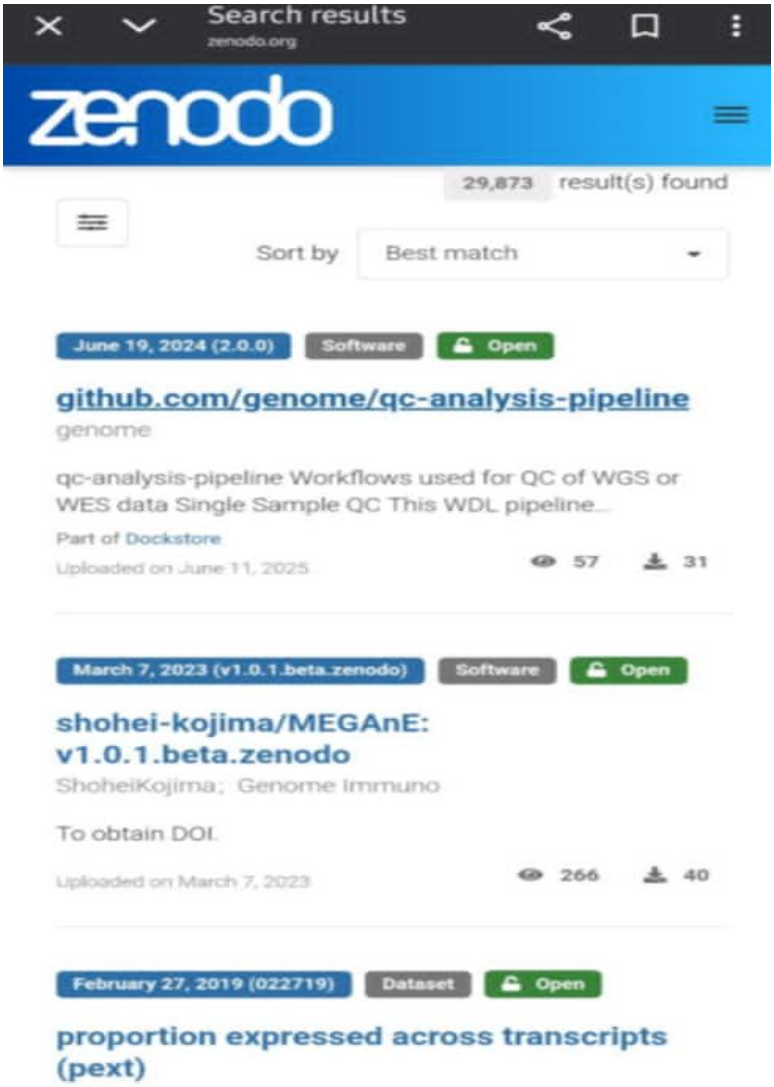


Figure 2 :search résulte for mot-clé :génome

2.3 sélection du dataset :

Parmi les résultats obtenus, un dataset pertinent a été sélectionné selon les critères suivants :

Type de document : Dataset

Domaine scientifique : Génomique

Métadonnées complètes et bien structurées

Accès libre (Open Access)

Figure 3 :page de dataset proportion expressed across transcripts

Le dataset sélectionné est intitulé Proportion expressed across transcripts. Il a été choisi car il est lié à l'analyse de l'expression génétique dans le domaine du génome

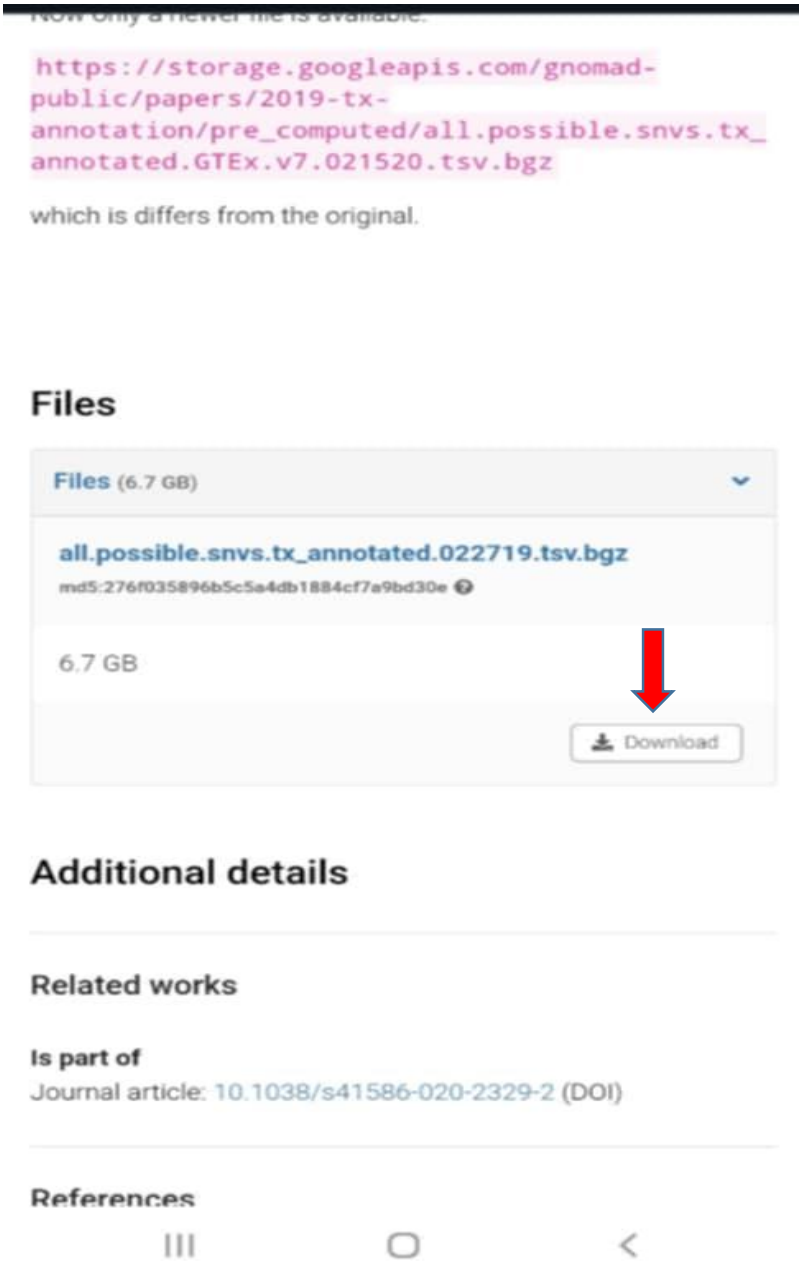


Figure 4 :téléchargement du dataset depuis la plateforme zenodo

Les fichiers du dataset ont été téléchargés à partir de la plateforme Zenodo afin de permettre leur analyse ultérieure

3/métadonnées du dataset

3_1 norme utilisée :

Les métadonnées ont été récupérées selon la norme Dublin Core, une norme internationale largement utilisée pour la description des ressources numériques

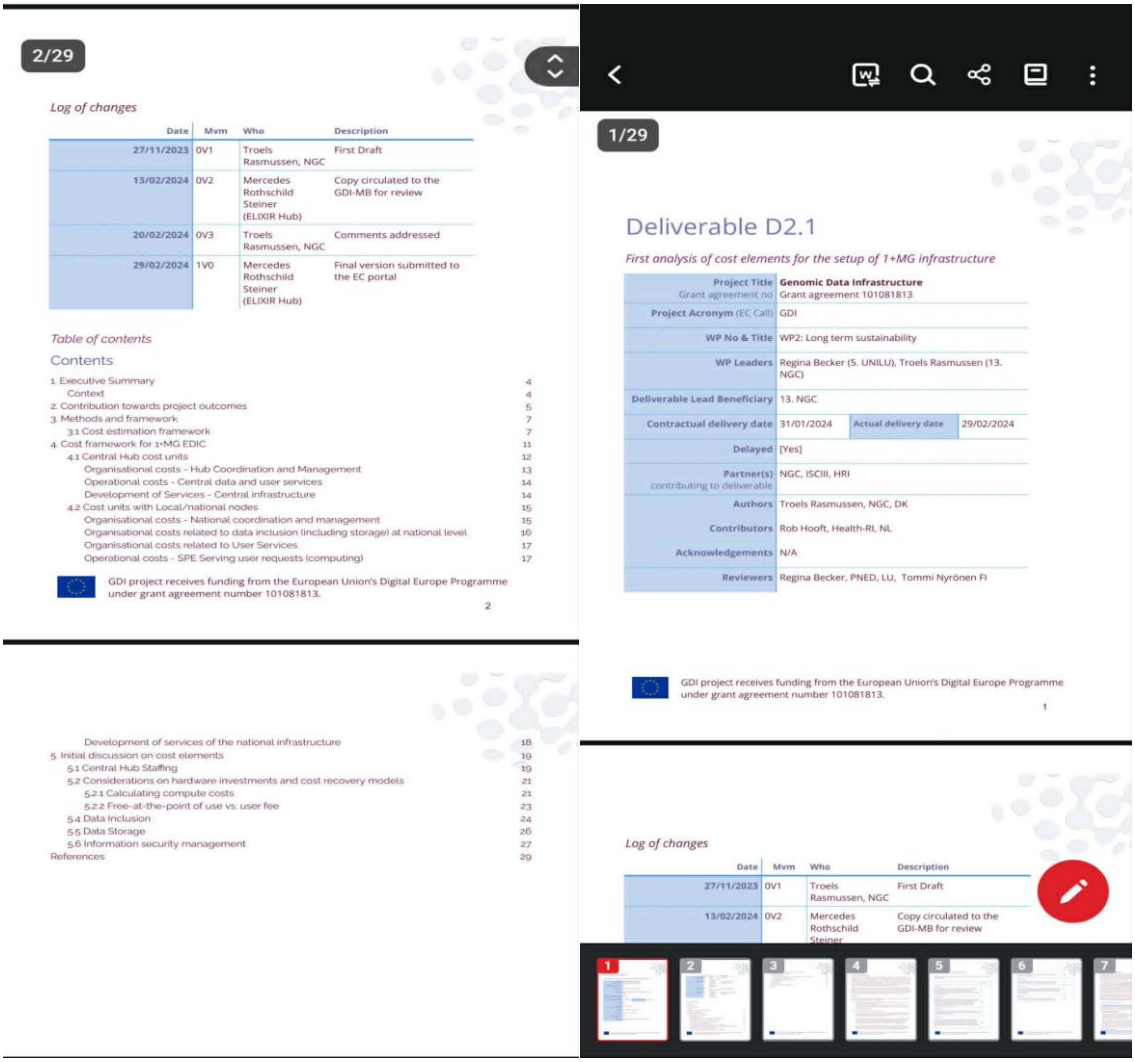


Figure 5 :dataset téléchargé

Tableau des métadonnées(dublin core) :

Elément (dubline core)	Information
Titre	Proportion expressed across transcripts (pext)
Créateur	Genome Aggregation Database Production Team Genome Aggregation Database Consortium
Subject	Genome gene expression transcripts pext
Description	Ce dataset fournit des informations sur la proportion d'expression des variants génétiques à travers différents transcrits, permettant une meilleure interprétation fonctionnelle des variantes génomiques
Publisher	Zenodo
Contributor	Genome Aggregation Database Consortium
Date	27 février 2019
Type	dataset
Format	TSV,BGZ,ZIP
Identifiant	DOI:10.5281/ Zenodo .4447230
Source	Genome aggregation Database (gnomAD)
Rights	Open access
Langue	English

Les métadonnées ont été extraites à partir de la page du dataset sur la plateforme Zenodo selon la norme Dublin Core. Ces informations permettent de décrire la ressource numérique et de faciliter sa citation et sa réutilisation scientifique

Conclusion :

Cette étude pratique a permis d'explorer la plateforme Zenodo, d'effectuer une recherche basée sur le mot-clé genome, de sélectionner un dataset pertinent (pext) et d'extraire ses métadonnées selon le standard Dublin Core. Zenodo constitue un outil fondamental pour la diffusion et la valorisation des données scientifiques en accès libre

Partie 3 :BONUS :GitHub

Etape 1 :créer un compte GitHub :

Étapes suivies :

- Accès au site officiel de GitHub via le lien suivant : https://github.com
- Cliquer sur le bouton "Sign up".
- Renseigner les informations demandées : Nom d'utilisateur (Username) Adresse e-mail Mot de passe Valider la vérification de sécurité (captcha).
- Confirmer l'adresse e-mail via le message reçu.

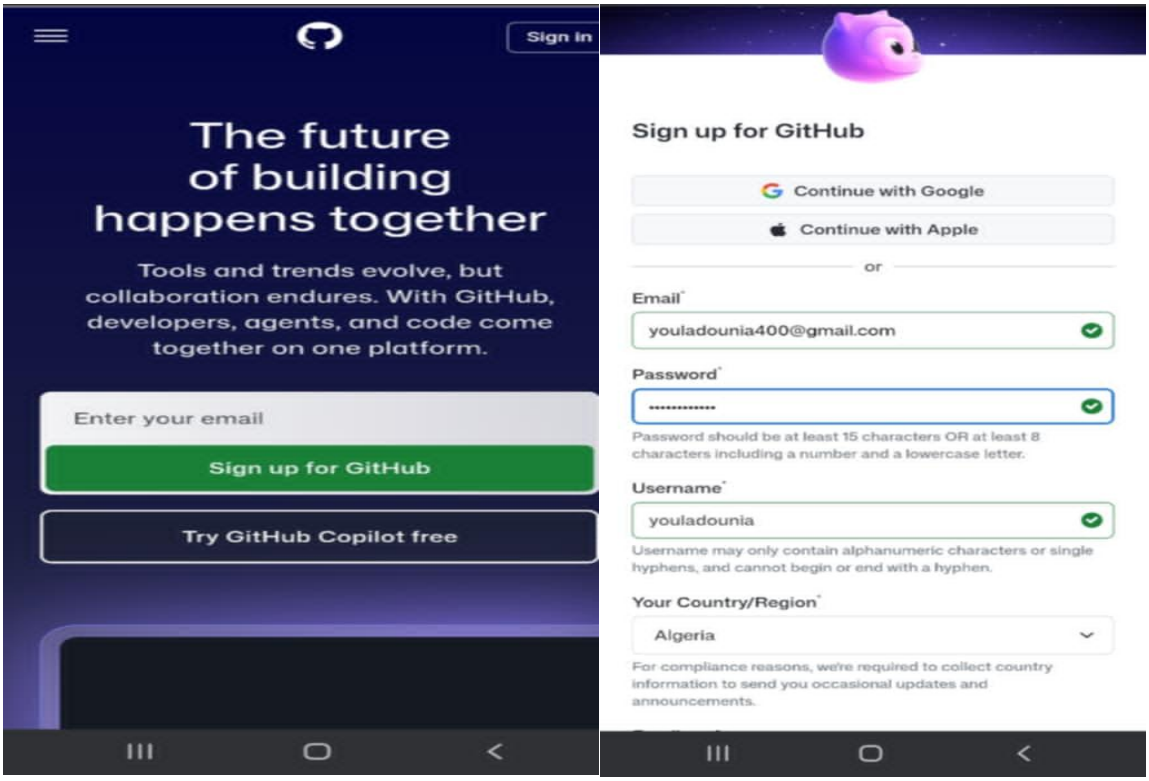
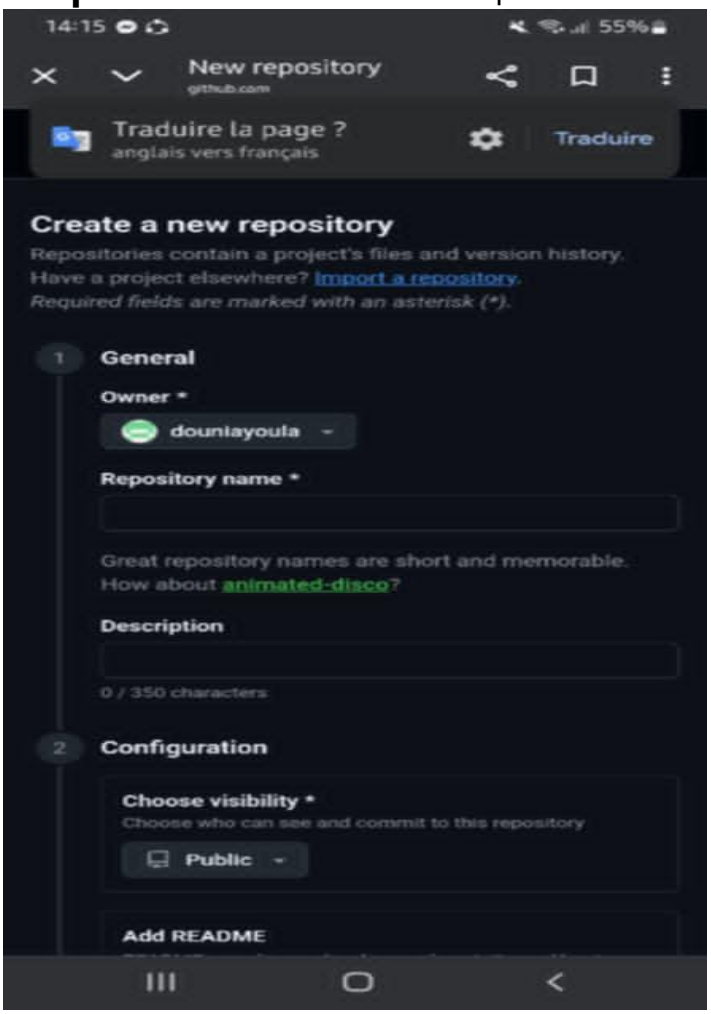


Figure 6 :création d'un compte sur la plateforme GitHub

Etape 2 : Créer un nouvel entrepôt



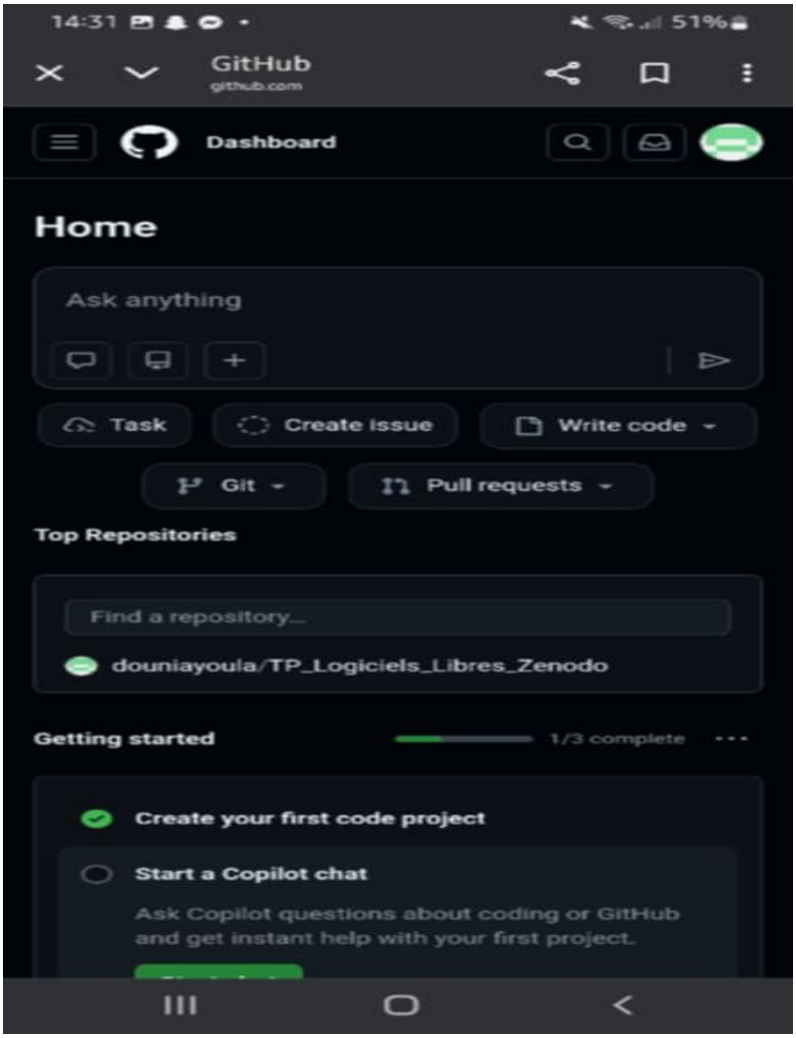


Figure 7 :création un dépôt (repository)

.