

Des ensembles de données inclusifs pour l'apprentissage automatique

Résistance aux antimicrobiens

Webinaire pour les candidats

10 juillet 2024

Intervenants

MERCY AYUB, ACTS

ADELA BRIANSÓ JUNQUERA, LACUNA FUND

SAMUEL WANJAU, ACTS

DR. OJENGE WINSTON, ACTS

Dr. ELIZABETH MUTUA, conseillère en qualité des données de l'ACTS



Objectifs de la réunion

- Fournir aux candidats potentiels une compréhension du Fonds Lacuna et des exigences en matière de propositions.
- Prévoir du temps pour répondre aux questions des candidats sur l'appel à propositions.



Ordre du jour

- 00:00** **Bienvenue, introduction,** examen de l'**ordre du jour**
- 00:10** **Présentation : Vue d'ensemble du Fonds Lacuna et exigences pour les appels d'offres, éléments d'une proposition solide**
- 00:40** **Présentation : Qualité des données et considérations relatives à l'hébergement**
- 00:50** **Présentation : Opportunité de mentorat et d'appariement**
- 01:00** **Questions et réponses**
- 01:25** **Prochaines étapes**
- 01:30** **Fin**

Une lacune est un vide ou une partie manquante d'un manuscrit.

Lacuna Fund soutient la création, l'expansion et la maintenance d'ensembles de données de formation et d'évaluation équitables qui permettent aux outils d'apprentissage automatique de mieux s'attaquer aux problèmes urgents dans les contextes de revenus faibles et moyens à l'échelle mondiale.

Principes qui guident Lacuna Fund :

- Accessibilité
- Fonds propres
- L'éthique
- Approche participative
- Qualité
- L'impact transformationnel



Bailleurs de fonds



Federal Ministry
for Economic Cooperation
and Development



Lacuna Fund's Structure de gouvernance

Comité de pilotage

Orientation stratégique et gouvernance globale du Fonds

Collaborateur du bailleur de fonds

Fournit au comité de pilotage des informations sur les priorités essentielles et constitue un forum pour une collaboration plus étroite avec les bailleurs de fonds.

Groupes consultatifs techniques

Étudier les processus de financement et sélectionner les propositions, fournir des informations sur les stratégies et les besoins spécifiques au domaine.

Secrétariat

Gérer les subventions et les rapports, assurer la facilitation, la communication et le soutien opérationnel.



Domaines pour les ensembles de données

Le Fonds Lacuna fournit des ressources pour la création, l'expansion ou la maintenance d'ensembles de données.

Appels à propositions en cours dans ces domaines :

- **Résistance aux antimicrobiens**
- **Traitement du langage naturel (NLP)**

Appels à propositions antérieurs dans ces domaines :

- Agriculture
- Langue
- Santé
- Climat

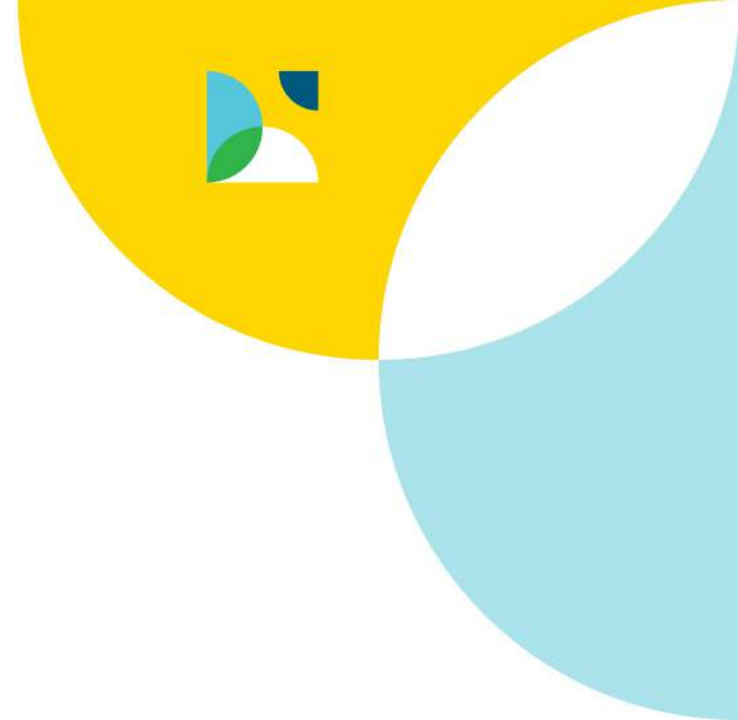
Futurs appels à propositions :

- Autres domaines dans lesquels le comité de pilotage identifie un besoin



Bailleurs de fonds pour l'AMR

L'appel à propositions AMR est rendu possible grâce au soutien généreux de WELLCOME.





Exigences pour l'appel d'offres

Résistance aux antimicrobiens (AMR)

Appel d'offres AMR - Objet

Objectif déclaré :

L'objectif de cet appel à propositions est de soutenir les efforts visant à développer des ensembles de données ouverts et accessibles pour les applications d'apprentissage automatique liées à la compréhension de la résistance aux antimicrobiens (RAM) dans les pays à revenu faible et intermédiaire (PRFI) d'Afrique, d'Amérique latine, d'Asie du Sud et d'Asie du Sud-Est.

Appel d'offres AMR - Besoin

Nous estimons qu'il est nécessaire de disposer d'ensembles de données de formation et d'évaluation qui permettront de mieux comprendre les effets de la RAM et de mettre au point des interventions visant à atténuer ces effets. Nous recherchons des ensembles de données identifiés par des experts locaux et conçus pour répondre à des besoins identifiés localement ; les exemples suivants ne sont donc donnés qu'à titre indicatif.

- Des ensembles de données pour améliorer la prévention et le contrôle des infections (IPC)
- Ensembles de données pour la découverte de médicaments
- Ensembles de données pour la surveillance mondiale de la résistance aux antimicrobiens et les approches anticipatives
- Ensembles de données pour le diagnostic
- Ensembles de données contenant des informations sur l'accès aux antimicrobiens et leur absence
- Ensembles de données améliorant la compréhension biologique fondamentale des agents pathogènes et de la résistance aux antimicrobiens
- Dans tous les ensembles de données, la sensibilité au genre, l'inclusion des principaux groupes vulnérables, en particulier ceux qui subissent le plus les effets de la RAM, ont été prises en compte.



Ce que nous recherchons

- Collecte et annotation de nouvelles données
- Annoter et publier des données existantes
- Élargissement et/ou maintenance des ensembles de données existants et amélioration de la facilité d'utilisation
- Relier et harmoniser les ensembles de données existants



Composantes de la proposition

- Informations sur le demandeur
- Exposé des motifs de la proposition
- Calendrier du projet
- Budget et exposé des motifs du budget
- Formulaire d'auto-déclaration de diligence raisonnable



Conditions d'éligibilité

- Les exigences spécifiques des bailleurs de fonds sont les suivantes
 - être une entité à but non lucratif, une institution de recherche, une entreprise sociale à but lucratif ou une équipe de ces organisations. Les individus doivent poser leur candidature par l'intermédiaire d'un sponsor institutionnel. Les partenariats sont vivement encouragés afin de renforcer la collaboration et de maximiser les avantages découlant de l'utilisation des ensembles de données, mais seul le candidat principal recevra des fonds.
 - L'entité doit avoir pour mission de soutenir le bien de la société, au sens large.

Conditions d'éligibilité

- Avoir son siège dans le pays ou la région où les données seront collectées. Le candidat principal doit être basé dans un pays à revenu faible ou intermédiaire et doit proposer un travail dans le même pays ou dans un pays à revenu faible ou intermédiaire d'Afrique, d'Amérique latine ou d'Asie du Sud ou du Sud-Est. La liste des pays éligibles est disponible à l'adresse suivante : <https://wellcome.org/grant-funding/guidance/low-and-middle-income-countries>.
- Les institutions basées dans d'autres pays ou régions peuvent poser leur candidature en tant que partenaires de l'institution chef de file. Comme indiqué ci-dessus, seul le candidat principal recevra des fonds.

Conditions d'éligibilité

- Disposer de toutes les autorisations nationales ou autres nécessaires pour mener à bien la recherche proposée. La procédure d'approbation peut être menée parallèlement à la demande de subvention, si nécessaire. Les frais d'approbation, le cas échéant, sont à la charge du candidat.
- Avoir la capacité technique - ou la capacité de développer cette capacité par le biais d'un partenariat décrit dans la proposition - de procéder à l'étiquetage, à la création, à l'agrégation, à l'expansion et/ou à la maintenance des ensembles de données, y compris la capacité d'appliquer les meilleures pratiques et les normes établies en matière de résistance aux antimicrobiens pour permettre à de multiples entités d'effectuer des analyses de haute qualité en matière d'IA/ML.

Dates clés

27 juin : Appel à propositions ouvert
- [Postuler - Lacuna Fund](#)

12 juillet : Date limite pour les questions.

Veillez soumettre vos questions à secretariat@lacunafund.org

29 juillet : Publication des réponses

26 juillet : Dernière chance de poser sa candidature pour un mentorat

23 août : Date limite de dépôt des propositions complètes

- **Novembre** : Le groupe consultatif technique décide de l'attribution des subventions
- **Décembre** : Les boursiers sont informés
- **Mai** : Annonce des prix



Éléments d'une proposition solide



Éléments d'une proposition solide



- **Équipe pluridisciplinaire** - expérience en science des données et en apprentissage automatique, AMR
- **Cas d'utilisation et engagement de la communauté**
- **Un énoncé clair du problème** et la manière dont l'ensemble de données ou l'agrégation proposés aideront à résoudre le problème.
- Spécificité concernant la taille de l'ensemble de données - la **taille et la qualité** doivent être **suffisantes** pour être utiles dans les applications futures.
- **Le budget** est adapté à la taille de l'ensemble de données produit

Éléments d'une proposition solide



- Considérations relatives à **l'équité** (genre, statut socio-économique, ethnicité, etc.)
- Prise en compte et planification des problèmes potentiels liés à la **protection de la vie privée et à l'éthique**
- Planifier la **gestion des données et l'octroi de licences**
- Travailler dans plusieurs domaines lorsque cela est possible et pertinent : temps, régions, agents pathogènes, résultats pour les patients...
- **Le budget** est adapté à la taille de l'ensemble de données produit

Engagement communautaire

- Décrire les consultations antérieures et/ou la collaboration proposée avec les bénéficiaires prévus.
 - Quand et où vous avez rencontré / rencontrerez les partenaires
 - Comment les partenaires seront-ils/elles impliqués ?
 - Identifier les besoins en données
 - Collecte de données, étiquetage
 - Gouvernance des données
 - Utilisation des données
 - Comment les partenaires bénéficieront-ils/elles de l'ensemble de données nouveau/élargi ?



Vie privée et éthique

Expliquez comment votre équipe abordera la question :

- a) le respect de la vie privée,
- b) le risque d'utilisation abusive en aval,
- c) les éventuels vecteurs de discrimination (par exemple, le sexe), et
- d) des conditions de travail justes et équitables, si des étiqueteurs rémunérés participent au projet.

Décrivez le processus que vous utiliserez pour détecter les problèmes éthiques potentiels (par exemple, un comité d'examen institutionnel, etc.)

Normes et partage des données

- Facile à trouver - facile à trouver sur une plateforme publique et largement utilisée
- Accessible - format ouvert (CCBY 4.0)
- Interopérabilité - format des données
- Réutilisable - métadonnées

- Durable - plan de maintenance
- Partagé - communauté d'utilisateurs engagée et plan de partage de l'ensemble des données une fois terminé

<https://www.go-fair.org/fair-principles/>

Budget et coûts autorisés

- Doit être soumis dans le modèle Excel.
- En cas de doute, expliquez !
- Les coûts pris en charge sont les suivants
 - Formation/renforcement des capacités ;
 - Collecte des données ; étiquetage ; AQ/CQ ou vérification ; et post-traitement ;
 - Modèle de base ;
 - Octroi de licences ;
 - Hébergement de l'ensemble des données ;
 - Publication des résultats en libre accès ;
 - Métadonnées ou déclaration de données ;
 - Les efforts de crowdsourcing, tels que les "label-a-thons".



Qualité des données

- Profiling des données - Il s'agit d'analyser les données pour en tirer des enseignements, notamment en ce qui concerne les types de données, les distributions, les valeurs manquantes, les valeurs aberrantes et les incohérences potentielles.
- Processus de nettoyage des données - Il s'agit de traiter les valeurs manquantes, de corriger les erreurs, de supprimer les valeurs aberrantes, de résoudre les incohérences et de normaliser les formats de données.
- Exactitude des données - vérifier l'exactitude des données en les comparant à des sources fiables ou à des connaissances d'experts.
- Complétude des données - S'assurer que l'ensemble des données contient tous les éléments nécessaires à la tâche d'apprentissage automatique.

Considérations sur la qualité des données

1. Hébergement et accessibilité

- La manière dont nous hébergeons les données et l'endroit où nous les hébergeons sont vraiment importants.
- L'utilisation d'un jeu de données peut dépendre de la manière dont il est hébergé et de l'endroit où il est hébergé.

Où votre jeu de données est-il actuellement hébergé ? Qu'est-ce qui a motivé cette décision ?

Hébergement et accessibilité

Quels sont les éléments à prendre en compte lors du choix d'une plateforme d'hébergement ?

- Domaine
- Cas d'utilisation
- Coût

Politique du Fonds Lacuna :

Flexible pour permettre la propriété locale des données et pour répondre aux cas d'utilisation envisagés pour les données. Guidés par les principes du Fonds Lacuna

Hébergement et accessibilité - Principes de Lacuna

| Accessibility



| Equity



| Ethics



| Participatory Approach



| Quality



| Transformational Impact





2. Documentation sur les ensembles de données

"Dans l'industrie électronique, il est courant d'accompagner chaque composant d'une fiche technique fournissant des caractéristiques de fonctionnement standard, des résultats de tests, des recommandations d'utilisation et d'autres informations.

De même, nous recommandons que chaque jeu de données soit accompagné d'une fiche de données documentant sa création, sa composition, ses utilisations prévues, sa maintenance et ses autres propriétés".

Avantages de la documentation des jeux de données



- Reproductibilité
- Transparence
- Facilité d'utilisation

La documentation de l'ensemble des données montrera le contexte, la structure et le contenu.

Documentation sur les jeux de données - Directives Lacuna

- Fiche technique
- file de métadonnées
- Identifier l'objet numérique (DOI)



3. Octroi de licences - Lignes directrices

- Politique de Lacuna en matière de propriété intellectuelle. Les ensembles de données et la propriété intellectuelle connexe développés avec des fonds de subvention seront :
- Être la propriété de l'entité sous-bénéficiaire
- faire l'objet d'une licence ouverte de la part de l'entité sous-bénéficiaire afin de maximiser le potentiel de diffusion responsable et d'utilisation en aval de cette propriété intellectuelle.



Licences... qu'est-ce qui est couvert ?

- Métadonnées, fiches techniques ou autres informations.
- Des conseils sont joints à l'ensemble de données pour en guider l'utilisation.
- Exemples de modèles ou autres outils d'information fournis avec les données.

Opportunité de mentorat : Programme de mentorat ACTS

Centre africain d'études technologiques (CAET)



Le programme de mentorat ACTS

- **Mission** : La mission d'ACTS est de renforcer les capacités et les politiques des pays et des institutions africains afin d'exploiter la science, la technologie et l'innovation pour le développement durable. Le programme de mentorat aidera les candidats à élaborer des candidatures de qualité et les guidera pour comprendre et suivre les meilleures pratiques de mise en œuvre de leurs projets.
- Une fois **n'est pas coutume** pour cet appel du Fonds Lacuna, ACTS offrira aux candidats du monde entier une opportunité spéciale de rejoindre le programme de mentorat, dans le cadre duquel les candidats et les équipes seront jumelés à un mentor.



Le programme de mentorat ACTS

- Postuler sur le **site ACTS**
[Programme de mentorat ACTS](#)
- Sélectionnez la participation "**Demande de mentorat**" **OU** "**Demande de mentorat**".
- Les deux formulaires comportent des champs similaires dans lesquels vous devrez fournir vos informations personnelles, votre fuseau horaire, votre langue, votre domaine de spécialisation, le domaine thématique de la proposition.
- Nous vous encourageons à préparer une ébauche détaillée ou un projet de proposition avant de poser votre candidature à un mentorat, ce qui vous permettra de connaître votre domaine thématique.
- Postulez au **moins 4 semaines avant la date d'échéance, avant le 26 juillet 2024**
- Les mentors seront attribués selon le principe du **premier arrivé, premier servi**.
- Tous les candidats sont tenus de **lire et de respecter le code d'éthique et de conduite du programme de mentorat**.

Le code de conduite de la FSCA

- Ne pas nuire
- Consentement libre, préalable et éclairé
- Confidentialité
- Sécurité
- Pas de coercition
- Participation des enfants (moins de 18 ans)
- Genre
- Inclusion

Le programme de mentorat ACTS

Rôles du mentor

- Fournir un mentorat et aider les candidats à comprendre l'objectif de l'appel.
- Établir des relations avec les mentorés et comprendre leurs faiblesses et leurs forces
- Aider les bénéficiaires à relire les propositions en fonction des besoins.
- Guider les mentorés sur les rôles et les conflits au sein du groupe qui pourraient survenir.
- Assister aux sessions ou aux discussions du groupe de proposition ou rencontrer séparément les chercheurs principaux.
- Discuter des travaux pertinents dans le domaine de la proposition afin de s'assurer que la proposition est contextualisée.





Des questions ?

Prochaines étapes

- **Soumettre des questions supplémentaires à secretariat@lacunafund.org avant le 12 juillet**
- **Réponses affichées** publiquement sur la page d'application du Fonds Lacuna le **29 juillet**
- **Postulez pour un mentorat dès que possible et avant le 26 juillet.**
- **Les propositions sont attendues le 23 août**