

Lacuna Fund : Our Voice on Data

Appel à propositions : Ensembles de données pour les langues en Afrique subsaharienne

23 Septembre 2020

1 – Introduction	0
Objet et objectifs du Fonds	1
Principes du Fonds	1
Philosophie en matière d’octroi de subventions	2
2 – Vue d’ensemble	3
Critères d’éligibilité applicables aux organisations	3
Processus de sélection et critères d’évaluation	3
Calendrier	5
3 – Objet et besoins	5
Objet	5
Besoins	6
4 – Informations relatives à la soumission des projets	8
Informations vous concernant	8
Exposé de la proposition	9
Lettres de soutien	11
Calendrier	11
Budget et explication du budget	12
Annexe A : formulaire d’autoévaluation du respect des obligations de diligence raisonnable	13

1 – Introduction

Cet appel à propositions financera la création, l'étiquetage, l'augmentation ou la tenue à jour d'ensembles de données pour l'apprentissage machine des langues en Afrique subsaharienne. Nous envisageons que ces ensembles de données rendront possibles des tâches spécifiques dans le traitement automatique du langage naturel (TALN) et contribueront à une recherche plus large dans le domaine de l'apprentissage machine avec pour but ultime de renforcer l'impact social.

Cet appel à propositions se termine le 6 novembre 2020. Les propositions ne seront acceptées que via le portail en ligne du site Web du Lacuna Fund (<https://lacunafund.org/fr/soumettre-un-projet/>). Le présent document contient davantage de détails sur les critères d'éligibilité, les critères de sélection, le calendrier, le processus de questions-réponses, l'objectif et les besoins ainsi que les exigences pour soumettre un projet.

Merci pour l'intérêt que vous portez au Lacuna Fund et pour votre engagement à combler les lacunes en matière de données afin de permettre à la communauté de l'apprentissage machine de mieux résoudre les problèmes urgents. Nous avons hâte de recevoir vos propositions !

Objet et objectifs du Fonds

Le Lacuna Fund soutient la création, le développement et la tenue à jour d'ensembles de données ouvertes et étiquetées* dans le respect des principes d'équité permettant une application fiable et plus équitable des outils d'apprentissage machine présentant une valeur sociale élevée dans les contextes de faibles et moyens revenus au niveau mondial.

Le Fonds poursuit plusieurs objectifs :

- Verser des fonds aux institutions en vue de créer, de développer et/ou de tenir à jour des ensembles de données qui combler les lacunes et réduisent les biais dans les données étiquetées utilisées pour l'apprentissage machine.
- Permettre aux populations mal desservies de profiter des avancées offertes par l'IA.
- Favoriser une meilleure compréhension par les communautés de l'apprentissage machine et du développement ainsi que les organisations philanthropiques de la manière de financer le plus efficacement et économiquement possible le développement et la tenue à jour d'ensembles de données étiquetées dans le respect des principes d'équité.

** Les données d'évaluation et de formation étiquetées et non étiquetées sont essentielles pour faire progresser le TALN en ce qui concerne les langues mal desservies. Le présent appel à propositions soutient la création, l'augmentation et la tenue à jour d'ensembles de données de formation et d'évaluation qui servent les objectifs mentionnés dans la section 3 – Objet et besoins.*

Principes du Fonds

Les principes suivants guident les opérations du Lacuna Fund.

- **Accessibilité** – Le Fonds s’engage à garantir que les ensembles de données étiquetées créés grâce à son financement sont accessibles de manière ouverte aux communautés mal desservies et leur bénéficient au service des objectifs énoncés précédemment. Les ensembles de données et la propriété intellectuelle dont ils font l’objet seront couverts par des licences de données ouvertes appropriées pour favoriser une utilisation responsable en aval. (Voir la [Politique en matière de propriété intellectuelle](#) du Fonds pour plus de détails.)
- **Équité** – Le Fonds vise à rendre l’IA plus équitable en créant des ensembles de données représentatifs de l’hémisphère Sud et des personnes de couleur dans le monde, ainsi que de leurs besoins. Ces ensembles de données ne doivent pas créer ou renforcer de préjugés.
- **Éthique** – Le Fonds financera la collecte de données d’une manière conforme aux normes éthiques du travail et exigera des bénéficiaires de subventions qu’ils décrivent les mesures qu’ils prendront pour protéger la vie privée et éviter tout préjudice dans la collecte, l’octroi de licences et l’utilisation des ensembles de données créés avec les fonds des subventions.
- **Approche participative** – Le Fonds s’efforce de répondre aux besoins des parties prenantes concernées en associant les développeurs, les chercheurs et les bénéficiaires locaux ainsi que les utilisateurs finaux à la gouvernance du projet ainsi qu’à la création de données.
- **Qualité** – Les données produites grâce aux efforts financés par Lacuna doivent être d’une qualité élevée permettant des applications bénéfiques pour la recherche, les communautés et l’industrie.
- **Impact transformationnel** – En finançant des ensembles de données qui comblent des lacunes fondamentales de l’IA, le Fonds entend libérer les avancées dont celle-ci est porteuse au bénéfice des communautés pauvres et mal desservies.

Philosophie en matière d’octroi de subventions

Le Lacuna Fund privilégie une approche collaborative et pilotée localement pour la création, le développement et la tenue à jour d’ensembles de données. Nous considérons que l’utilité et l’actualisation pérennes des données ouvertes imposent de s’appuyer sur une communauté investie et concernée par ces données.

Le Lacuna Fund espère financer des ensembles de données qui contribuent aux multiples applications à forte valeur sociale, que ce soit au travers de la recherche, de l’innovation commerciale ou de l’amélioration des services du secteur public. **Bien que la section 3 « Objet et besoins » expose les besoins définis par le Groupe consultatif technique (GCT), le Lacuna Fund accueille toutes les idées novatrices dans ce domaine mettant en évidence un avantage clairement articulé aligné sur les critères de sélection présentés ci-après.**

Le Lacuna Fund est soutenu par la Fondation Rockefeller, Google.org, le Centre de recherche pour le développement international du Canada et l’agence de coopération internationale allemande pour le

développement (GIZ) pour le compte du ministère allemand de la Coopération et du Développement économique (BMZ).

2 – Vue d’ensemble

Critères d’éligibilité applicables aux organisations

Le Lacuna Fund entend rendre son financement accessible à un maximum d’organisations dans l’univers de l’IA pour le bien social et cultiver les capacités et les organisations émergentes dans ce domaine.

Pour pouvoir bénéficier d’un financement, les organisations doivent remplir les critères suivants :

- Être une entité à but non lucratif, un institut de recherche, une entreprise à but lucratif ou une équipe composée de ce type d’organisations. Les particuliers doivent envoyer leurs propositions via un promoteur institutionnel. Les partenariats sont encouragés, mais seul le candidat principal recevra directement des fonds.
- Avoir une mission de soutien du bien sociétal, au sens large.
- Être une organisation dont le siège se situe en Afrique ou ayant un partenariat important avec une ou plusieurs organisations dont le siège se situe en Afrique.
- Disposer de toutes les autorisations requises, nationales ou autres, pour poursuivre les recherches proposées. Le cas échéant, le processus d’autorisation pourra être mené en parallèle de la proposition de subvention. Les éventuels frais d’autorisation sont à la charge du candidat.
- Disposer de la capacité technique pour procéder à l’étiquetage, à la création, au développement et/ou à la tenue à jour des ensembles de données, y compris la capacité à mettre en œuvre les bonnes pratiques et les normes établies dans le domaine visé (par exemple, le TALN) afin que plusieurs entités puissent utiliser des ensembles de données afin de créer des travaux analytiques de haute qualité pour l’IA/l’apprentissage machine.

Processus de sélection et critères d’évaluation

Le Lacuna Fund et ses partenaires procèdent à une sélection initiale sur la base des critères d’éligibilité applicables aux organisations et du respect de l’ensemble des exigences par la proposition. À l’issue de cette sélection initiale, un Groupe consultatif technique composé d’experts, d’utilisateurs de données et de parties prenantes pour le domaine visé évalue les propositions sur la base des critères de sélection énoncés ci-après. Les membres du Groupe consultatif technique ne peuvent pas répondre à un appel à propositions pour lequel ils sont évaluateurs (voir la [Politique du Lacuna Fund en matière de conflit d’intérêts](#)).

- **Qualité** – L’organisation ou l’équipe proposant le projet est constituée d’experts qualifiés en : a) apprentissage machine ; b) gestion des données et octroi de licences en matière de données ; et c) le cas échéant, le domaine d’intérêt.
- **Impact transformationnel** – Le projet : a) crée une valeur supplémentaire dans un ensemble de données existant ; b) crée un nouvel ensemble de données de formation de grande valeur à destination d’une population mal desservie ou pour apporter une réponse à un problème ayant trait aux [objectifs de développement durable \(ODD\)](#) ; c) améliore la représentativité d’un ensemble de données existant et la prise en compte par cet ensemble de données des contextes de revenus faibles et moyens ; ou d) améliore la durabilité d’un ensemble de données largement utilisé et conforme aux principes d’équité.
- **Équité** – Le projet est assorti d’un argumentaire convaincant en faveur du changement démontrant que les ensembles de données amélioreront l’apprentissage machine et seront mis en œuvre afin d’aider les populations pauvres, vulnérables ou d’autres populations mal desservies par la technologie linguistique actuelle.
- **Approche participative** – Lorsque l’ensemble de données a une portée géographique (par exemple, les ensembles de données linguistiques et géospatiales), l’équipe est majoritairement implantée dans la région concernée et/ou entretient des liens étroits avec les acteurs locaux pour garantir une tenue à jour et un usage pérennes de l’ensemble de données par la communauté locale.
- **Éthique** – Le projet peut passer avec succès un examen sur les questions d’éthique (par exemple, par un comité d’examen institutionnel) portant sur les aspects suivants : a) les questions liées au respect de la vie privée ; b) le risque d’utilisation abusive de l’ensemble de données en aval ; c) les vecteurs de discrimination potentiels (par exemple, la discrimination fondée sur le genre) ; et d) les conditions de travail justes et équitables si des étiqueteurs rémunérés participent au projet. (Voir le [Code d’éthique de l’ACM](#) pour les normes générales.)
- **Efficacité** – Le porteur de projet a tenu compte des ensembles de données existants et propose l’utilisation d’outils et de techniques efficaces de collecte et d’étiquetage de données permettant d’accélérer la collecte, le nettoyage, la tenue à jour et le partage des données.
- **Faisabilité** – Le projet est réalisable en fonction du budget et de l’étendue des travaux proposés.
- **Accessibilité** – L’ensemble de données sera largement mis à disposition dans le cadre d’une licence ouverte ou, en cas d’impossibilité, un argumentaire convaincant expose les raisons d’un système d’octroi de licence plus strict dans le but de protéger la vie privée ou de prévenir tout préjudice. Cela comprend un plan sur la manière de garantir que la source des données à partir de laquelle l’ensemble de données est créé n’est pas soumis à une licence ou à un droit d’auteur qui empêche de rendre l’ensemble de données accessibles via une licence ouverte.

- **Durabilité** – Le projet comporte un plan visant à garantir la durabilité et la tenue à jour future de l'ensemble de données, par exemple par une communauté spécifique ou un groupe de parties intéressées (organisations à but lucratif ou non lucratif), et un modèle de gouvernance robuste pour l'ensemble de données ouvert.

Calendrier

Calendrier

Appel à propositions diffusé publiquement sur le site Web de Lacuna Fund	23 septembre 2020
Date limite pour les questions/réponses Veuillez adresser vos questions à secretariat@lacunafund.org	7 octobre 2020
Réponses anonymes diffusées publiquement sur le site Web de Lacuna Fund	14 octobre 2020
Date limite pour le dépôt des candidatures	6 novembre 2020

Période de questions/réponses : toutes les questions concernant cet appel à propositions doivent être adressées par courrier électronique à secretariat@lacunafund.org en précisant « Question Appel à propositions Langues 2020 » dans l'objet. Les questions adressées d'ici le 7 octobre seront anonymisées et leur réponse sera publiée le 14 octobre dans un document posté sur la [page « Soumettre un projet »](#) du site Web de Lacuna Fund.

3 – Objet et besoins

Cet appel à propositions financera des ensembles de données étiquetées et non étiquetées pour permettre des travaux nouveaux et améliorés dans le domaine du traitement automatique du langage naturel (TALN) pour les langues d'Afrique subsaharienne. La mission du Lacuna Fund est de soutenir la création, l'étiquetage, le développement et la tenue à jour d'ensembles de données porteurs de changement afin de combler les lacunes et d'améliorer l'exactitude dans le domaine de l'apprentissage machine.

Objet

La capacité à communiquer et à être compris dans sa propre langue est un droit fondamental et une condition préalable à l'inclusion numérique et sociétale. Les techniques de traitement automatique du langage naturel (TALN) ont permis des applications essentielles à cet effet : par exemple, améliorer

l'éducation, l'inclusion financière, les soins de santé, l'agriculture, la communication et la réponse aux catastrophes.

Cependant, les ensembles de données mis à la disposition du public sont rares ou n'existent parfois même pas pour la plupart des langues africaines¹. Quand des ensembles de données existent, ils s'appuient souvent sur des textes religieux, missionnaires ou judiciaires, produisant alors une langue désuète et créant des préjugés. À travers les textes, les discours et d'autres ensembles de données, les lacunes dans les ensembles de données disponibles de manière ouverte ont empêché de réaliser des avancées grâce aux technologies du TALN. Les recueils de données et de discours étiquetés restent un élément clé de ces lacunes, de même que la disponibilité des données qui peuvent être utilisées dans l'apprentissage par transfert ou les approches semi-supervisées.

Cet appel à propositions commencera à répondre à ce besoin de données linguistiques, en incluant des données étiquetées et des recueils qui comblent d'autres lacunes clés liées aux données. Les efforts de Lacuna Fund dans le domaine du TALN s'appuient sur une dynamique récente en faveur de la création d'outils de TALN meilleurs et plus ouverts dans les langues africaines, instillée par les membres de la communauté de l'apprentissage machine, dans le cadre, notamment, d'ateliers académiques, de collaborations volontaires, de programmes académiques innovants et d'autres efforts récents.

Besoins

Cet appel à propositions vise à financer largement les ensembles de données d'évaluation et de formation ouvertes pour le TALN des langues africaines. Le GCT reconnaît la diversité et l'importance des besoins en matière de TALN des langues africaines ainsi que le besoin d'avoir des ensembles de données multilingues. Les propositions doivent faire progresser l'état actuel des données et augmenter le potentiel de développement des outils de TALN dans la ou les langues pour lesquelles les efforts sont proposés.

Le GCT reconnaît le besoin d'avoir des ensembles de données qui permettent une meilleure exécution des tâches clés du TALN dans les langues africaines, notamment, mais pas uniquement, les éléments suivants :

- les recueils de discours, notamment la reconnaissance vocale automatisée qui permet aux personnes analphabètes et à d'autres groupes de personnes défavorisées d'avoir accès à des informations et/ou à des services ;
- les recueils de textes étiquetés et non étiquetés pour les utiliser comme données de formation ;
- les corpus multilingues pour la traduction automatique ;
- les recueils contribuant aux tâches basiques du TALN, comme la reconnaissance d'entités nommées, l'étiquetage morphosyntaxique, le plongement de mots, etc. ;

¹ <https://arxiv.org/pdf/2004.09095.pdf>

- les ensembles de données contribuant aux tâches essentielles du TALN en aval, comme la réponse aux questions et l'IA conversationnelle, les ensembles de données d'analyse des sentiments ou la technologie pour l'éducation des langues ;
- les ensembles de données permettant d'améliorer l'exécution des tâches du TALN sur des textes ou des discours avec alternance codique.

Plus généralement, il faut également :

- une augmentation des ensembles de données existants dans tous les domaines pour réduire les préjugés (comme les préjugés sexistes ou d'autres types de préjugés ou de discriminations) ou une amélioration de la facilité d'utilisation des technologies de TALN dans des contextes de revenus faibles ou moyens ;
- plus de données de référence pour les tâches de TALN dans les langues mal desservies ou pour étoffer les modèles multilingues ;
- des ensembles de données innovants, comme le sous-titrage audio ou vidéo ou d'autres interactions image-texte ;
- la création ou l'augmentation d'ensembles de données de textes ou de discours propres à un domaine, comme des ensembles de données numériques, des noms d'endroits ou des combinaisons de mots ou de phrases spécifiques qui permettent des applications avec un impact social important.
 - À titre d'exemple, une application potentielle pourrait être le repérage par mot clé à la radio pour les interventions humanitaires, la reconnaissance vocale pour les intrants agricoles ou un autre besoin qui exige un ensemble de données spécifique.

Veuillez noter que cet appel à propositions soutient uniquement la création, le développement et la tenue à jour de données d'évaluation et de formation dans le domaine de l'apprentissage machine. Bien que cela comprenne le soutien à des approches d'apprentissage semi-supervisées et actives pour la collecte ou la production de données, Lacuna Fund est incapable de soutenir le développement d'une application ou d'un modèle.

Même si le Groupe consultatif technique a défini certains besoins clés dans le domaine mentionné ci-dessus, les candidats doivent se sentir libres de proposer des idées novatrices dans le domaine mentionné dans la section « Objet » et conforme aux critères d'évaluation de cet appel à propositions (voir « Philosophie en matière d'octroi de subventions » à la page 3 pour plus de détails).

Toutes les propositions doivent :

- Étudier la manière de garantir que les données collectées sont représentatives et, si possible, impartiales.
- Garantir le respect des dispositions légales et préciser les droits d'auteur ou de propriété intellectuelle pertinents pour tout texte source.

- Utiliser les infrastructures existantes ou communes (voir la page « Ressources » sur le site Web de Lacuna Fund pour une liste non exhaustive).
- Envisager des moyens innovants et rentables pour collecter et tenir à jour des données de qualité, notamment dans le cadre de la collecte communautaire, de la production par les pairs sur une base commune, d'autres formes de crowdsourcing, de la gamification et de l'utilisation de techniques d'apprentissage actif.
- Démontrer une connaissance des ensembles de données existants et tenter d'assurer l'interopérabilité avec ces ensembles de données.
- Veiller à ce que vos métadonnées soient bien décrites. (Voir « [Déclarations de données pour le traitement du langage naturel](#) » pour plus d'informations sur les meilleures pratiques. Une déclaration de données n'est PAS nécessaire pour cette proposition.)

4 – Informations relatives à la soumission des projets

REMARQUE : le site Web de Lacuna Fund propose [diverses ressources](#), notamment des références pertinentes sur la qualité des données, la documentation et le format, afin d'aider les soumissionnaires à préparer un dossier de candidature concurrentiel.

La soumission des propositions doit être faite exclusivement sur le portail spécifique disponible à l'adresse <https://lacunafund.org/fr/soumettre-un-projet/>. Une description des questions relatives aux candidatures est proposée ci-après à titre d'information uniquement. **Les sections suivantes comprennent :**

- des informations concernant le candidat (accessibles sur le portail) ;
- un exposé de la proposition ;
- des lettres de soutien (facultatives, obligatoires si l'utilisation d'un ensemble de données de source non publique est proposée) ;
- le calendrier, le budget et une explication du budget ;
- un formulaire d'autoévaluation du respect des obligations de diligence raisonnable.

Informations vous concernant

Dans cette section, le candidat est invité à fournir :

- un résumé de la proposition de 200 à 250 mots ;
- des détails concernant la ou les institution(s) et/ou l'équipe présentant la proposition ;
- l'endroit où le travail aura lieu ;
- les CV des membres essentiels de l'équipe.

Dans le futur, le candidat devra fournir des informations concernant :

- les processus de contrôle éthique de la ou des institutions affiliées ;
- la capacité de l'équipe à obtenir les approbations nationales.

Exposé de la proposition

Veillez limiter l'exposé de votre proposition à 10 pages, références non incluses, avec des marges de 2,5 cm et une police de caractères de 11 points minimum. Les annexes ou supports de l'exposé de la proposition au-delà des 10 pages ne seront pas examinés.

Dans cette section, le candidat est invité à télécharger un exposé cohérent, au format PDF ou Word, traitant des aspects décrits ci-après :

Qualifications – Décrivez la ou les organisations ou également le ou les partenariats présentant la proposition ainsi que les qualifications particulières pour entreprendre le travail proposé.

Ensemble de données et cas d'utilisation proposés : veuillez résumer brièvement l'ensemble de données que vous souhaitez créer, augmenter ou tenir à jour et le besoin spécifique en matière d'apprentissage machine que l'ensemble de données satisferait dans le contexte de la ou des langues particulières ou également des tâches de TALN.

Votre ensemble de données permettrait-il, par exemple, à d'autres personnes de construire une partie de l'étiquetage morphosyntaxique dans une langue sans outils basiques ou pourrait-il servir de référence améliorée pour un modèle de traduction automatique dans une langue où certaines données sont déjà disponibles, mais où les modèles sont difficiles à évaluer ? Donnez-nous votre solution dans le contexte du ou des besoins spécifiques. Voir section 3 : « Objet et besoins » pour plus de détails.

Caractéristiques et produits à livrer en ce qui concerne les données et la documentation proposées – Veuillez fournir des informations complémentaires sur les efforts que vous proposez et expliquer comment la quantité/qualité des données proposée, les méthodes de collecte et d'autres détails rendent les données utilisables dans le contexte opérationnel de cas d'utilisation potentiels. Veuillez inclure les éléments suivants, suffisamment détaillés pour que le Groupe consultatif technique puisse évaluer votre proposition en ce qui concerne la faisabilité, le potentiel de transformation et la fiabilité technique :

- quantité, types et format des données et/ou des étiquettes ainsi qu'une base et une taille d'échantillon ou un programme permettant de garantir la représentation, le cas échéant ;
- des techniques d'étiquetage et de collecte de données proposées et des informations sur l'interopérabilité, en tenant compte des infrastructures communes ou existantes ;
- des programmes permettant d'évaluer et d'atténuer les erreurs et les préjugés (par exemple, les préjugés sexistes ou d'autres préjugés) ;

- des indicateurs utilisés pour évaluer les résultats souhaités de la création de données (par exemple, indicateurs d'équité dans l'ensemble de données, AQ/CQ par rapport à un référentiel, etc.) ;
- toutes les difficultés prévues ou les incertitudes concernant la collecte de données et les contremesures proposées.

Voie(s) vers l'impact et les bénéficiaires visés – Expliquez comment l'étiquetage, la création, l'augmentation ou la tenue à jour de l'ensemble de données proposé contribuera à atteindre un impact (par exemple, en utilisant le cadre des objectifs de développement durable). Décrivez les bénéficiaires prévus de la ou des applications permises par l'ensemble de données et donnez un aperçu des consultations précédentes et/ou de la collaboration proposée avec les bénéficiaires prévus.

Le cas échéant, décrivez comment les produits peuvent susciter des pistes de recherche ou des applications à but non lucratif ou commerciales multiples et durables. Par exemple, associer [des données de traduction parallèles avec des images](#) pourrait permettre des recherches futures dans différents domaines. Soulignez toutes les contraintes pratiques auxquelles cette voie pourrait être confrontée (par exemple, pénétration d'Internet).

Accessibilité, gestion des données et octroi de licence – Veuillez décrire :

- les plans concernant le stockage, la propriété et l'accessibilité des données, conformément aux principes de Lacuna. Par exemple, le format de données proposé doit être bien documenté et inclure les informations nécessaires à l'utilisateur final pour utiliser les données ;
- tout problème anticipé lié aux droits d'auteur pour les données sources et à la collaboration avec le détenteur des droits d'auteur ;
- les plans concernant l'octroi de licences afin de favoriser une utilisation responsable en aval. Selon les principes de Lacuna Fund, l'ensemble de données et toute propriété intellectuelle y afférente, par exemple les méthodes de collecte, les fiches de données, la manière de charger ou de lire les ensembles de données ou les autres informations permettant de garantir l'exploitabilité doivent être mis à disposition dans le cadre d'une licence ouverte par attribution (CC-BY 4.0 ou similaire). Si un système d'octroi de licence plus restrictif est proposé, il convient de fournir une justification ;
- si vous avez l'intention d'utiliser un ensemble de données existant pour votre projet, veuillez indiquer si votre équipe a reçu les autorisations nécessaires de la part du propriétaire de l'ensemble de données, selon lesquelles cet ensemble de données peut être diffusé conformément à [la politique de Lacuna Fund en matière de propriété intellectuelle](#) ou fournir un justificatif pour une autre structure de licence.

Risques, incluant l'éthique et la confidentialité – Déterminez les risques potentiels, notamment, mais pas uniquement, en matière de confidentialité ou d'éthique et décrivez les mesures envisagées pour atténuer ces risques. Notamment :

- Décrivez comment vous comptez obtenir un consentement éclairé le cas échéant (cela doit inclure la notification des cas d'utilisation future potentiels des données).

- Décrivez comment vous comptez garantir l'équité au sein du personnel du projet, notamment, mais pas uniquement, en matière de rémunération équitable pour l'étiquetage et la collecte ainsi que la parité pour les agents de terrain.
- Présentez un plan pour l'anonymisation des informations personnelles identifiables (IPI) et la conformité avec les lois relatives au respect de la vie privée ; le cas échéant. En l'absence de cadre juridique national, la proposition doit indiquer les bonnes pratiques correspondantes ou y faire référence.
- Discutez des effets négatifs potentiels lors de la production et de l'utilisation de l'ensemble des données et des étapes pour les atténuer.

Plan de durabilité – Décrivez comment les ensembles de données étiquetées seront tenus à jour, intégrés et/ou développés au-delà du financement initial (par exemple, grâce aux applications d'AM issues du projet, par une communauté spécifique ou un groupe de parties intéressées doté d'un modèle de gouvernance robuste pour l'ensemble de données ouvertes).

Considérations liées à la pandémie de COVID-19 – Veuillez indiquer si vous avez besoin d'adapter votre travail en raison des restrictions liées à la COVID-19. Dans l'affirmative, décrivez les mesures envisagées pour vous adapter aux restrictions et, dans le cas du travail de terrain, pour garantir la sécurité du personnel du projet.

Lettres de soutien

Des lettres de soutien sont nécessaires de la part du propriétaire de tout ensemble de données existant qu'une proposition a l'intention d'utiliser ou d'annoter, si les informations sur l'octroi de licences ouvertes pour cet ensemble de données ne sont pas déjà mises à la disposition du public (Par exemple, si une proposition a pour but de numériser des articles de journaux, une lettre de soutien de la part de l'éditeur du journal est nécessaire ; mais si l'ensemble de données source est déjà mis à la disposition du public et/ou si les conditions de la licence sont claires, une lettre de soutien n'est pas nécessaire). La lettre doit indiquer que l'ensemble des données produites peuvent être diffusées conformément à [la politique de Lacuna Fund en matière de propriété intellectuelle](#).

Les lettres de soutien des institutions collaboratrices sont facultatives. Si elles ne sont pas téléchargées sur le portail des candidatures, elles seront demandées plus tard aux auteurs des propositions acceptées.

Calendrier

Dans cette section, le candidat est invité à soumettre un calendrier pour l'exécution des principales activités et des livrables du projet, sous forme de tableau. Le calendrier peut inclure, sans toutefois s'y limiter, la formation du personnel, la collecte de données, l'étiquetage, l'assurance qualité, la validation/le nettoyage et la publication des données. Parmi les livrables figurent par exemple, de manière non exhaustive, les portions de l'ensemble de données servant à démontrer la faisabilité, l'ensemble de données complet et la documentation annexe ou les méthodes de collecte à diffuser de manière ouverte.

Tous les calendriers doivent indiquer la date à laquelle les données seront disponibles publiquement avec toute la documentation. Les ensembles de données ne peuvent pas être visés par un embargo jusqu'à la publication.

Les projets proposés doivent être achevés dans un délai de 12 mois à compter de la date de notification de l'attribution (les notifications sont prévues pour le mois de décembre 2020/janvier 2021 approximativement).

Budget et explication du budget

Le portail de candidature en ligne demandera au candidat de télécharger un budget formaté suivant le modèle de budget de Lacuna Fund et qui est disponible sur la [page de candidature](#) sous « Appels à projets ouverts » et sur le portail du candidat. Les budgets doivent être présentés en dollars américains (USD). Nous prévoyons des budgets dans une fourchette de 10 000 à 50 000 USD pour les projets de petite à moyenne taille, et des budgets pouvant atteindre 250 000 USD pour les projets de grande ampleur et complexes. Il convient de noter que le Groupe consultatif technique évaluera la faisabilité et la pertinence du budget ainsi que sa cohérence avec l'exposé de la demande de subvention dans le cadre des critères de sélection. Sans s'y limiter, les budgets incluront les coûts concernant :

- le renforcement des capacités ayant trait à la collecte des données et à l'assurance/au contrôle qualité ;
- la collecte des données ;
- l'étiquetage des données ;
- l'AQ/CQ ou la vérification, dont l'engagement de linguistes pour réaliser l'étiquetage ou garantir la qualité ;
- le post-traitement des données ;
- la publication des données ;
- l'octroi de licences, dont des conseils juridiques pour garantir le respect des droits d'auteur et de la propriété intellectuelle ;
- les coûts liés aux collaborations et aux consortiums ;
- les méthodes communautaires de collecte de données ou les efforts de crowdsourcing, comme les « label-a-thons ».

Les financements ne peuvent **pas** être utilisés pour payer directement des droits de douane, d'importation ou d'autres droits ou taxes perçus à l'importation des marchandises ou équipements dans un pays ou territoire. *Les coûts indirects ne doivent pas excéder 12 % du budget proposé.*

Voir la fiche d'instructions dans le modèle de budget pour de plus amples informations concernant les directives budgétaires, dont les informations sur les coûts de personnel admissibles.

Explication du budget : le modèle de budget prévoit un espace permettant d'indiquer toute hypothèse budgétaire ou de fournir davantage de contexte.

Annexe A : formulaire d'évaluation du respect des obligations de diligence raisonnable

Ce formulaire et tout autre document doivent être transmis via le portail de soumission des candidatures de Lacuna Fund, accessible à l'adresse lacunafund.org/apply.

Le Meridian Institute doit se conformer aux exigences applicables aux contributeurs du fonds ainsi qu'aux exigences légales applicables aux organisations offrant un financement qui sont basées aux États-Unis.

Dans la mesure où la loi le permet, toutes les informations fournies sont conservées et traitées de façon confidentielle et ne peuvent être divulguées à des tiers sans notification et accord préalables.

1. Veuillez indiquer la forme juridique de votre organisation et la législation/réglementation applicable. Veuillez également indiquer le mois et l'année de début d'activité de votre organisation sous cette entité.
2. Votre organisation est-elle dotée d'une politique relative aux conflits d'intérêts et existe-t-il des cas de conflit d'intérêts dont le Meridian Institute devrait avoir connaissance ?
3. Un fonctionnaire de l'État ou une entité gouvernementale ont-ils un intérêt financier, de gestion ou une participation majoritaire dans votre société/organisation ? Dans l'affirmative, fournissez des détails ainsi que leur niveau d'intérêt ci-dessous.
4. Veuillez indiquer les noms complets de tous les dirigeants de votre société/organisation. (Note : « dirigeant » désigne les cadres, associés, propriétaires, administrateurs, mandataires ou autres personnes exerçant un contrôle quelconque sur votre société/organisation.)
5. Votre société/organisation emploie-t-elle des fonctionnaires de l'État ? Dans l'affirmative, veuillez indiquer ci-dessous ou dans une page ou un document distinct les noms et postes occupés par les fonctionnaires de l'État en question.
6. Un dirigeant de votre organisation a-t-il un parent proche fonctionnaire de l'État (un parent proche désigne une mère, un père, une sœur, un frère, un(e) époux(-se) ou un enfant) ? Dans l'affirmative, veuillez indiquer ci-dessous ou dans une page ou un document distincts le nom, le lien de parenté, le titre, les responsabilités et le service gouvernemental ou l'agence de l'État dans le(la)quel(le) travaille(nt) le ou les fonctionnaire(s) de l'État en question.
7. Votre société/organisation, une filiale ou entité affiliée de votre société/organisation ou un dirigeant de votre société/organisation ont-ils déjà été condamnés pour un crime ou ont-ils déjà fait l'objet d'une enquête pénale, d'une mise en examen ou d'une procédure similaire ? Dans l'affirmative, fournissez des informations détaillées ci-dessous.
8. Au cours des trois dernières années, votre société/organisation ou toute filiale ou entité affiliée de votre société/organisation ont-elles fait l'objet d'un contentieux terminé ou en cours ou d'une enquête gouvernementale ? Dans l'affirmative, fournissez des informations détaillées ci-dessous.

9. Y a-t-il des informations complémentaires concernant votre société/organisation ou tout employé de votre société/organisation susceptibles d'aider le Meridian Institute à s'acquitter de ses obligations de diligence raisonnable en matière de lutte contre la corruption ? Dans l'affirmative, fournissez des informations détaillées ci-dessous.
10. La personne soussignée, étant dûment autorisée à répondre à ce questionnaire et à fournir de nouvelles assurances concernant les questions suivantes, certifie ce qui suit :
- a) À ma connaissance, toutes les informations fournies dans le présent formulaire de diligence raisonnable anticorruption sont sincères, exactes et complètes.
- b) J'ai pris connaissance des informations concernant la loi anticorruption de 2010 du Royaume-Uni (<http://www.justice.gov.uk/downloads/legislation/bribery-act-2010-guidance.pdf>), le code pénal australien (<http://www.oecd.org/daf/anti-bribery/anti-briberyconvention/2027148.pdf>) et la loi américaine relative à la corruption dans les transactions à l'étranger (<http://www.sec.gov/investor/alerts/fcpa.pdf>), sur les sites Web indiqués, et je connais les exigences de ces textes légaux relatifs à la lutte contre la corruption.
- c) J'ai lu et pleinement compris le code de conduite du Meridian Institute et, à titre de condition pour travailler avec le Meridian Institute, j'accepte de me conformer à ce code de conduite.
- d) Ni moi ni ma société n'avons jamais versé d'argent, approuvé un tel geste ou autrement offert quelque chose de valeur à un fonctionnaire de l'État, directement ou indirectement, à des fins illégitimes, de corruption ou illégales, et n'avons aucune intention de le faire à l'avenir.
- e) Ni moi ni ma société n'avons jamais créé une fausse facture ou autrement manipulé des documents dans le but d'effectuer un paiement déguisé à un fonctionnaire de l'État à toutes fins, et n'avons aucune intention de le faire à l'avenir.
- f) Mes/nos pratiques comptables décrivent à tout moment avec précision et de manière exhaustive tous les paiements effectués, y compris les paiements en faveur de fonctionnaires de l'État.

NOTE : « Fonctionnaire de l'État » désigne toute personne, élue ou désignée pour occuper une charge ou une fonction exécutive, législative, administrative ou judiciaire dans une entité publique, y compris une agence internationale. En outre, « Fonctionnaire de l'État » inclut toute personne assurant des fonctions publiques dans toute branche du gouvernement national, fédéral, local ou municipal de tout pays ou territoire ou qui exerce une fonction publique, dans le cadre d'un emploi ou en sous-traitance, pour toute entité, agence ou entreprise publique de ce pays ou territoire, y compris les entreprises appartenant à l'État ou contrôlées par l'État. La définition de « Fonctionnaire de l'État » inclut également tout représentant d'un parti politique ou tout candidat à une fonction politique.