

# Présence de la science et de la science ouverte sur Wikipédia et les projets Wikimedia

Wikimédia France recherche une prestation pour la collecte et l'analyse de données issues de la recherche sur Wikipédia.

## 1. Présentation du projet

### 1.1 L'association

Wikimédia France est l'association française, créée le 23 octobre 2004, qui œuvre pour le libre partage de la connaissance au travers des projets Wikimedia (<https://www.wikimedia.fr/les-projets-wikimedia/>) (Wikipédia, le Wiktionnaire, Wikimedia Commons, Wikidata, etc.) Pour remplir cet objectif, elle se fixe trois missions :

- L'accroissement des contenus (en qualité et en quantité) et pour cela, elle organise des journées contributives ou des opérations libres, elle aide à la numérisation et au téléversement. Dans cette perspective, elle signe des partenariats avec des organisations aux objets sociaux divers (musées, universités, collectivités territoriales...).
- L'animation de communautés intéressées par les projets Wikimedia et à cette fin, elle soutient les contributeurs dans leurs projets individuels ou collectifs, en proposant des moyens humains, financiers et matériels.
- La montée en compétences de publics-cibles et pour cela, elle propose des formations à destination des acteurs éducatifs, culturels et du grand public et des bénévoles sur différentes thématiques comme les usages de Wikipédia en classe et la contribution aux projets Wikimedia.

## 1.2 Enjeux

Wikimédia France et le Ministère de l'enseignement supérieur et de la recherche sont engagés depuis 2022 dans un partenariat autour de la science ouverte. (<https://www.ouvrirlascience.fr/un-partenariat-pour-renforcer-les-liens-entre-wikipedia-et-lenseignement-superieur-et-la-recherche-francaise-dans-un-contexte-de-science-ouvert/>).

Dans le cadre du 2e plan national pour la science ouverte ([https://www.enseignementsup-recherche.gouv.fr/sites/default/files/content\\_migration/document/2e-plan-national-pour-la-science-ouverte-2021-2024-7794.pdf](https://www.enseignementsup-recherche.gouv.fr/sites/default/files/content_migration/document/2e-plan-national-pour-la-science-ouverte-2021-2024-7794.pdf)), les projets Wikimédia et notamment Wikipédia sont cités comme exemples permettant d'évaluer l'impact sur la société de la consultation en accès libre des travaux de recherche (thèses, article de revue, études...).

Le ministère met en ligne un baromètre (baromètre français de la science ouverte <https://barometredelascienceouverte.esr.gouv.fr/>) afin de consulter l'évolution de la science ouverte en France. Ce baromètre présente des indicateurs issus de plusieurs sources et jeux de données (<https://barometredelascienceouverte.esr.gouv.fr/a-propos/methodologie>) et est mis à jour annuellement.

## 1.3 Périmètre

Afin de mieux cerner l'impact des travaux de recherche à travers Wikipédia, le MESR et Wikimédia France engagent un travail d'analyse et de collecte de données en vue d'alimenter le baromètre de la science ouverte avec des données provenant de Wikipédia.

Actuellement, les bénévoles qui contribuent à Wikipédia utilisent différents moyens pour indiquer qu'une référence est en accès libre ([https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Acc%C3%A8s\\_url](https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Acc%C3%A8s_url) entre autres). Plusieurs modèles existent et diffèrent selon la version linguistique de Wikipédia et les habitudes de chacun.

Deux versions linguistiques sont à l'étude dans le cadre de la prestation : la version francophone et la version anglophone de Wikipédia. Afin de répondre à ces enjeux et problématiques, la prestation se réalisera en deux

phases : une première phase correspondant à une étude de faisabilité sur les deux versions linguistiques et après choix de l'une des deux langues ou des deux, une deuxième phase de collecte et d'extraction des données pertinentes à avec le développement d'un module spécifique.

Pour réaliser à bien la prestation, les dumps de Wikipédia - en accès libre - seront utilisés en priorité.

## 2. Cahier des charges

### 2.1 Phase 1 : Étude de faisabilité

- Périmètre
  - L'étude de faisabilité doit être réalisée sur les versions linguistiques française et anglaise de Wikipédia
- Objectifs

A partir des dumps de Wikipédia, l'étude de faisabilité devra permettre d'établir les éléments suivants :

- Liste des modèles (<https://fr.wikipedia.org/wiki/Aide:Mod%C3%A8le>) utilisés pour représenter les publications scientifiques
- Liste des modèles utilisés pour représenter les publications scientifiques en accès ouvert
- Identification de références ou sources bibliographiques ([https://fr.wikipedia.org/wiki/Aide:Ins%C3%A9rer\\_une\\_r%C3%A9f%C3%A9rence](https://fr.wikipedia.org/wiki/Aide:Ins%C3%A9rer_une_r%C3%A9f%C3%A9rence)) pointant vers des publications scientifiques qu'il faudra pouvoir distinguer de publications non académiques (presse, ouvrages de vulgarisation, documents de communication institutionnelle, etc.). Le prestataire adoptera une définition des publications scientifiques en concertation avec le comité de pilotage.
- Identification d'un sous-ensemble de références ou sources bibliographiques issues de la recherche française. La liste des références issues de la recherche française depuis 2013 est à trouver dans le Baromètre de la science ouverte, dont les données sont

publiques<sup>1</sup>. La liste des références avant 2013 est à trouver dans OpenAlex<sup>2</sup>.

- Etude de la possibilité de récolter la liste des données suivantes, à qualifier selon un degré de faisabilité à définir par la suite :
  - url de la page Wikipédia
  - sujet principal de la page (article scientifique, biographie ...)
  - date de création de la page Wikipédia
  - sujet principal de l'article
  - nombre de consultations de la page (nombre de visiteurs annuel)
  - date et auteur de l'ajout d'une mention à une publication scientifique française en open access
  - type de référence (source bibliographique, média, revue...)
  - Place de la référence dans l'article (résumé, notes, références, bibliographie, tableaux, illustrations...)
  - lien de la source
- Possibilité d'effectuer une collecte annuelle et documentation des scénarios techniques pour le faire
- Choix d'une version linguistique ou de deux versions linguistiques à explorer pour collecter des données citées au préalable
- Choix de deux années de référence pour exploiter et collecter des données citées au préalable
- Rédaction d'un Plan de gestion des données.
- Réalisation de présentations explicatives des avancées de l'étude : notes écrites faisant état de l'avancement, difficultés rencontrées et moyens de contournement, tableau de données pouvant être collectées et qualifiées selon le degré technique de faisabilité et liste des modèles recensés ainsi que tout autre élément pertinent

---

<sup>1</sup>[https://data.enseignementsup-recherche.gouv.fr/explore/dataset/open-access-monitor-france/information/?disjunctive.year&disjunctive.rors&disjunctive.oa\\_host\\_type&disjunctive.oa\\_colors&refine.bso\\_local\\_affiliations=130025661](https://data.enseignementsup-recherche.gouv.fr/explore/dataset/open-access-monitor-france/information/?disjunctive.year&disjunctive.rors&disjunctive.oa_host_type&disjunctive.oa_colors&refine.bso_local_affiliations=130025661) Il faudra filtrer sur le champ

bso\_country = fr

<sup>2</sup>[https://api.openalex.org/works?filter=institutions.country\\_code:FR&per\\_page=200&page=1](https://api.openalex.org/works?filter=institutions.country_code:FR&per_page=200&page=1) (en itérant sur les pages)

pour prendre connaissance du projet par les équipes du MESR et de Wikimedia France

- Rédaction d'un rapport présentant en anglais la méthodologie et les premiers résultats, qui a vocation à être rendu public par le Comité pour la science ouverte.

## 2.2 Phase 2 : Développement et collecte des données

- Périmètre
  - La collecte des données recherchées s'effectuera sur la version française, anglaise ou les deux de l'encyclopédie Wikipédia en fonction des résultats de l'étude de faisabilité et en concertation avec le Ministère de l'enseignement supérieur et de la recherche
  - Les années utilisées pour l'analyse des dumps seront définis à la fin de la phase 1 entre toutes les parties prenantes.
- Objectifs
  - Développement d'un script d'analyse des dumps de Wikipédia
  - Documentation du code et publication sous licences libres (GNU GPL, EUPL, CeCILL, MIT...) permettant la réutilisation des données et la configuration d'outils tierces par l'un ou l'autre des partenaires ainsi que toute autre structure intéressée par la réutilisation des données (laboratoire de recherche par exemple). Le script sera hébergé sur le compte GitHub de Wikimedia France.
  - Extraction et mise en forme des données à destination des commanditaires (tableau qualifié des données)
  - Améliorer le format des références sur les projets Wikimedia
- Point d'attention
  - Il conviendra d'utiliser le langage Python.

## 3. Cadre de réponse

Il est possible de candidater en consortium pour les deux phases du projet.

Les candidatures peuvent être envoyées par courrier électronique ou par courrier postal avant la date limite des candidatures.

### 3.1 – Calendrier et organisation

#### 3.1.1 Phase de candidature

- Date de publication de l'appel d'offre : 19 janvier 2024
- Date limite de candidature : 16 février 2024
- Date de réponse : 1<sup>er</sup> mars 2024

#### 3.1.2 Phases du projet

##### Phase 1 : Étude de faisabilité

- Date de début de mission : 1<sup>er</sup> mars 2024
- Points d'étape :
  - 1 réunion hebdomadaire avec l'équipe de Wikimedia France
  - 2 réunions par mois avec les équipes du Ministère de l'Enseignement supérieur et de la Recherche
- Date de fin de mission : 31 mai 2024

##### Phase 2 : Collecte des données

- Date de début de mission : 1<sup>er</sup> juin 2024

- Points d'étape :
  - 1 réunion hebdomadaire avec l'équipe de Wikimedia France
  - 1 réunion par mois avec les équipes du Ministère de l'Enseignement supérieur et de la Recherche
- Date de fin de mission : 30 septembre 2024

## 3.2 – Dossier de candidature

### 3.2.1 Liste des documents du dossier de candidature

- Note d'intention du prestataire par rapport à la mission proposée
- Présentation de la structure, de l'équipe, portfolio des projets similaires incluant si possible des exemples d'analyse et de visualisation de données
- Présentation des capacités techniques du prestataire
- Calendrier prévisionnel
- Budget du projet détaillé et ventilé

### 3.2.2 Déposer un dossier de candidature

Nous vous invitons à envoyer les pièces mentionnées en 3.2.1. à l'adresse suivante : [ao-bso@wikimedia.fr](mailto:ao-bso@wikimedia.fr) ou par courrier à l'adresse suivante : Wikimedia France, 28 rue de Londres 75009 Paris.

Vous pouvez également contacter la personne responsable du projet pour des questions sur l'appel d'offres :  
Mathieu Denel, chargé de mission éducation,  
[mathieu.denel@wikimedia.fr](mailto:mathieu.denel@wikimedia.fr)  
Téléphone : 07 62 92 42 01

## 3.3 – Critères d'éligibilité et modalité de sélection

Le comité de sélection sera particulièrement attentifs aux critères suivants :

- Note technique (60 %) permettant notamment de jauger de la capacité du prestataire à documenter le code, créer un plan de gestion de données, utiliser le langage privilégié et rendre accessible les avancées de la prestation.
- Budget (25 %) détaillé et ventilé
- Calendrier (15 %) indiquant les échéances principales

### 3.4 – Modalités de règlement

Le versement des prestations s'effectuera en plusieurs fois selon les phases :

- Phase 1 :
  - 20% à la signature du contrat
  - 40% après validation de la fin de la phase 1
- Phase 2 :
  - 20% au démarrage de la phase 2
  - 20% après validation de la fin de la phase 2