

# Matilda

---

 [openscience.pasteur.fr/2024/10/14/matilda/](https://openscience.pasteur.fr/2024/10/14/matilda/)

CeRIS - Institut Pasteur

14 octobre 2024

Matilda est un **moteur de recherche académique multidisciplinaire** basé sur des données ouvertes et sur un logiciel open-source. Son objectif est de faciliter l'accès et l'utilisation des **données de références et de citations**. Il ne s'agit pas d'une base de données bibliométriques permettant de réaliser des analyses poussées (les données de citations restent à améliorer, les affiliations ne sont pas présentées...) mais d'un outil pour aider les scientifiques à **découvrir des références bibliographiques**.

Le développement de Matilda est coordonné depuis 2019 par Didier Torny (CNRS). Le projet est porté par le Centre de sociologie de l'innovation et Huma-Num et a bénéficié d'un appui financier de l'ANR et du CNRS. Il se place en alternative open-source à :

- Google Scholar, outil dont les données sont en libre accès mais dont les sources et la hiérarchisation des références sont opaques ;
- *Web of Science* et *Scopus*, outils commerciaux dont les données sont accessibles sur abonnement, mais plus adaptés pour des analyses bibliométriques ;
- OpenAlex, autre outil ouvert dont on vous parlait [ici](#), qui peut être utilisé comme outil de découverte mais aussi pour de la bibliométrie ;
- PubMed, autre base de données bibliographiques mais spécialisée dans les domaines des sciences biomédicales et sciences de la vie.

Les caractéristiques de Matilda :

– Matilda compile des données provenant de nombreuses sources ouvertes : ArXiv, PubMed, Crossref, RePEc, HAL, Unpaywall et Orcid.

– Concernant sa couverture :

- Matilda se limite aux **documents textuels académiques** (cela exclut les logiciels, vidéos, photos, images, audio...). Tous les textes sont traités de façon égale : aucune exclusion ou hiérarchisation en fonction de la langue, du type de document... Sa couverture est donc potentiellement plus étendue que les autres moteurs de recherche pour les langues autres que l'anglais.
- La **couverture temporelle** de Matilda est pour le moment limitée : les documents référencés dans Matilda sont ceux ayant été publiés au plus tôt en 2019 sur les sources Matilda ou ayant été cités par un document lui-même publié après 2019. Seule la source HAL comprend aujourd'hui l'ensemble de l'historique, mais Matilda contiendra à terme la totalité de chacune des sources.

– La **fréquence des mises à jour est très élevée**. Les sources sont moissonnées quotidiennement et l'intégration des références dans Matilda se fait en moyenne en 2 jours.

– Concernant les fonctionnalités de recherche :

- **Matilda contient tous les PDF légalement disponibles**. Il est possible d'étendre sa recherche à leurs contenus ou non. En cas de recherche dans le PDF, les résultats affichent la contextualisation des mots recherchés.
- Le modèle de données de Matilda décrit des œuvres, leurs versions et les auteurs. Si l'auteur a un ORCID, l'outil permet d'accéder à sa fiche et à ses publications liées. Par contre, Matilda ne référence pas les laboratoires de recherche car ils ne comportent aucun identifiant dans les publications.
- Les sources étant multilingues, la recherche peut se faire en différentes langues mais ne ramènera que les documents de la langue utilisée.

Enfin, Matilda propose d'autres fonctionnalités intéressantes comme l'export massif de références en BibTeX ou la possibilité de se créer des alertes par flux RSS pour être informé de toute mise à jour associée aux résultats d'une requête.

### **Pourquoi le nom de Matilda ?**

Le nom de l'outil fait référence à l'effet Matilda, c'est-à-dire la minimisation, voire la spoliation, récurrente de la contribution des femmes dans la recherche scientifique. L'effet Matilda est souvent comparé à l'effet Matthieu selon lequel la reconnaissance pour une nouvelle découverte est plus susceptible d'être accordée aux chercheurs déjà très reconnus qu'à ceux qui le sont moins.