

# OpenAlex au service des analyses bibliométriques et des communautés

Étude bibliométrique sur le Domaine des Écoles Polytechniques Fédérales (EPF)

Simon Willemin (ETH Zurich)

LabExplorer : Identifier les experts EPFL grâce aux données ouvertes et à l'IA

Gaël Bernard (EPFL)

Journée francophone OpenAlex, 2 décembre 2025

**ETH** zürich

**EPFL**

# Étude bibliométrique sur le Domaine des Écoles Polytechniques Fédérales (EPF)

Journée francophone OpenAlex, 2 décembre 2025

**ETH** zürich

**EPFL**

# Contenu

1. Éléments de contexte
2. Préparation d'une offre reposant sur des données ouvertes (OpenAlex)
3. L'étude bibliométrique sur les institutions du Domaine des EPF de 2026 (en préparation)

# Éléments de contexte

# Le Domaine des Écoles Polytechniques Fédérales (EPF)

- Deux écoles polytechniques fédérales
  - EPFL (École Polytechnique Fédérale de Lausanne)
  - ETH Zurich (École Polytechnique Fédérale de Zurich)
- Quatre établissements de recherche fédéraux
  - Eawag (Institut fédéral des sciences et technologies de l'eau)
  - Empa (Laboratoire fédéral d'essai des matériaux et de recherche)
  - PSI (Institut Paul Scherrer)
  - WSL (Institut fédéral de recherches sur la forêt, la neige et le paysage)

# Analyses bibliométriques en Suisse

## Analyses bibliométriques couvrant plusieurs institutions suisses

- [Bibliometric study of the ETH Domain](#) (2022)
  - Rapport réalisé par CWTS (Center for Science and Technology Studies, Université de Leiden), mandaté par le Conseil des EPF
  - Principales sources de données : données **propriétaires** (“CWTS version of the Web of Science database”) et données fournies par les institutions
- [Les publications scientifiques en Suisse, 2008-2022 : Une analyse bibliométrique de la recherche scientifique en Suisse](#) (2024)
  - Rapport réalisé par le SEFRI (Secrétariat d’État à la formation, à la recherche et à l’innovation)
  - Principale source de données : données **propriétaires** (Clarivate Analytics)

## Autres rapports, articles et ressources

- Machado, A., Hoppmann, L., Knaus, J., Palzenberger, M. (2016). [Analysis of the international journal publishing activities in Switzerland with special emphasis on gold open access publishing](#)
  - Analyse réalisée par le Max Planck Digital Library, mandaté par le Fonds national Suisse de la recherche scientifique (FNS) et swissuniversities
  - Principales sources de données : données **propriétaires** (Elsevier) et non propriétaires (DOAJ)
- Hug, S. E., Brändle, M. P. (2017). [The coverage of Microsoft Academic: analyzing the publication output of a university](#), *Scientometrics*, 113, 1551-1571.
  - Comparaison de la couverture des publications d’une université suisse (Université de Zurich) dans des bases de données bibliométriques
  - Bases de données **ouvertes** (Microsoft Academic, désactivé en 2021) et **propriétaires** (Scopus, Web of Science)
- Swiss Open Access Monitor
  - [Journal Monitor](#) (repose depuis 2024 sur des données **ouvertes**, principalement OpenAlex)
  - [Repository Monitor](#) (repose sur des données fournies par les institutions et disponibles en **libre accès**)

# Arguments pour une transition vers des outils ouverts (2023-2024)



## Le CNRS se désabonne de la base de publications Scopus

11 janvier 2024

INSTITUTIONNEL



“Le désabonnement de la base bibliographique Scopus est une **première étape** pour s’affranchir des bases commerciales” ; “Les statistiques d’usages montrent que les chercheurs et chercheuses du CNRS ont eu recours beaucoup plus fréquemment aux services du Web of Science qu’à ceux de Scopus. Pour accompagner cette **transition** vers de nouveaux outils, l’abonnement au *Web of Science* sera quant à lui **maintenu.**” (CNRS, 11.01.2024)



## Le CNRS s’émancipe du Web of Science

01 décembre 2025

INSTITUTIONNEL



## Sorbonne Université se désabonne du Web of Science

Menant une politique ambitieuse et volontariste depuis de nombreuses années, Sorbonne Université est pionnière en matière de science ouverte. Poursuivant son engagement en faveur de l’ouverture de données relatives à la recherche, elle fait le choix de ne pas renouveler son abonnement à la base de publications Web of Science, ainsi qu’aux outils bibliométriques de Clarivate en 2024. En s’affranchissant résolument des produits bibliométriques propriétaires, elle ouvre la voie à l’exploration de nouveaux outils ouverts, libres et participatifs.

## Introducing the Leiden Ranking Open Edition

January 30, 2024 • CWTS development • 8 min read

This post introduces the Open Edition of the CWTS Leiden Ranking, published today by CWTS in collaboration with the Curtin Open Knowledge Initiative, Sesame Open Science, and OurResearch. The Leiden Ranking Open Edition is the first fully transparent university ranking.

“While our organisations are making further improvements to the Leiden Ranking Open Edition, CWTS will **continue to release** annual updates of the **traditional** Leiden Ranking based on Web of Science data. In the next one or two years, we expect the traditional Leiden Ranking and the Open Edition to **co-exist.**” (CWTS, 10.01.2024)



## The CWTS Leiden Ranking 2025 - More open, more inclusive, more informative

October 29, 2025 • CWTS Development and Focal Area Information & Openness • 6 min read

“We in fact released two editions of the ranking, the Traditional Edition and the Open Edition.” (CWTS, 29.10.2025)

- **Respect des engagements en faveur de l’ouverture**

“Engagement en faveur de l’ouverture des données relatives à la recherche” (Sorbonne)

“Le CNRS s’est [...] engagé en faveur de la science ouverte” (CNRS)

- **Mise en valeur d’outils ouverts, libres et participatifs**

Sorbonne Université “ouvre la voie à l’exploration de nouveaux outils ouverts, libres et participatifs” (Sorbonne)

- **Transparence**

CWTS “travaille actuellement à une version entièrement transparente et reproductible de son classement, en s’appuyant sur les données ouvertes de CrossRef et d’OpenAlex” (Sorbonne)

“One of the most challenging problems of bibliometric indicators: the lack of transparency of these indicators due to their dependence on proprietary data” (CWTS)

- **Reproductibilité**

CWTS “développe une version transparente et reproductible de son classement mondial des universités” (CNRS)

- **Ownership**

“Permettre aux personnels de recherche de se réappropriier les résultats de leurs propres travaux” (Sorbonne)

“Se réappropri[er] les résultats de [ses] propres travaux” (CNRS)

- **Accessibilité à la société**

“Ouvrir la science à la société” (Sorbonne)

“Rend[re] la science accessible à l’ensemble de la société” (CNRS)

- **Durabilité**

“Consolidation d’une alternative durable et internationale, en s’appuyant notamment sur l’outil OpenAlex” (Sorbonne)

“Les économies [...] réalisées permettront de soutenir et de consolider des solutions ouvertes et durables” (CNRS)

# Préparation d'une offre reposant sur des données ouvertes (OpenAlex)

# Rapport d'auto-évaluation du Domaine des EPF

- Le **rapport d'auto-évaluation du Domaine des EPF** est réalisé tous les quatre ans :
  - [Self-Assessment Report: Intermediate Evaluation 2023 of the ETH Domain](#) (2022)
  - [Self-Assessment Report: Intermediate Evaluation 2019 of the ETH Domain](#) (2018)
  - [Self-Assessment Report: Intermediate Evaluation 2015 of the ETH Domain](#) (2014)
- Il repose notamment sur une **étude bibliométrique sur les institutions du Domaine des EPF**. En 2022, l'étude a été réalisée pour la cinquième fois consécutive par le Center for Science and Technology Studies (CWTS) de l'Université de Leiden aux Pays-Bas et reposait sur des données propriétaires :
  - [Bibliometric study of the ETH Domain](#) (2022)

# Spécifications (appel d'offre)

- L'étude bibliométrique de 2026 doit contenir des indicateurs sur :
  1. La productivité et l'impact
  2. Les collaborations et les réseaux
  3. L'accès ouvert
  4. Le genre
- L'étude doit prendre en compte les types de publications suivants :
  1. Les articles de revues évalués par les pairs (*peer reviewed journal articles*)
  2. Les actes de conférence (*conference proceedings*)
- L'étude doit couvrir une période de 10 ans (2014-2023)
- L'étude doit être réalisée dans un délai de 13 mois

# Nos ambitions (au-delà des spécifications)



Réalisation d'une étude en interne



Utilisation de données ouvertes (OpenAlex)

# Arguments mis en avant



Réalisation d'une étude en interne



Utilisation de données ouvertes (OpenAlex)

Arguments	Explications
Connaissance interne	Une analyse interne consolide et sécurise l'expertise interne au sein des institutions concernées
Durabilité	L'établissement d'un workflow interne pour l'étude bibliométrique permet d'offrir de telles analyses régulièrement (non seulement tous les quatre ans, mais également annuellement, par exemple)
Coûts	Moins de cash-out (redistribution des ressources existantes)
Walk the talk (science ouverte)	Les institutions du Domaine des EPF soutiennent officiellement la science ouverte et les données de la recherche ouvertes. Avec une telle analyse, le Domaine des EPF peut passer de la parole aux actes et réaliser ses objectifs proclamés en publiant ouvertement les résultats de l'analyse et la méthodologie. À cet égard, la crédibilité de l'étude bibliométrique du Domaine des EPF augmente

# Risques et réponses ou stratégies de mitigation



Réalisation d'une étude en interne



Utilisation de données ouvertes (OpenAlex)

Risques	Réponses ou stratégies de mitigation
Neutralité et crédibilité	Une expertise interne n'est pas nécessairement reconnue au même degré qu'une expertise externe. Ce risque est mitigé en réalisant l'analyse avec des données ouvertes et en publiant le code de manière à garantir que tous les indicateurs et toutes les figures sont reproductibles. De plus, l'analyse bibliométrique fait partie d'une auto-évaluation et peut être acceptée comme une analyse réalisée de manière interne
Faisabilité	Les institutions disposent des ressources nécessaires à la réalisation de l'étude en interne avec OpenAlex, une base de données bibliométrique ouverte lancée en 2022
Continuité avec les éditions précédentes de l'étude bibliométrique	Seul le prestataire antérieur est en mesure de fournir une analyse en tout point comparable à celle qui a été fournie lors des précédentes études. Tout autre prestataire, interne comme externe, ne peut le faire
Qualité des données	Des données fournies et validées par les institutions peuvent être utilisées comme base de l'analyse

# L'étude bibliométrique sur les institutions du Domaine des EPF de 2026 (en préparation)

# L'équipe qui a soumis le projet

- Omar Ballester (EPFL)
- Gaël Bernard (EPFL)
- Mahmoud Hemila (ETH Zurich)
- Simon Willemin (ETH Zurich)

# Work Packages

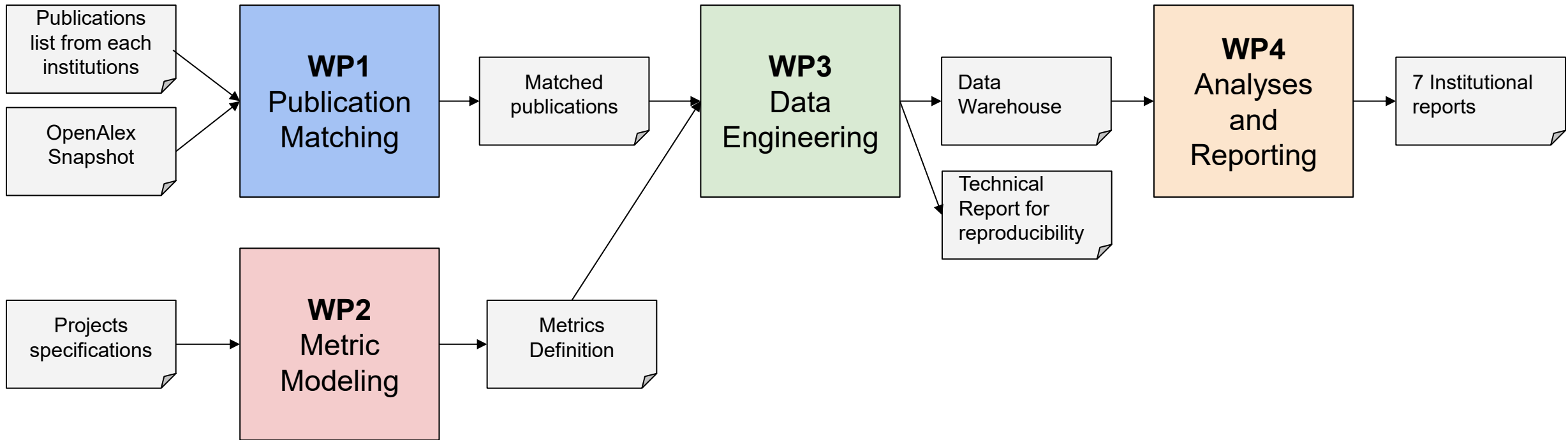
**WP1**  
Publication  
Matching

**WP3**  
Data  
Engineering

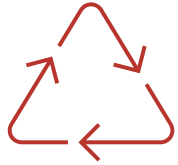
**WP4**  
Analyses  
and  
Reporting

**WP2**  
Metric  
Modeling

# Work Packages



# Étude bibliométrique : Quelques exemples de défis ou points de discussion



## Réutilisation du travail réalisé lors des **précédentes éditions**

- Accès à l'étude de 2022
- Pas d'accès au code utilisé pour implémenter les indicateurs de l'étude de 2022



## **Types de publications** inclus dans les spécifications

- Prise en compte des articles de revues évalués par les pairs (*peer reviewed journal article*)
- Prise en compte des actes de conférences (*conference proceedings*)



## Actualisation des données externes

- Opposition entre **stabilité des données** (utiliser un snapshot disponible dès les premières phases du projet) et **promesse d'amélioration significative de la qualité** (publication du premier "Walden" snapshot en novembre 2025)

# Au-delà de l'étude bibliométrique



Du **code réutilisable** permettant de chercher le(s) identifiant(s) OpenAlex d'une publication à partir du DOI, du titre, des auteurs et du nom du journal dans un snapshot OpenAlex et de calculer les métriques incluses dans l'étude



Un **rapport technique** séparé décrivant précisément les métriques et le code associé sera publié en libre accès



L'**entrepôt de données** (DWH) à partir duquel les métriques sont implémentées n'est pas limité au Domaine des EPF (p. ex. métriques normalisées). Possibilité de le réutiliser pour d'autres institutions avec un effort limité



La comparaison entre les données fournies par les institutions et les données OpenAlex a permis d'identifier des erreurs qui ont été ou qui pourront être corrigées de manière à **améliorer la qualité** des données ouvertes

Rendez-vous en 2026 pour une  
étude bibliométrique du Domaine des EPF  
reposant sur OpenAlex

Merci de votre attention

Avez-vous des questions ?

Contact : [eth-bibliometric-study@library.ethz.ch](mailto:eth-bibliometric-study@library.ethz.ch)