

10th National Meet & Greet · Swiss Medical Librarians

Bridging Institutional and Open Repositories

How EPFL Uses OpenAlex and AI to Serve
Its Academic Community

University of Bern
26 August 2026

Gaël Bernard
Data Scientist
Institutional data and analytics
gael.bernard@epfl.ch



EPFL at a glance

**14'072***

Students

**372**

Professors

**6'530**Employees
(incl. Phd)**+130**

Nationalities

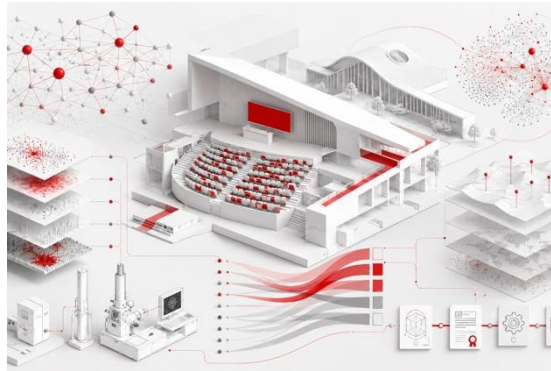
**Including BSc, MSc, PhD and Continued Education*

Institutional Data and Analytics



Strategic Impact Analysis

- Advisory to the Presidency
- Institutional data stories
- Benchmarking and sustainability



Academic Analytics

- Teaching and student success
- Research and funding insights
- Technology transfer indicators



Data Gov & Coordination

- Data quality processes
- Stewardship and standards
- National data coordination

Two kinds of repositories

Institutional repository

local



One institution's record

Curated by people

Decides who and what counts

Infoscience (EPFL) · SERVAL (UNIL)
BORIS (Bern) · Research Collection (ETHZ) · ...

Open repository

global



One index for the whole world

Aggregated automatically

Discovery and benchmarking

OpenAlex · OpenAIRE

One provides **trusted, curated data**.

The other enables **global discovery and benchmarking**.

One bridge, two crossings

Crossing 1

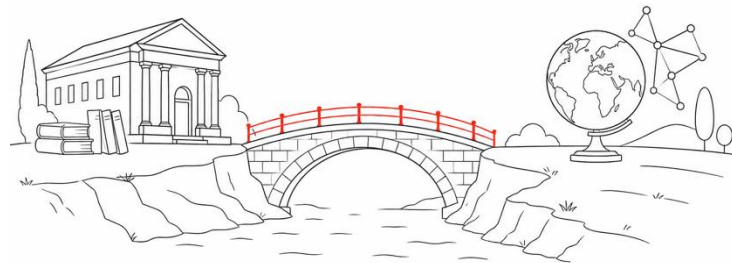
LabExplorer

using the local repository
(augmented with the global one)

Crossing 2

Bibliometric Reports

using the global repository
(checked against the local lists)



A project manager at the EPFL AI Center writes:

Part of my job is to point the right funding opportunities to the right faculty members.

I spent the last couple of weeks exploring LLMs to browse faculty expertise. The system shows potential, but further improvements would require **better data...**

LabExplorer

LabExplorer

Looking for experts on topics like Renewable fuels of non-biological origin or AI for personalized healthcare ?
Enter your query below to see a ranked list of EPFL labs working in the field.

Visual prosthetics

▶ Run LabExplorer

Last update: 5 August 2026 — LabExplorer now uses a local model provided by [RCP \(Qwen/Qwen3.6-35B-A3B\)](#) instead of OpenAI. Questions? Feel free to reach out!

Nom unité	Nom	Prénom	fraction_pos_paper
Chaire Fondation Bertarelli de technologie neuroprosthétique	Lacour	Stéphanie	28.71%
Laboratoire de neurotechnologies intégrées	Shoaran	Mahsa	26.56%
Chaire Fondation Bertarelli en neuro-ingénierie translationnelle	Micera	Silvestro	16.61%
Laboratoire d'intelligence et d'apprentissage visuels	Roshan Zar	Amir	4.76%
Unité du Prof. Schrimpf - SV	Schrimpf	Martin	4.35%
Unité du Prof. Hummel	Hummel	Friedhelm C	3.65%
Groupe SCI STI SC	Carrara	Sandro	3.52%
Laboratoire de psychophysique	Herzog	Michael	3.23%
Groupe SCI STI AS	Skrivervik	Anja	2.36%

Impact

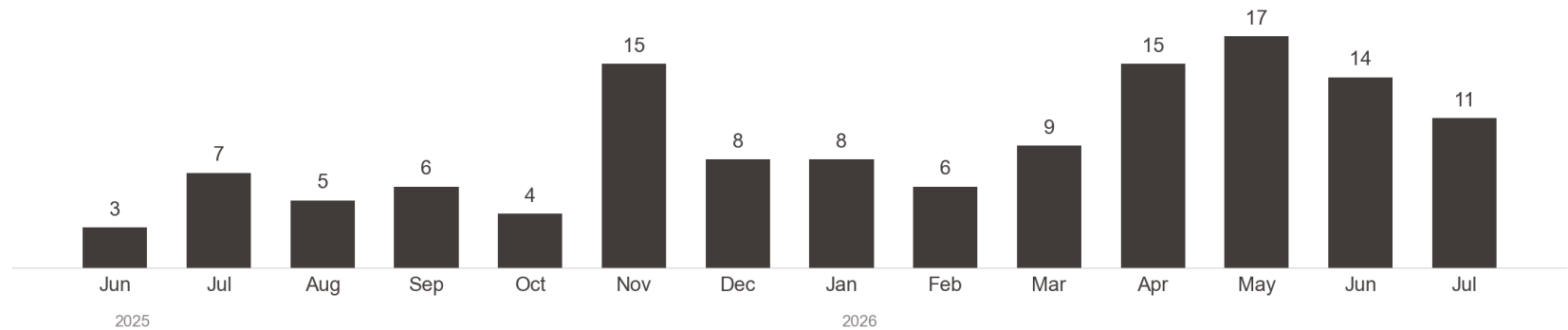
985

queries on LabExplorer
in 13 months

596

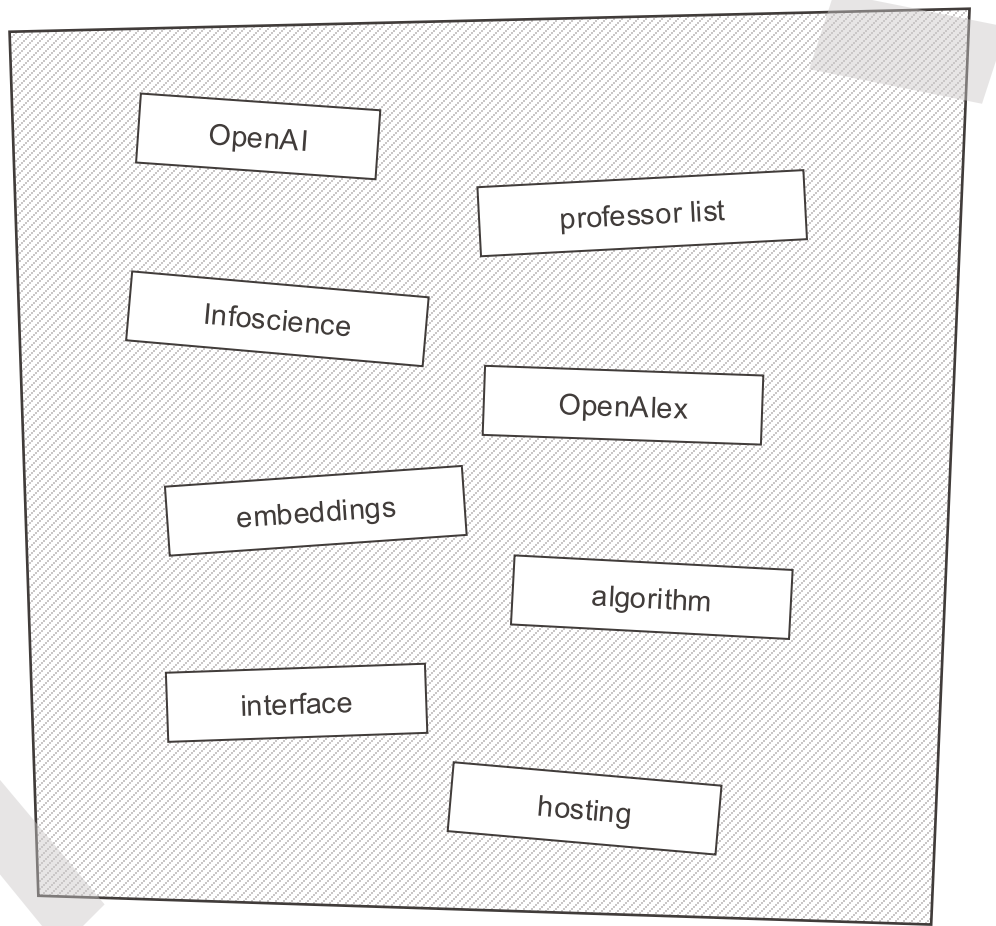
recommendation emails
two campaigns with the Research Office

Unique users per month



The monolith

LabExplorer 2025



- MVP in a few weeks (AI)
- One single block
- Coupled, fragile: and the right choice to start

The by-products

ListeProf

the link

shared with the library

ALIRA

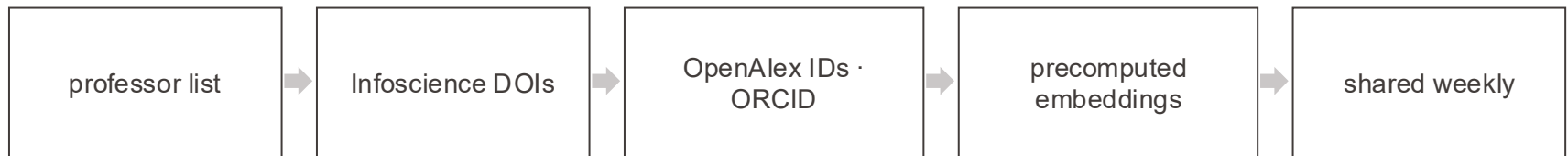
the AI core

pip install alira

LabExplorer

the interface

what remains



Every professor linked to the global identifiers.
Shared with the library: the links flow back into Infoscience.



The matching engine, extracted and released open source.

Domain agnostic: any text corpus, not only papers.

Screened all 69,625 companies of the Vaud commercial register:

2,605 discovered related to sport.

`pip install alira · github.com/epfl-p-data/alira`

The product is now developed by someone else.

A power-user team liked LabExplorer enough to build **LabExplorer²** on top of it:
their own interface, driven through a small API.

It now generates a large share of the traffic.

One bridge, two crossings

Crossing 1

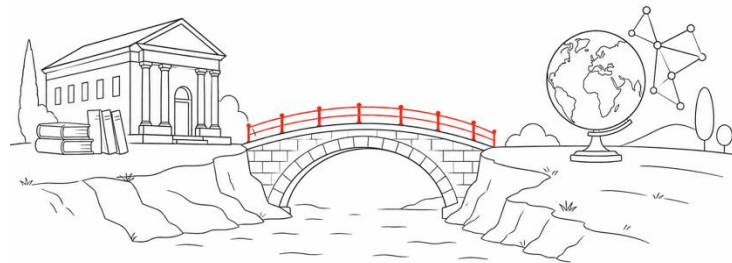
LabExplorer

using the local repository
(augmented with the global one)

Crossing 2

Bibliometric Reports

using the global repository
(checked against the local lists)



Bibliometric reports

Bibliometric study of EPFL (2014–2023)

An open bibliometric analysis of the institutions of the ETH Domain

July 26, 2026

Bibliometric study of the ETH Domain (2014–2023) An open bibliometric analysis of the institutions of the ETH Domain July 26, 2026 Report for the ETH Board Contributors Project Team (Lead Authors): Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL), Former Project Team Members: Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich), Sounding Board: Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL), E-mail: eth-bibliometric-study@library.ethz.ch	Bibliometric study of ETH Zurich (2014–2023) An open bibliometric analysis of the institutions of the ETH Domain July 26, 2026 Report for the ETH Board Contributors Project Team (Lead Authors): Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL), Former Project Team Members: Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich), Sounding Board: Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL), E-mail: eth-bibliometric-study@library.ethz.ch	Bibliometric study of the Paul Scherrer Institute (PSI) (2014–2023) An open bibliometric analysis of the institutions of the ETH Domain July 26, 2026 Report for the ETH Board Contributors Project Team (Lead Authors): Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL), Former Project Team Members: Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich), Sounding Board: Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL), E-mail: eth-bibliometric-study@library.ethz.ch
--	--	---

Bibliometric study of WSL (2014–2023) An open bibliometric analysis of the institutions of the ETH Domain July 26, 2026 Report for the ETH Board Contributors Project Team (Lead Authors): Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL), Former Project Team Members: Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich), Sounding Board: Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL), E-mail: eth-bibliometric-study@library.ethz.ch	Bibliometric study of Empa (2014–2023) An open bibliometric analysis of the institutions of the ETH Domain July 26, 2026 Report for the ETH Board Contributors Project Team (Lead Authors): Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL), Former Project Team Members: Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich), Sounding Board: Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL), E-mail: eth-bibliometric-study@library.ethz.ch	Bibliometric study of Eawag (2014–2023) An open bibliometric analysis of the institutions of the ETH Domain July 26, 2026 Report for the ETH Board Contributors Project Team (Lead Authors): Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL), Former Project Team Members: Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich), Sounding Board: Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL), E-mail: eth-bibliometric-study@library.ethz.ch
---	--	---

Report for the ETH Board

Contributors

Project Team (Lead Authors)
Former Project Team Members
Sounding Board
E-mail

Alexander Ehlert (ETH Zurich) and Gaël Bernard (EPFL)
Mahmoud Hemila (ETH Zurich), Omar Ballester (EPFL) and Simon Willemin (ETH Zurich)
Alexander Nebel (EPFL), David Johann (ETH Zurich), Michael Bachmann (Lib4RI) and Sarah Gerster (EPFL)
eth-bibliometric-study@library.ethz.ch

Key metrics

≈ 270 M

records in
OpenAlex

≈ 57 M

in scope
(criteria)

135,992

ETH Domain
2014-2023

7

reports
avg 71 pages

From the whole world to seven reports.

Scope numbers only. The reports are still under validation by the ETH Board and will probably become public soon.

The publication lists

**Each institution sent its own list
of publications.**

More control over the data entering the analysis.

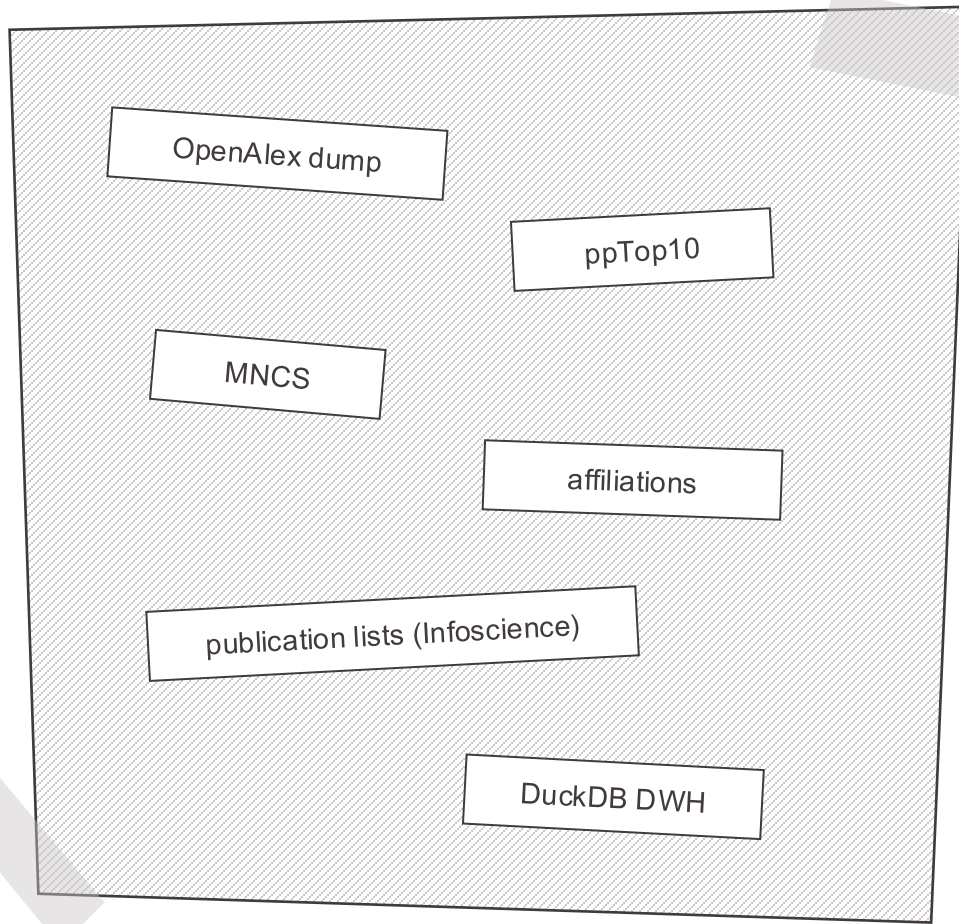
≈ 94%

of the publications sent (2014-2023)
are found in OpenAlex

One global corpus for every institution.
Every figure is checked against these lists.

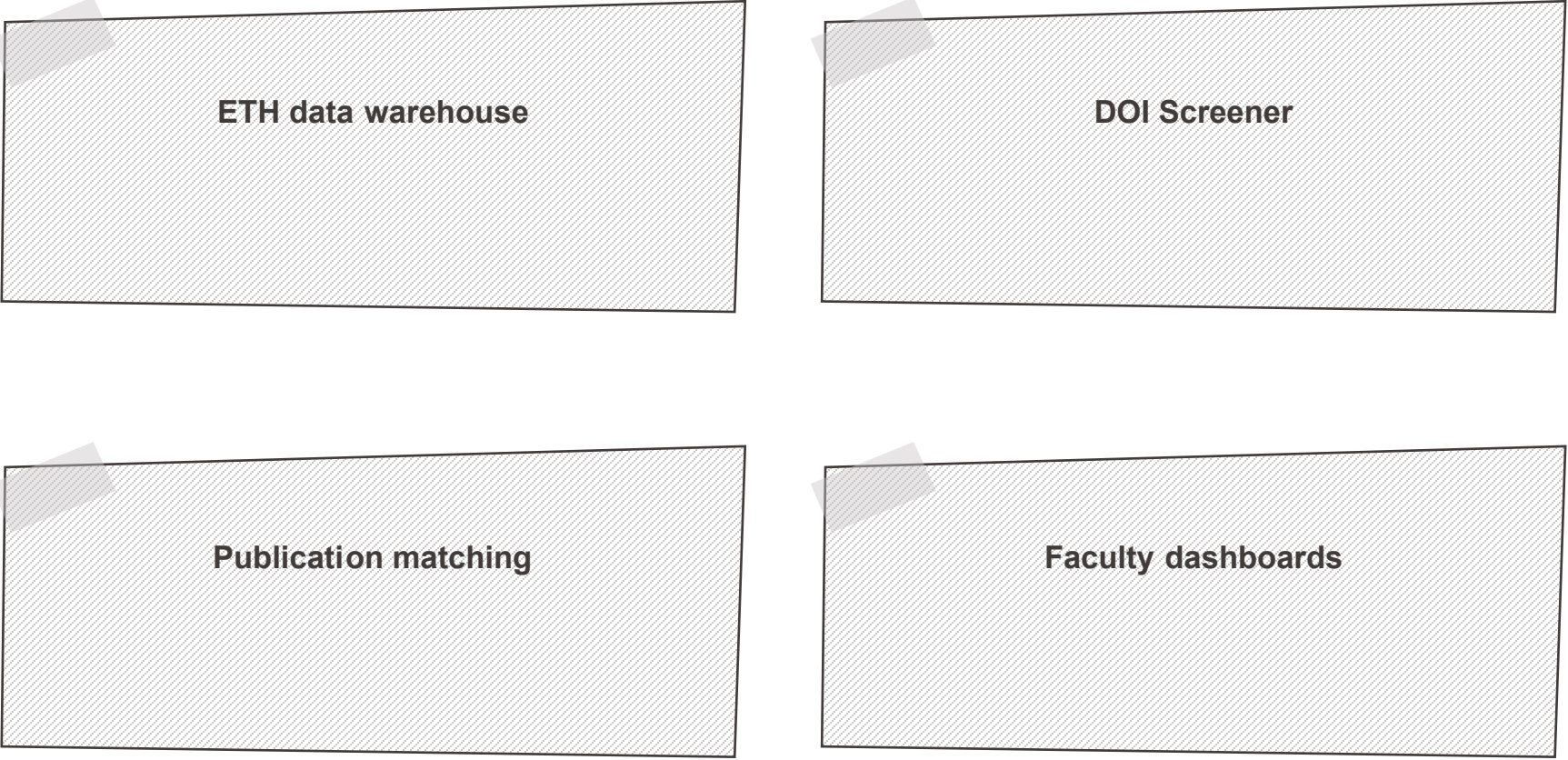
The monolith

ETH Domain 2026



- The first time we run this exercise ourselves
- Until now: an external provider (CWTS Leiden)
- Still one block

The by-products



ETH data warehouse

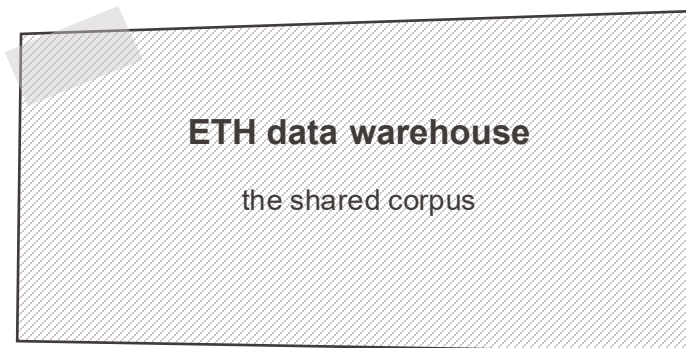
DOI Screener

Publication matching

Faculty dashboards

Side products we can reuse

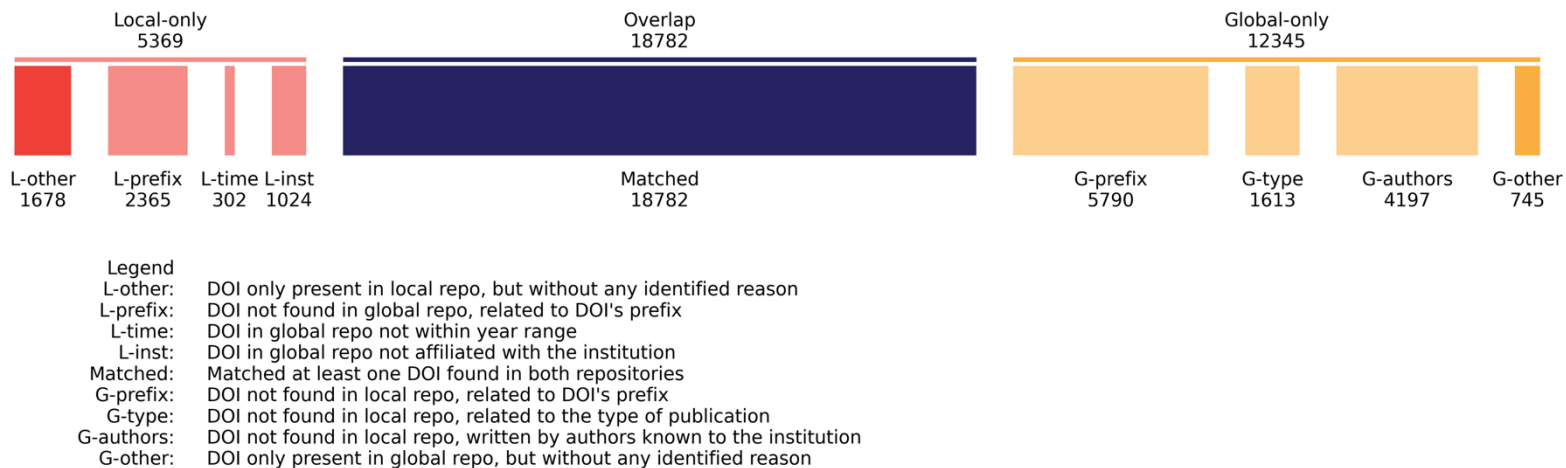
ETH data warehouse



The full OpenAlex dump, about 600 GB,
rebuilt as one queryable DuckDB warehouse.

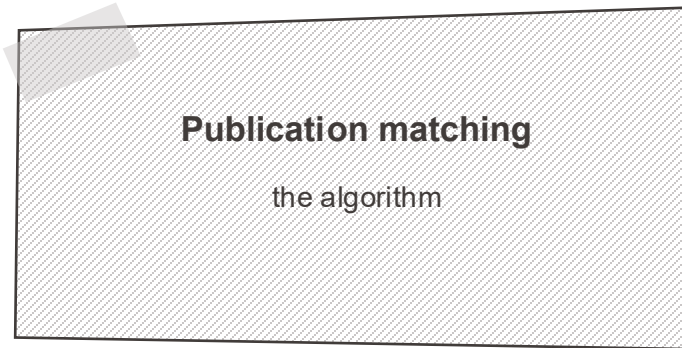
One warehouse for the six institutions.
Every metric in the report reads from it.

DOI Screener



Compares a local repository with OpenAlex and explains,
for every publication missing on either side, why it is missing.

Publication matching

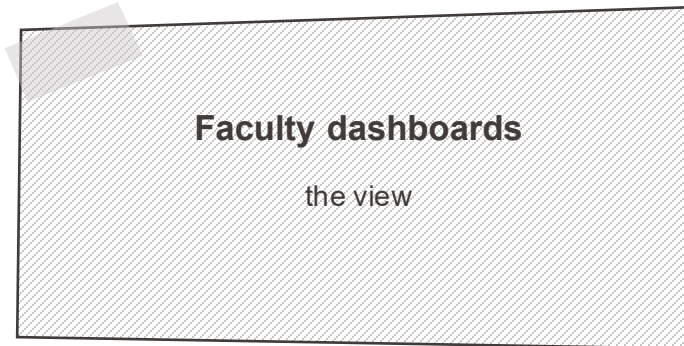


Take the institutional lists ($\approx 100,000$ publications for EPFL) and find each one in the OpenAlex dump: 270 million records.

The hard part: **fuzzy text matching that scales.**

not published yet · probably valuable for libraries far beyond the ETH Domain

Faculty dashboards



The same warehouse, cut per faculty.

Not started yet: EPFL believes it could be a valuable side outcome of the bibliometric study.

Both ends of the bridge

LabExplorer

started from the institutional repository

Infoscience: curated, trusted quality
But new professors are barely in it yet

The open repository (OpenAlex) fills the gap

ETH Domain report

started from the open repository

OpenAlex: one corpus for six institutions
But figures are trusted only against the institutional lists

The institutional repository is the ground truth

Not one side against the other:

in these two projects, the institutional and the open repository were complementary.

Hackathon at EPFL

Swiss Open Academic Data

A Swiss community around open academic data:
OpenAlex, OpenAIRE, ORCID.

Hackathon

2 September 2026 · EPFL

registration open · soad.ch

Hands-on, all levels. Come build with us.

Conclusion

AI makes building fast.

What lasts is the components
we can **reuse, share and open-source.**

gael.bernard@epfl.ch