

# Institutional Scientific Output and Citation Coverage in OpenAlex

The case of the Czech Technical University

**Jan Dvořák<sup>1,2,3</sup> – Tomáš Chudlarský<sup>1,2</sup> – Josef Špaček<sup>1</sup>**

<sup>1</sup> Czech Technical University in Prague

<sup>2</sup> Charles University

<sup>3</sup> euroCRIS



# Czech Technical University

- Largest technical university in CZ
- Established 1707
- 18,000 students
- 6,200 (4,200 FTE) staff
  - 3,900 (2,500 FTE) academics and researchers
- 8 faculties
- 6 university institutes
- Computing and Information Centre
- Central Library



# Institutional CRIS — V3S & EZOP

## **V3S**

- 278K outputs
- 32K recognitions
- 910K citations

## **EZOP**

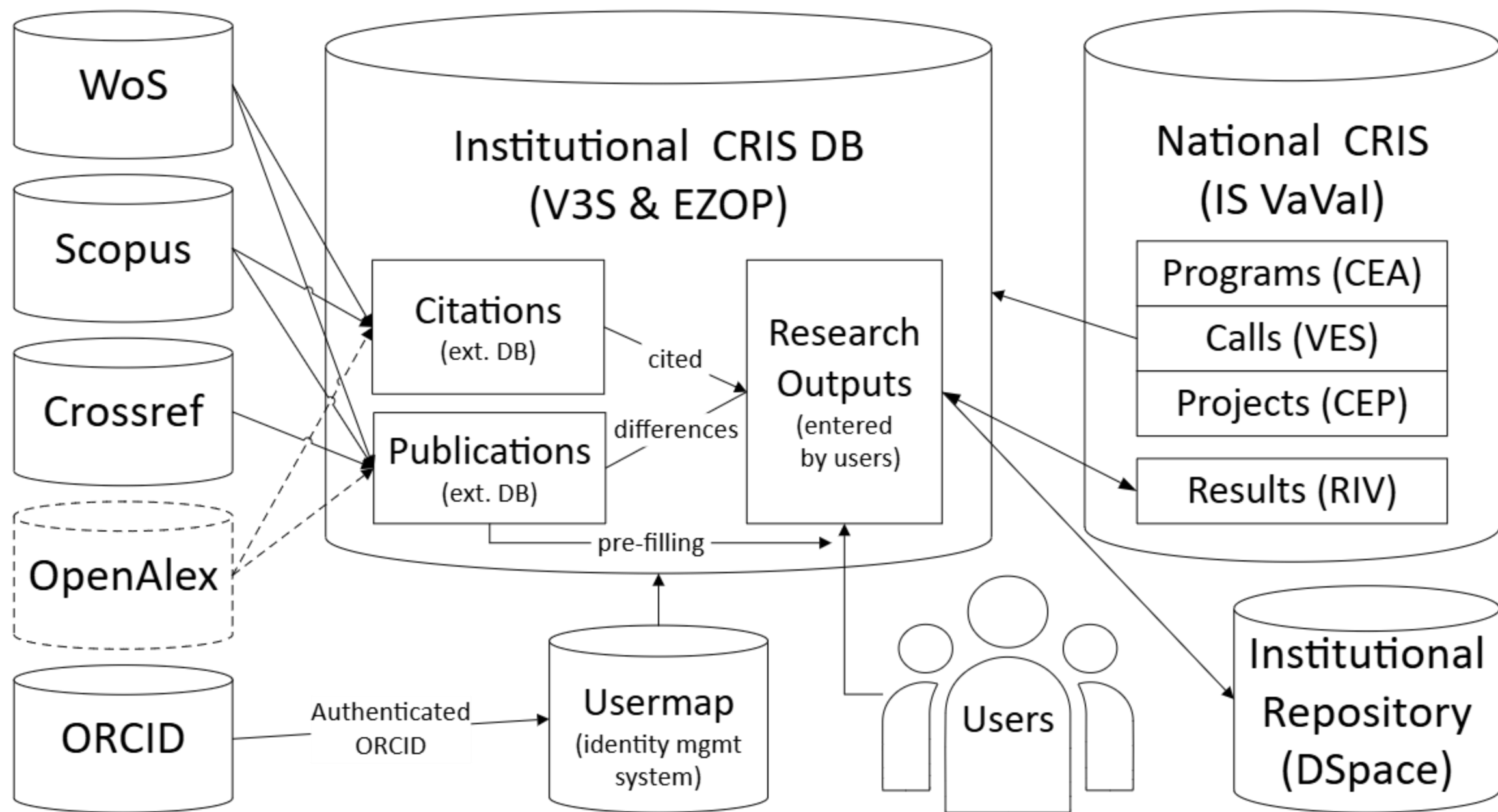
- 24K projects
- 8K contracts for contract research

## In-house built

- 1995 – Oracle Forms (console) / Oracle db
- 2003 – PHP / Oracle db
- 2016 – 2x Java EE app / 1x Oracle Exadata db

# Expectations from a Citation Database

- Comprehensive coverage of our researchers' publication output
  - Affiliations
- Reliable information on citations





- Est. 2022, successor of Microsoft Academic Graph (discont. 2021)
- Non-profit, by OurResearch (UnPaywall, ...)
- Multiple entity types:
  - Scholarly works (~480M, metadata, incl. preprints and datasets)
    - With citation links
  - Researchers
  - Institutions
  - Funding (grants, awards, ...)
  - Funders
  - Sources (journals, repositories, ...)
  - Publishers
  - ...
- Crossref is sourced

# Czech Technical University in Prague

 Institution 

**Homepage:** [cvut.cz](https://cvut.cz) 

**Alternate name:** České vysoké učení technické v Praze

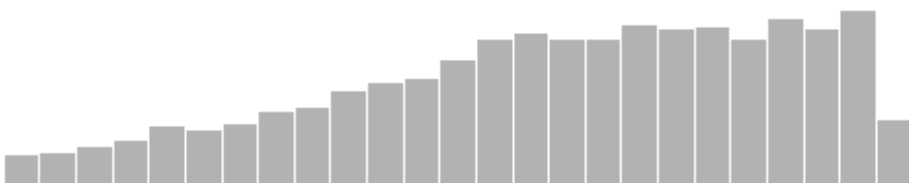
**Child institutions:** [Research Infrastructure for Experiments at CERN](#), [VR-1 – Training Reactor for Research Activities](#), [Czech National Infrastructure for Biological Data](#), [Research Infrastructure for Fermilab Experiments](#), [Brookhaven National Laboratory - Participation of the Czech Republic](#) **+2 more**

**ROR:** 03kqpb082

**Works count:** 49 650

**Citations count:** 4 072 000

 Year 



More...

 Topic 

|                                                       |       |
|-------------------------------------------------------|-------|
| Particle physics theoretical and experimental studies | 1 837 |
| High-Energy Particle Collisions Research              | 1 034 |
| Solid State Laser Technologies                        | 636   |
| Laser-Plasma Interactions and Diagnostics             | 529   |
| Concrete and Cement Materials Research                | 505   |



## MizAR 60 for Mizar 50

PDF

2023 · Jakubův, Jan, Chvalovský, Karel, et al. · *DROPS (Schloss Dagstuhl – Leibniz Center for Informatics)* ·  75 682

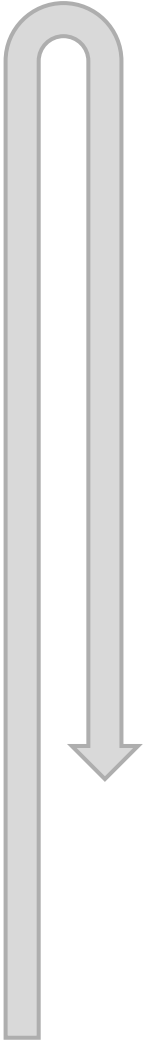
## Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC

2012 · G. Aad, T. Abajyan, et al. · *Physics Letters B* ·  10 670

## The ATLAS Experiment at the CERN Large Hadron Collider

2008 · G Aad, E Abat, et al. · *Journal of Instrumentation* ·  3 978

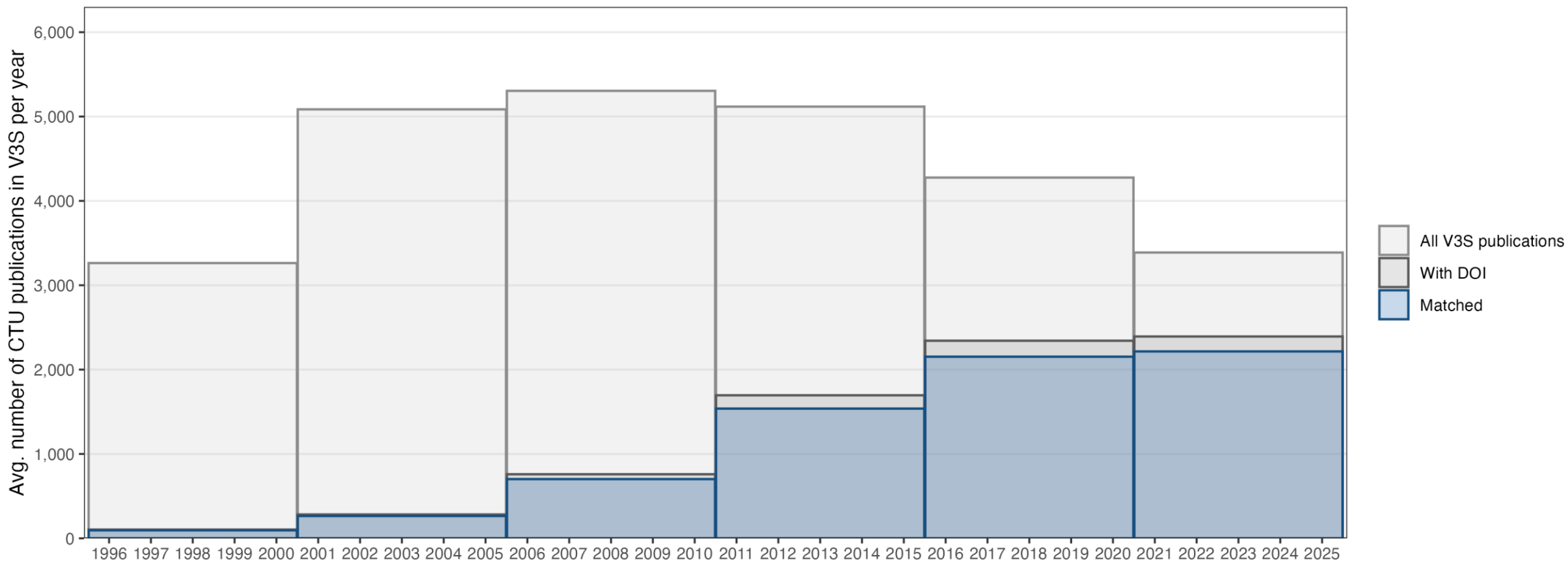
# Integrating OpenAlex in V3S

- 
- Lookup the institution record (through ROR ID) in OpenAlex
  - Fetch the work ids with an affiliation to the institution from OpenAlex
  - Fetch the work records by the ids from ^
  - Attempt linking to research outputs in the CRIS
  - Fetch the work records by DOIs from the CRIS that were not linked
  - Attempt linking to research outputs in the CRIS
  - Fetch the ids of works citing the linked research outputs
  - Fetch the unique work records from ^
  - Insert citing records in the CRIS

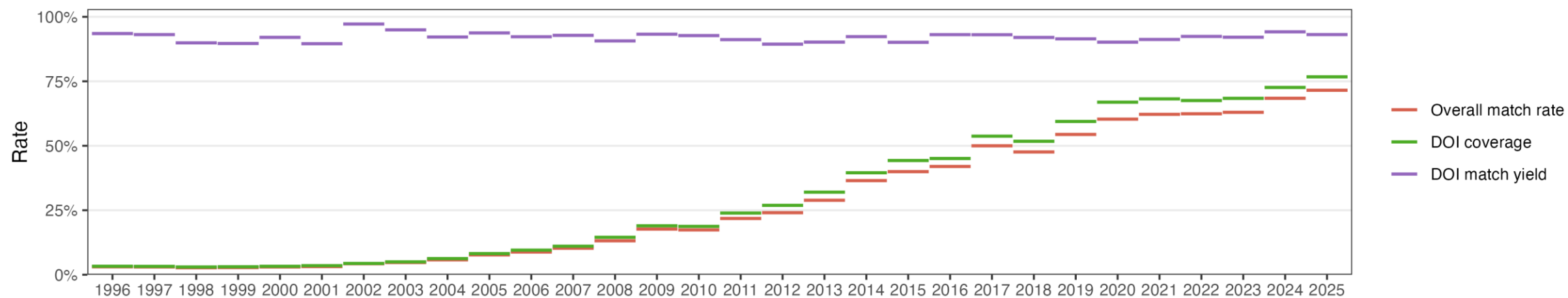
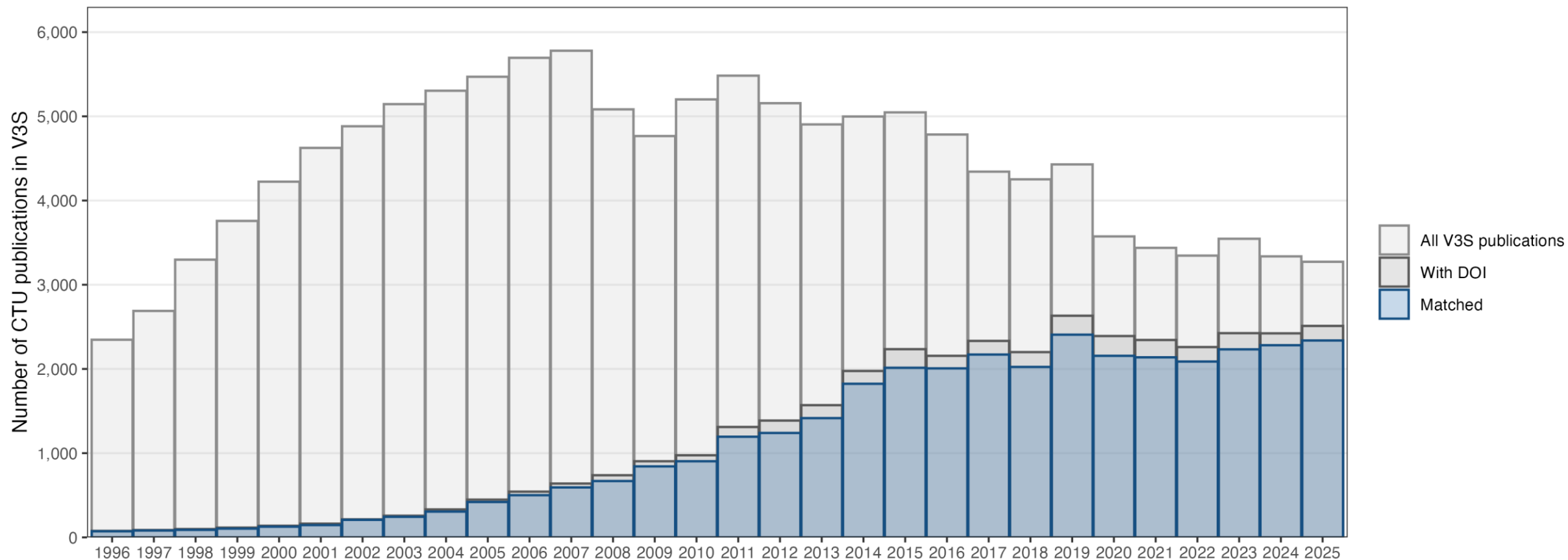


Publication coverage

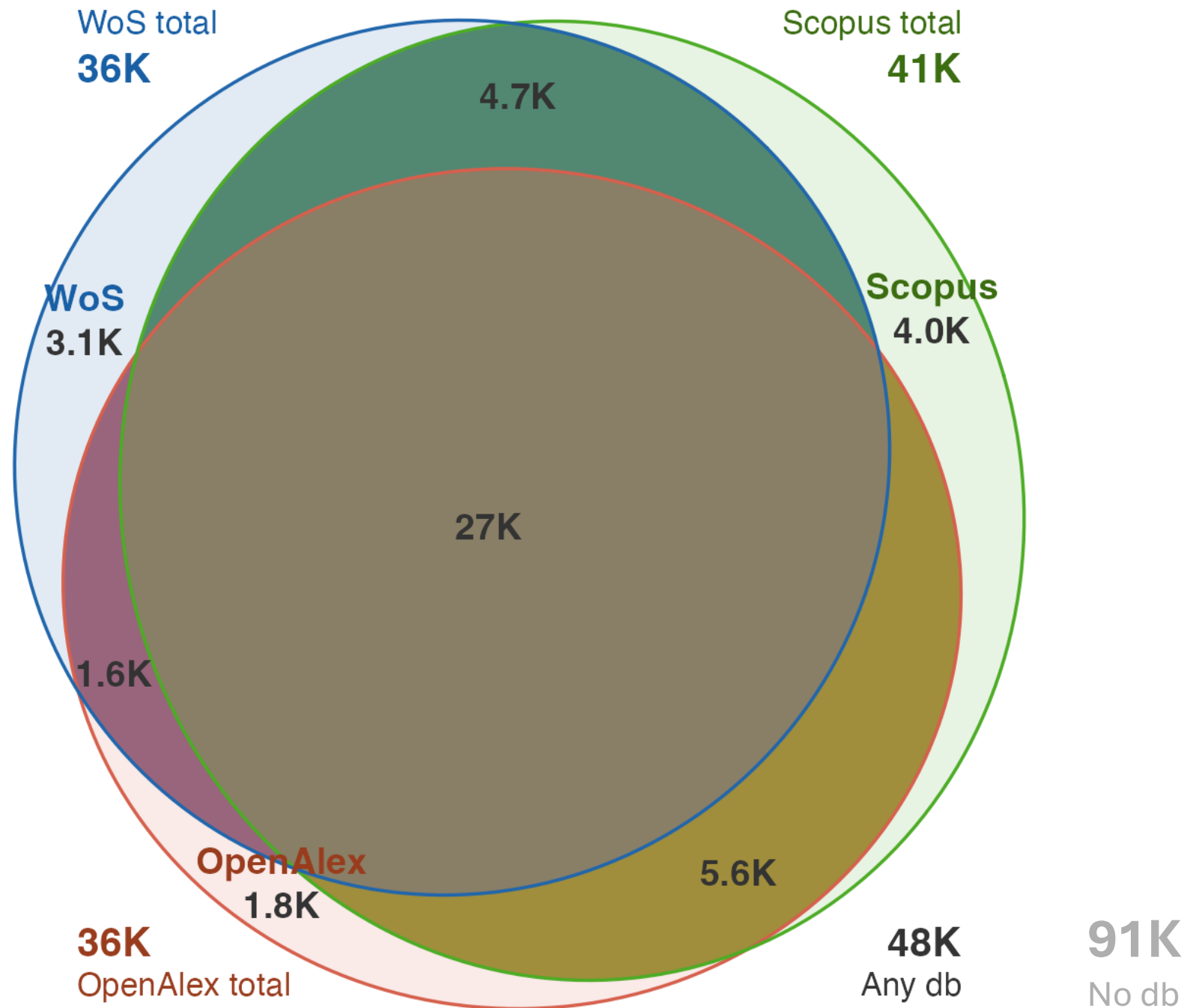
V3S publications coverage in OpenAlex (1996–2025): 5-year windows



**V3S publications coverage in OpenAlex (1996–2025)**



Which  
databases are  
linked  
together?



# OpenAlex: DOIs are essential for the linkage

- So we checked and improved the DOIs in V3S
- Then again ...
- And again ...

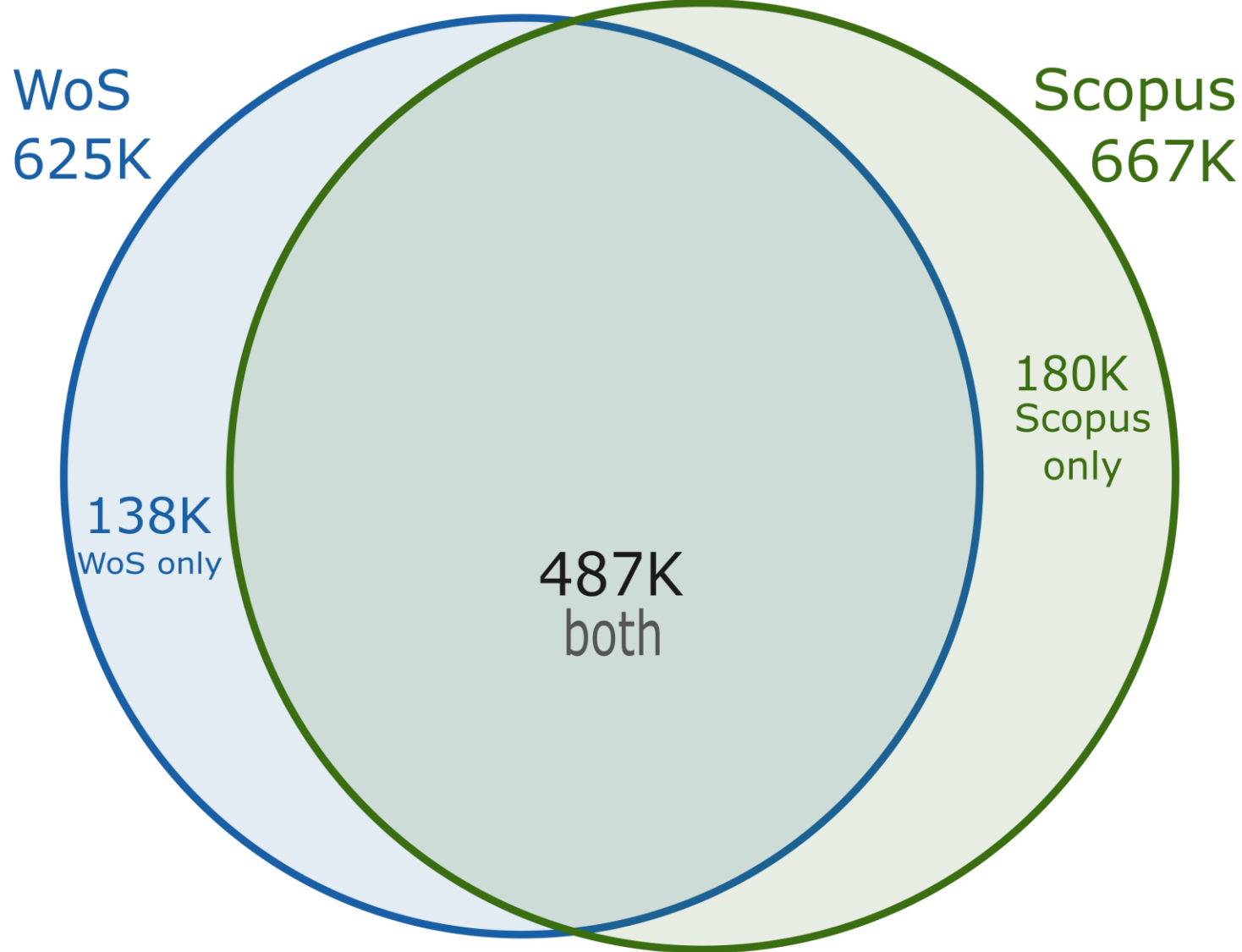
(to be repeated)

We check specifically:

- Title (similarity threshold) — subtitles generally not in OpenAlex
- Year ( $\pm 2$ )
- Types compatible — OpenAlex type "article" includes conf. papers

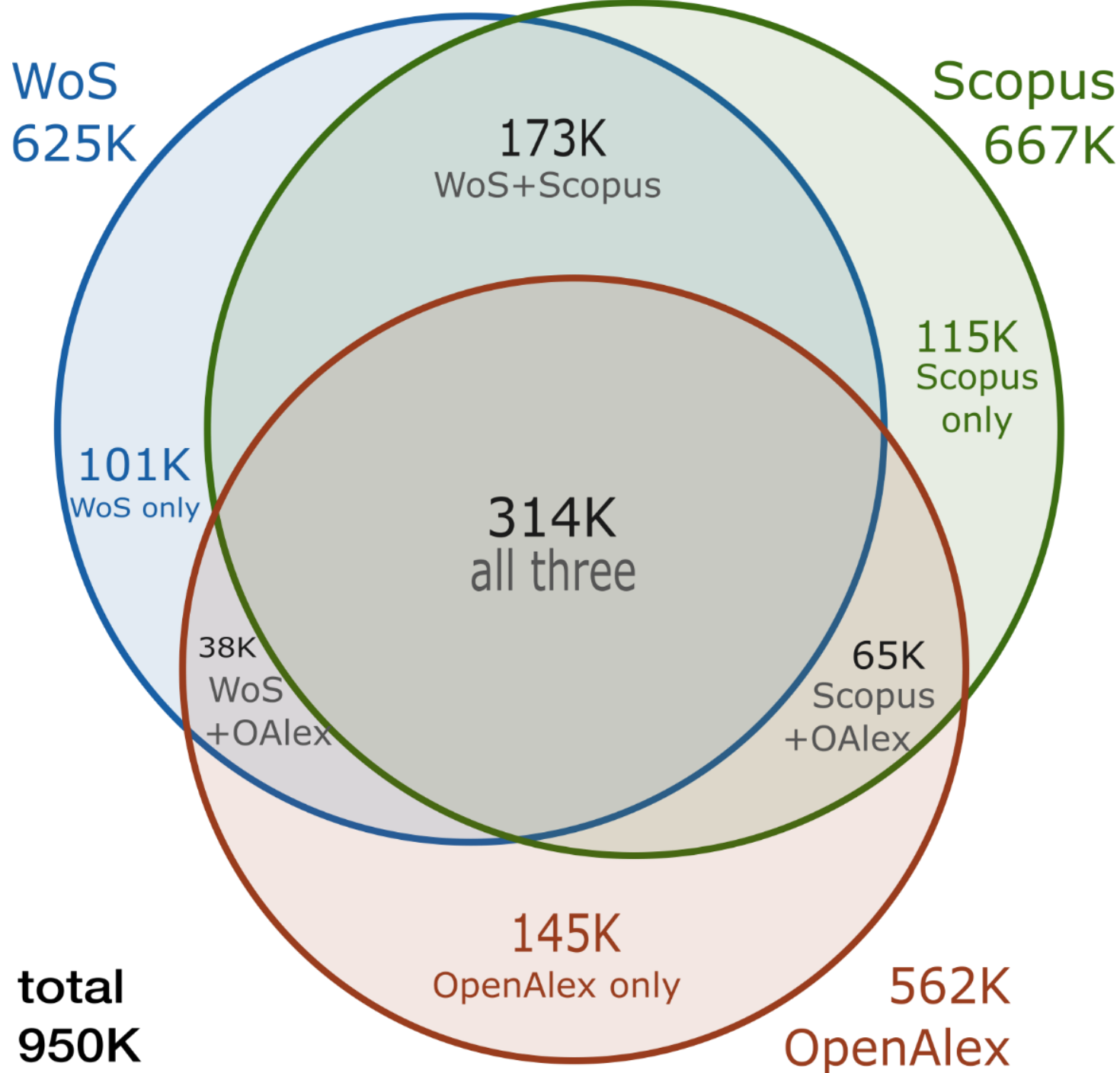
Citation coverage

# Citation Coverage before OpenAlex



total  
805K

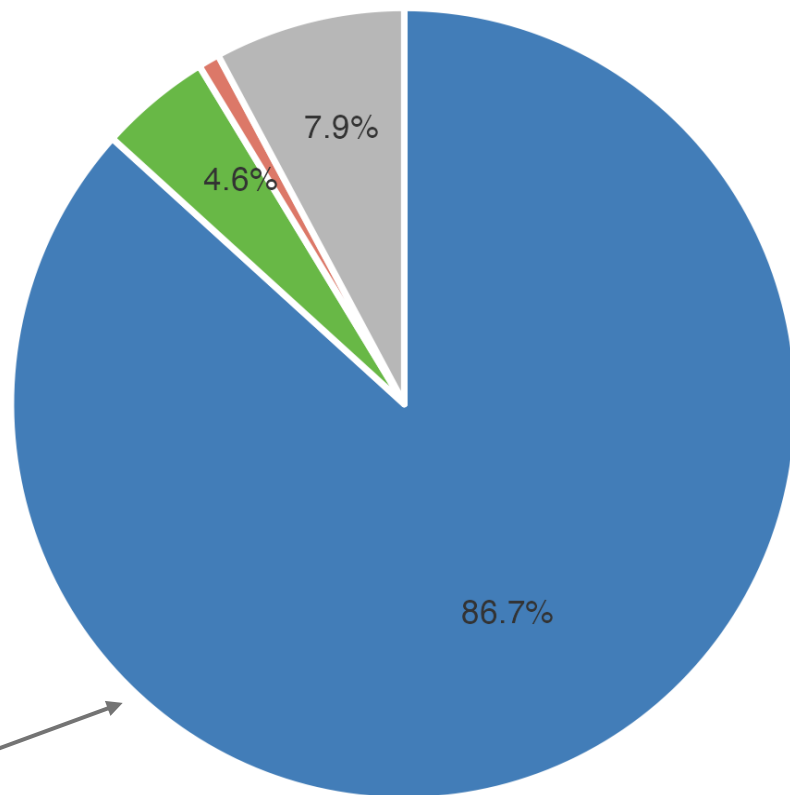
# Citation Coverage with OpenAlex



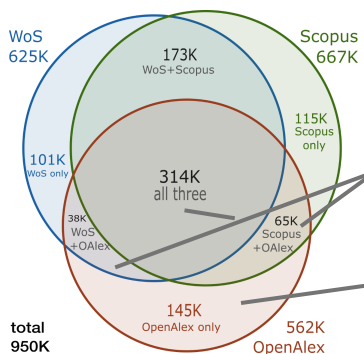
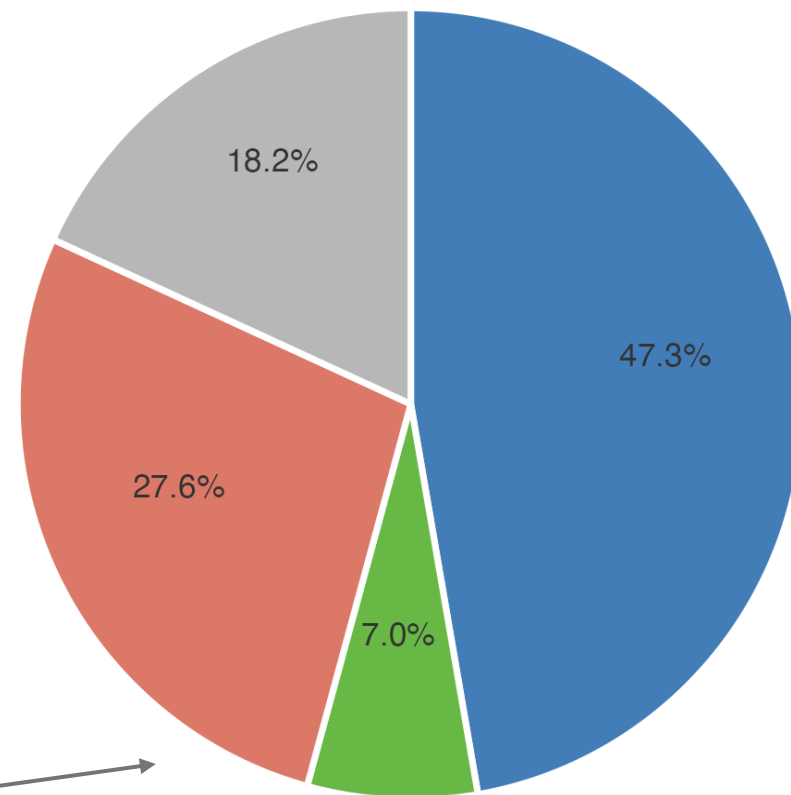


# What sources do OpenAlex citations come from ?

417K ~ 74%  
OpenAlex  $\cap$  other database



145K ~ 26%  
OpenAlex only



Journal

Conf. Proceedings /  
Book Series

Repository /  
e-Book Platform

(unspecified)

So what can OpenAlex do for us?

# Experience with OpenAlex

- Clearly designed, reliable, fast API
  - Very responsive user support for the API
  - Not so for curation requests
  - Coverage looks good, esp. if DOIs for older works are added
- 
- However, some high-level figures raise suspicion

# Czech Technical University in Prague

Institution

institution's profile in OpenAlex

Homepage: [cvut.cz](http://cvut.cz)

Alternate name: České vysoké učení technické v Praze

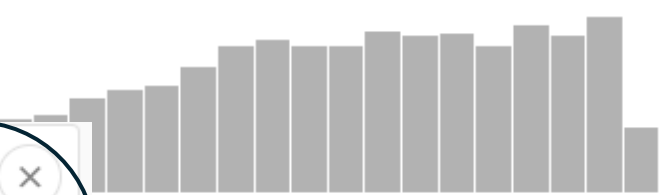
Child institutions: [Research Infrastructure for Experiments at CERN](#), [VR-1 – Training Reactor for Research Activities](#), [Czech National Infrastructure for Biological Data](#), [Research Infrastructure for Fermilab Experiments](#), [Brookhaven National Laboratory - Participation of the Czech Republic](#) +2 more

ROR: 03kqpb082

Works count: 49 650

Citations count: 4 072 000

Year



Search 480M scholarly works

works



boolean

xpac



Institution

is

Czech Technical University in Prague



+ Add filter Clear

OpenAlex search by institution

about 47 460 works

793 700 citations



Open Access



44,8%  
21 250

|                       |       |
|-----------------------|-------|
| ical and experimental | 1 837 |
| ollisions Research    | 1 034 |
| ologies               | 636   |
| ns and Diagnostics    | 529   |
| aterials Research     | 505   |

about 47 450 works

# Is this really our most cited paper?

## MizAR 60 for Mizar 50

2023 · Jakubův, Jan, Chvalovský, Karel, et al. · *DROPS (Schloss Dagstuhl – Leibniz Center for Informatics)* ·

75 682

PDF

Web of Science™

Smart Search

Advanced Search

Research Assistant ⓘ

Sign In ▾

Register



MENU



MizAR 60 for Mizar 50MizAR 60 for Mizar 50

Full text at publisher

Export ▾

Add To Marked List

< 1 of 1

## MizAR 60 for Mizar 50

By

Jakubuv, J (Jakubuv, Jan) <sup>[1]</sup>; Chvalovsky, K (Chvalovsky, Karel) <sup>[1]</sup>; Goertzel, Z (Goertzel, Zarathustra) <sup>[1]</sup>  
; Kaliszyk, C (Kaliszyk, Cezary) <sup>[2]</sup>, <sup>[3]</sup>; Olsak, M (Olsak, Mirek) <sup>[4]</sup>; Piotrowski, B (Piotrowski, Bartosz) <sup>[1]</sup>  
; Schulz, S (Schulz, Stephan) <sup>[5]</sup>; Suda, M (Suda, Martin) <sup>[1]</sup>; Urban, J (Urban, Josef) <sup>[1]</sup>

???

### Citation Network

In Web of Science Core Collection

4  
Citations

Create citation alert

# Conclusions

- At the system level, OpenAlex is great. Get your DOIs right!
- Data quality issues under the surface. *Caveat emptor!*
- Institution's publication output can be verified using the CRIS.
- But not the citation links!