

SCUOLA  
NORMALE  
SUPERIORE

# Comparing Institutional Repository and OpenAIRE Research Products.

Scuola Normale Superiore Case Study

**A. Cignoni**, D. Marotta, D. Tamagno  
EUNIS 2026 - 3rd June 2026

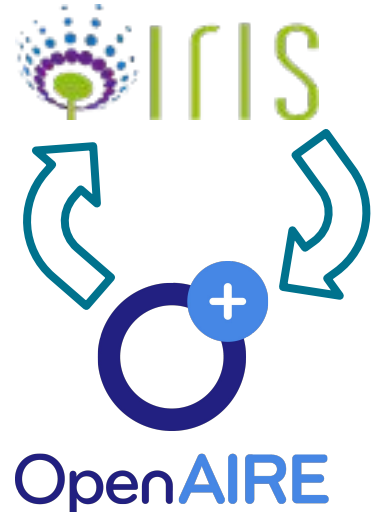
Slides available on Zenodo: <https://doi.org/10.5281/zenodo.20406388>



## IRIS and OpenAIRE: improving together?

---

- SNS's **Institutional Repository**: IRIS
  - Proprietary and commercial service dependent
- **OpenAIRE**: Open Scholarly Communication Infrastructure
  - It offers scientific publications metadata through APIs
- **Research questions**:
  1. Can OpenAIRE support the population of institutional repositories?
  2. How reliable are OpenAIRE affiliations against IRIS SNS data?



## Collecting metadata is not that simple



- **Publication from Institutional Repository (PfIR)**  
Publication in IRIS

- **Author List (AL)**  
Currently affiliated researchers with ORCID



- **Affiliated Products (AP)**  
Publication with at least one SNS author
- **Products from Author List (PfAL)**  
Publication linked to ORCIDs from AL



**OpenAIRE (OA)**  
Set-theoretic union of AP and PfAL

## A simple Python script

---

### Implementation logic:

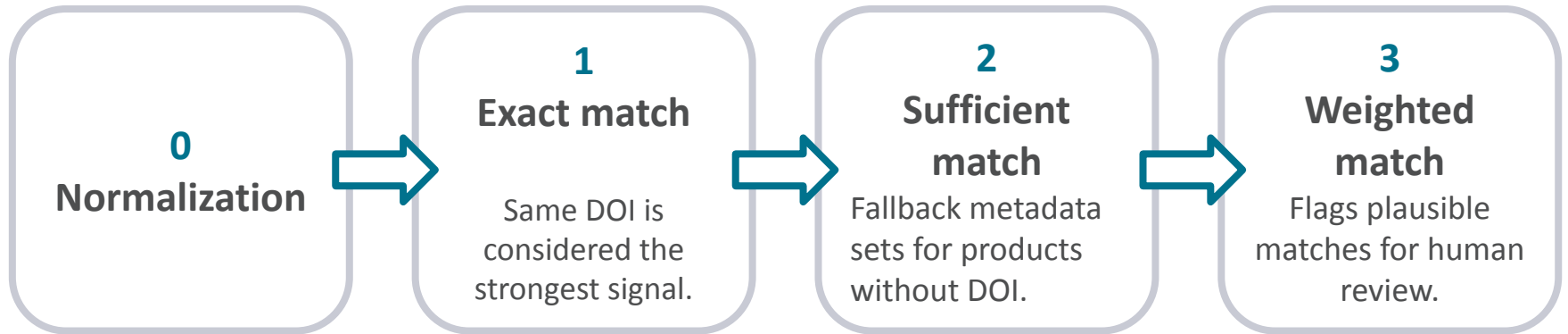
- Simplicity: end-users are administrative staff
- Reusability: input from any repository that conforms to the data model



Code available on GitHub: <https://github.com/alessandrocignoni-sns/IRVOApy>

# Identifying publications in commons

**Design choice:** no fuzzy distance algorithms; uncertain cases left for manual validation.



## 2. Sufficient rules:

1. suff1 (Book chapters without DOI): title + publisher + container + year
2. suff2 (Papers in journals without DOI): title + ISSN + volume/issue
3. suff3 (Products without DOI): authors + title + year

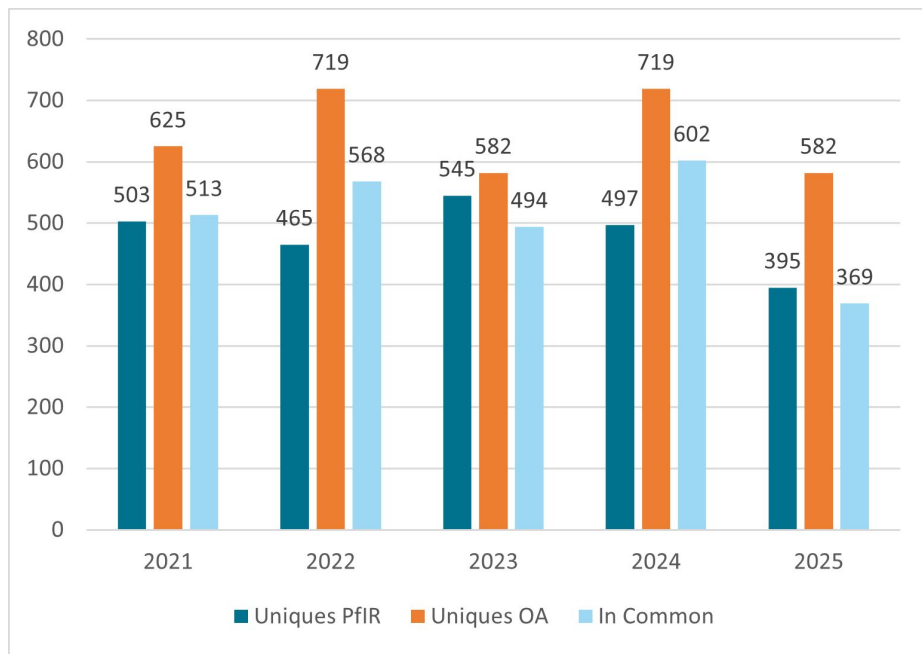
## Benchmarking OpenAIRE using SNS

---

- **Year by year analysis:** year metadata always matches.
- **Covered period:** 2021-2025.
- **AL:** SNS-affiliated authors with an ORCID, regardless of their role.
- **PfIR:** products stored in IRIS of the covered period, regardless of author's role.
- **Extraction date:** 19th February 2026.

Input and output datasets available on Zenodo: <https://doi.org/10.5281/zenodo.18716059>

## OpenAIRE could substantially enrich IRIS

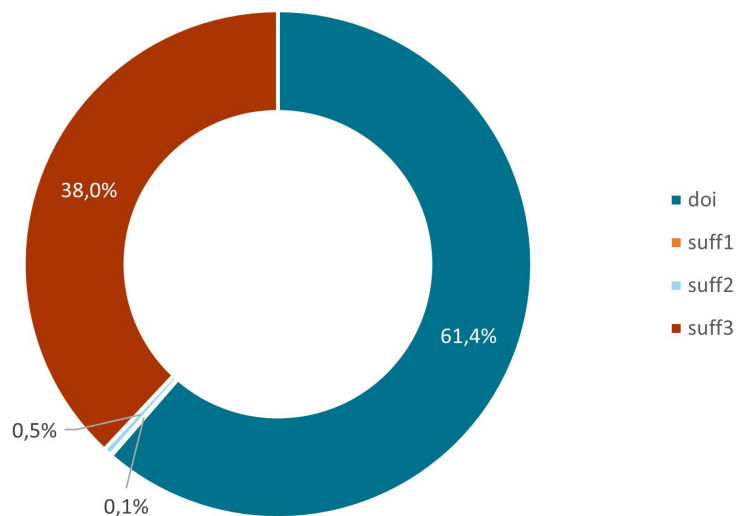


**65%**

potential increase in  
IRIS from OA Uniques

- OA and PfIR are comparable in magnitude
- 2025 is lower in both sources
- Bulk import not safe yet, **human oversight** is needed

## Most automatic matches come from DOI

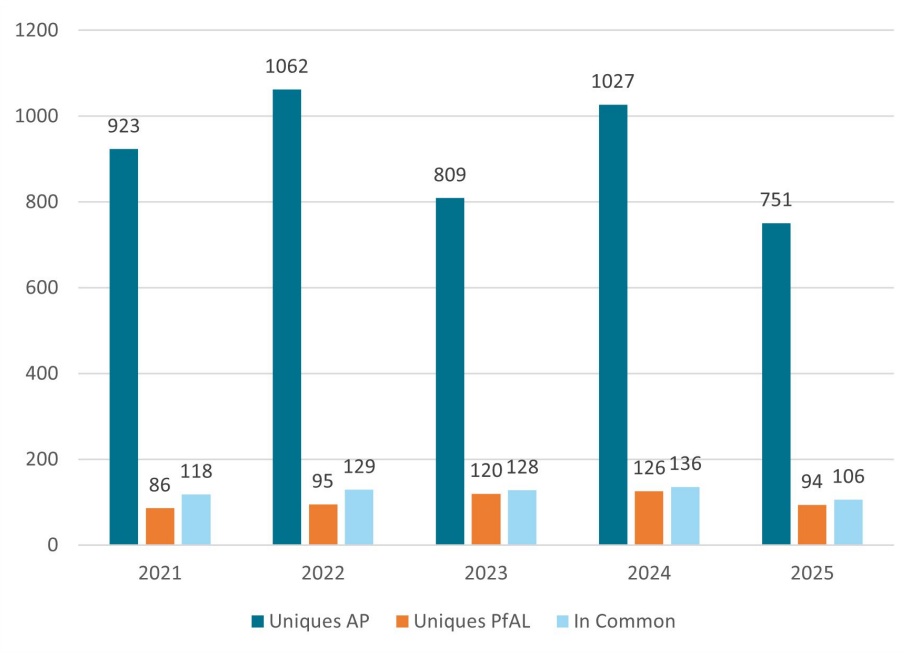


**38%**

of matches are from  
products without DOI

- DOI matching is prioritized.
- Book chapters without DOI likely not harvested by OpenAIRE.
- suff3 can create **false positives** without venue/type checks.

## Affiliation tracking is the bottleneck



**30%**

of AL have some  
products in OpenAIRE

- OpenAIRE records organizations on research products, but not authors' affiliations.
- OpenAIRE do not assigning publications from names.

## How does OpenAIRE perform?

---

### Can OpenAIRE support the population of institutional repositories?

- Substantial number of possible integrable products
- Not fully equivalent to IRIS
- Mismatches imply human validation
- Metadata quality and product identification limit bulk import

### How reliable are OpenAIRE affiliations against IRIS SNS data?

- Products affiliation is not reliable
- There is no name-only attribution
- AP and PfAL are just proxies:
  - AP includes no longer affiliated researchers
  - PfAL is based on ORCID

## Promising, but not yet automatic

OpenAIRE will be a viable path toward **institutional repositories independence** from commercial platform.



Before bulk import can be implemented:

### OpenAIRE

- Person-level affiliation tracking is being implemented.
- Data quality can be improved with ORCID exposure in institutional repositories.
- Improve DOI handling and product-type mapping



### Matching script

- User interface
- Weight recalibration through statistical methods
- Reporting incorrect metadatas to OpenAIRE