



NATIONAL
INFORMATION
PROCESSING
INSTITUTE

Ludzie Nauki: A national CRIS for Poland

Przemysław Korytkowski

Marek Michajłowicz



This work is licensed under CC BY-NC-ND
4.0.

Viena, 16-05-2024

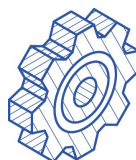
National Information Processing Institute



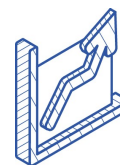
**Software House
for the Ministry**



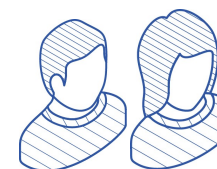
R&D



AI/ML/NLP



Data Science



**Almost 500
employees**

Our products

POLON

OSF
OŚRODEK STRUMIENIA
FINANSOWANIA

radon

NP
NAUKA
POLSKA

Studia

JSA

PBN

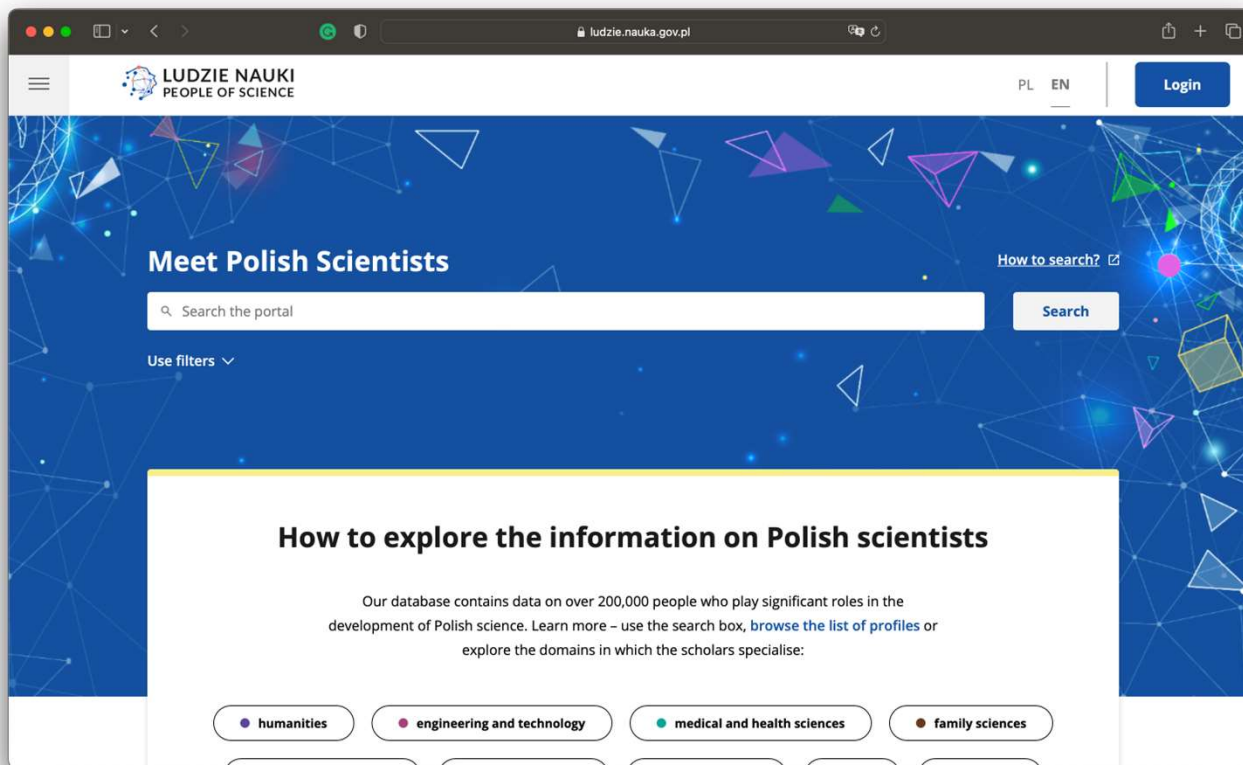
NAVOICA

NAVA

ELA

polon^{2.0}
SEDN
SYSTEM
OCENY
JAKOŚCI
NAUKOWEGO

https://ludzie.nauka.gov.pl/



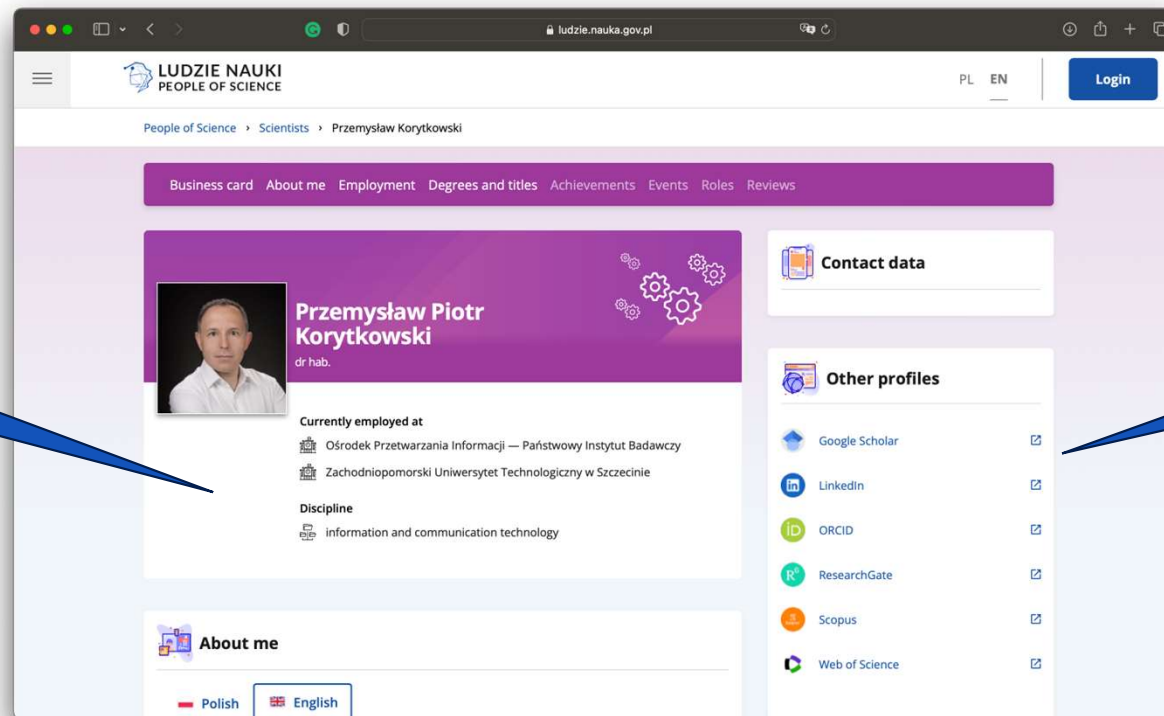
Ministry of Education and Science
Republic of Poland

Open Government Data

FAIR data:

- findable
- accessible
- interoperable
- reusable

Profile



Employment
history

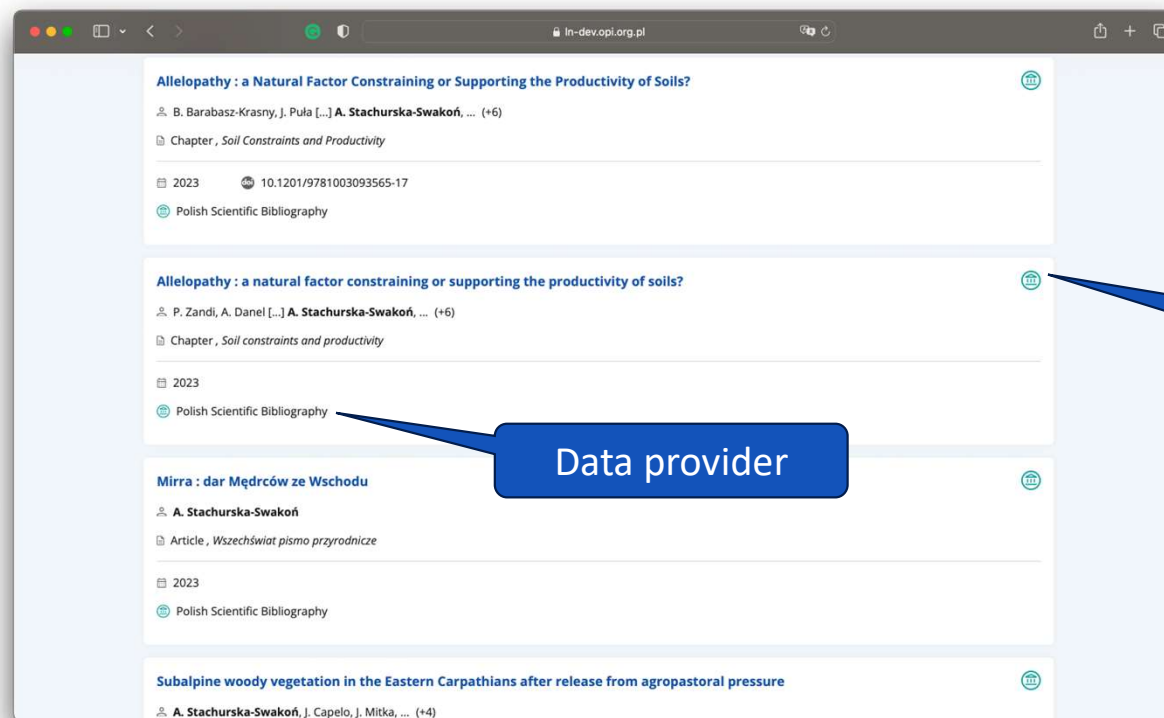
Links to external
profiles

Reliable information

The screenshot shows a web browser window with the URL `ludzie.nauka.gov.pl`. The page is titled "Scientific degrees and titles" and lists three entries, each with a year, a degree title, and details about where it was awarded and registered. Two blue callout boxes point to specific parts of the entries: "Source database" points to the "Degree awarded by title" field, and "Data provider" points to the "Degree registered by title" field.

Year	Degree Title	Degree awarded by title	Degree registered by title
2017	Habilitated Doctor	technology / computer and information science	Wydział Informatyki Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
2005	Doctor	Nauki techniczne / Informatyka	Politechnika Szczecińska Ośrodek Przetwarzania Informacji — Państwowy Instytut Badawczy
2002	Master of Science in Engineering	Informatyka	Politechnika Szczecińska Ośrodek Przetwarzania Informacji — Państwowy Instytut Badawczy

Reliable information



National CRIS ecosystem in Poland



1999-2018

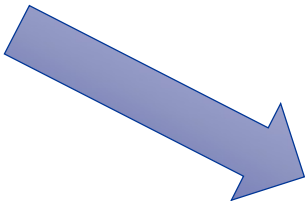
- scientific degrees obtained
- scientific disciplines/fields
- history of employment at Polish institutions
- doctoral supervisions
- dissertation reviews
- history of academic duties



<https://nauka-polska.pl>



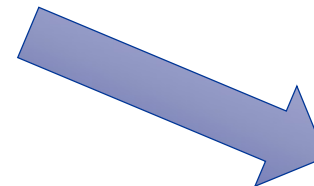
Institutional CIRS



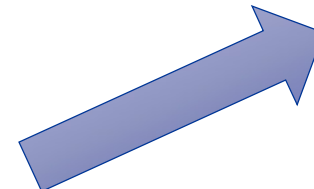
polon^{2.0}

Since 2013

- scientific degrees obtained
- scientific disciplines/fields
- history of employment at Polish institutions
- patents (national, EPO, PCT)
- projects
- doctoral supervisions
- dissertation reviews
- research infrastructure
- ORCID



Ministry



PBN

Polska Bibliografia Naukowa

Since 2013

- scientific publications metadata

polon^{2.0}
SEDN SYSTEM
EWALUACJI
DOROBKU
NAUKOWEGO

Research evaluation

National CRIS ecosystem in Poland

polon^{2.0}

Since 2018

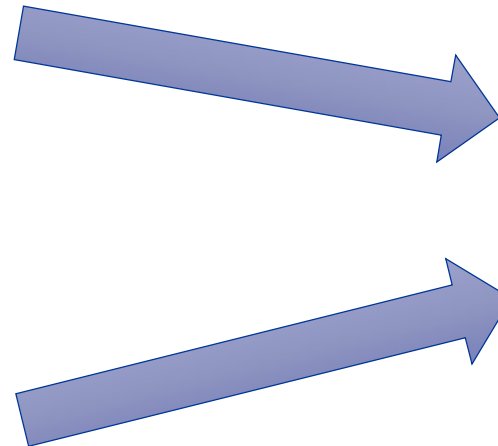
- scientific degrees obtained
- scientific disciplines/fields
- history of employment at Polish institutions
- patents (national, EPO, PCT)
- projects
- doctoral supervisions
- dissertation reviews
- research infrastructure
- ORCID



PBN

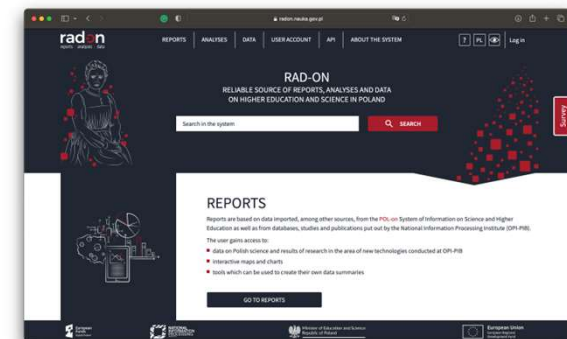
Polska Bibliografia Naukowa

- scientific publications metadata



radon

reports | analyses | data



<https://radon.nauka.gov.pl>

National CRIS ecosystem in Poland

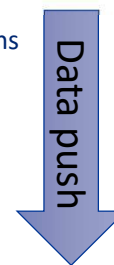
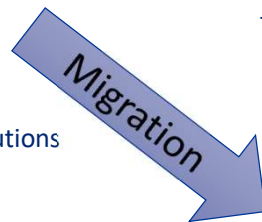


1999-2018

- scientific degrees obtained
- scientific disciplines/fields
- history of employment at Polish institutions
- doctoral supervisions
- dissertation reviews
- history of academic duties

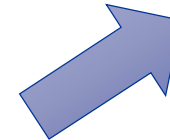
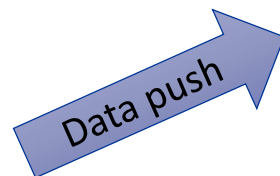


- scientific publications



Since 2018

- scientific degrees obtained
- scientific disciplines/fields
- history of employment at Polish institutions
- patents (national, EPO, PCT)
- projects
- doctoral supervisions
- dissertation reviews
- research infrastructure
- ORCID



Research funding organisations



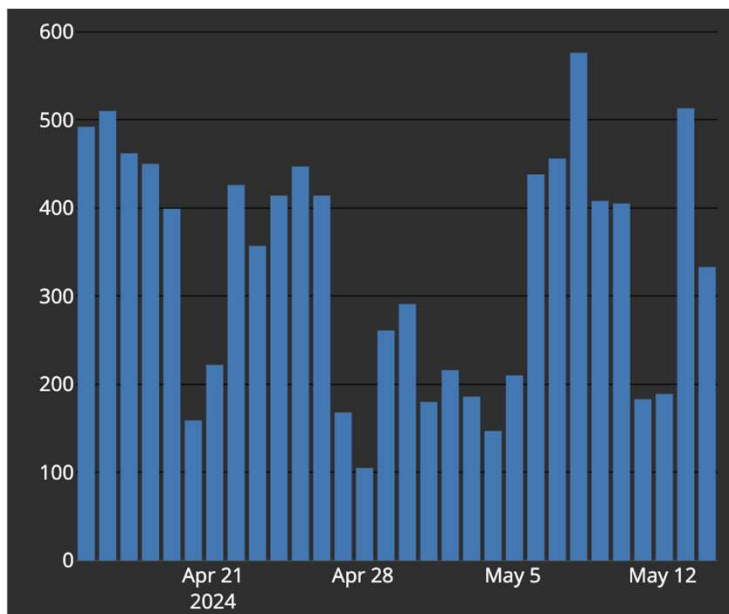
The goals of *Ludzie Nauki*

- To increase the scope of data on science in Poland that is freely available to the public.
- To provide Polish scientists with a single reliable data source.
- To increase the visibility of scientists by establishing a single 'source of information' on their scientific achievements.
- To exploit the data collected in several national systems over the years and consume it in various areas, such as research evaluations, grant applications, and scientific promotions.
- To provide administrative support by reducing reporting obligations.
- To enable businesses and members of the public to contact and search for information on scientists.

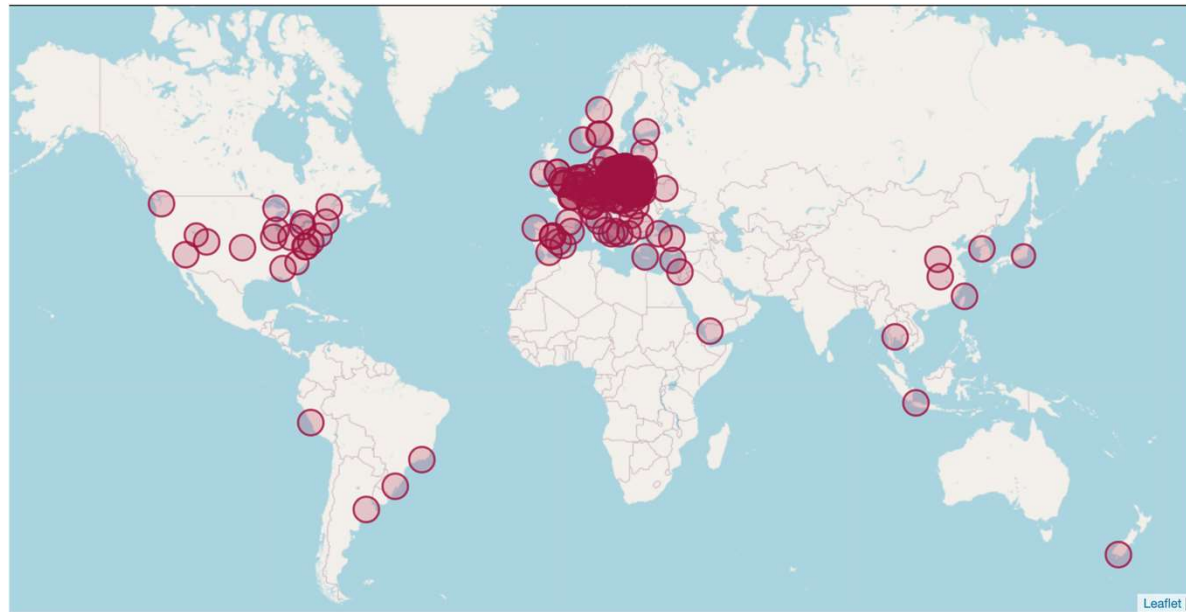
Ludzie Nauki in numbers

206 050	researchers' profiles	32 000	projects
641 000	journal research papers	102 000	monographs
319 000	chapters	11 600	patents
5 000+	PhD dissertations per year	+ more	

Usage statistics



Daily number of visits to the homepage



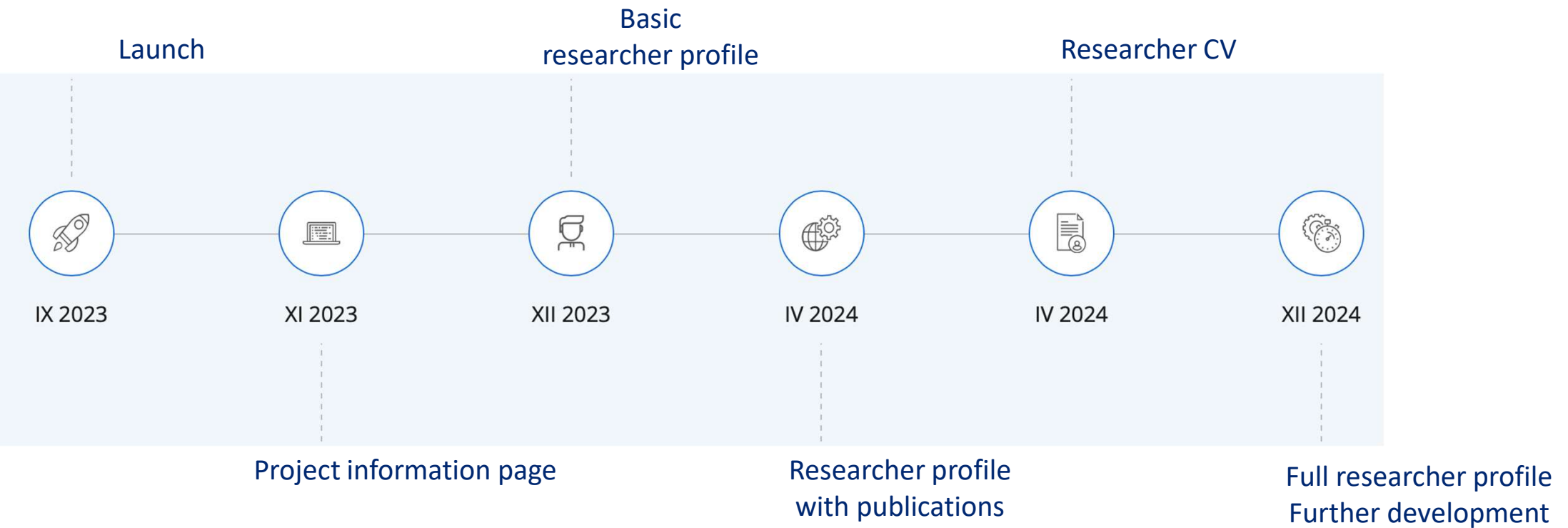
Interoperability



ORCID



Project timeline



Challenges

- Deduplication of publication metadata
- Interoperability
- Semantic search engine



Deduplication of publication metadata

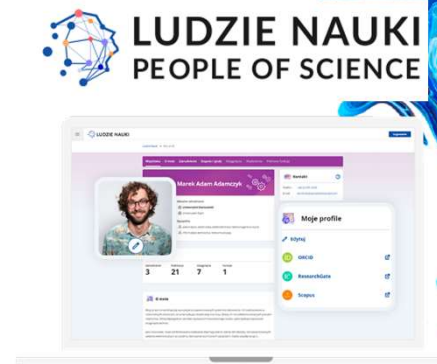


verification
and enrichment

Golden record

deduplication

PostgreSQL



Hadoop + Spark Cluster:

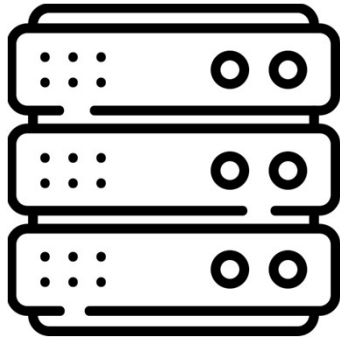
- Artificial Intelligence
- Natural Language Processing



In the future:



Hadoop + Spark Cluster

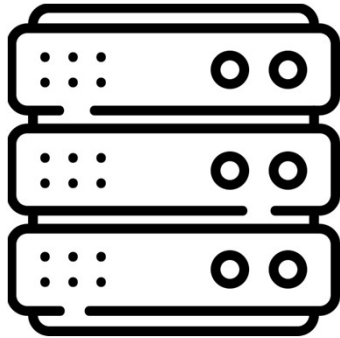


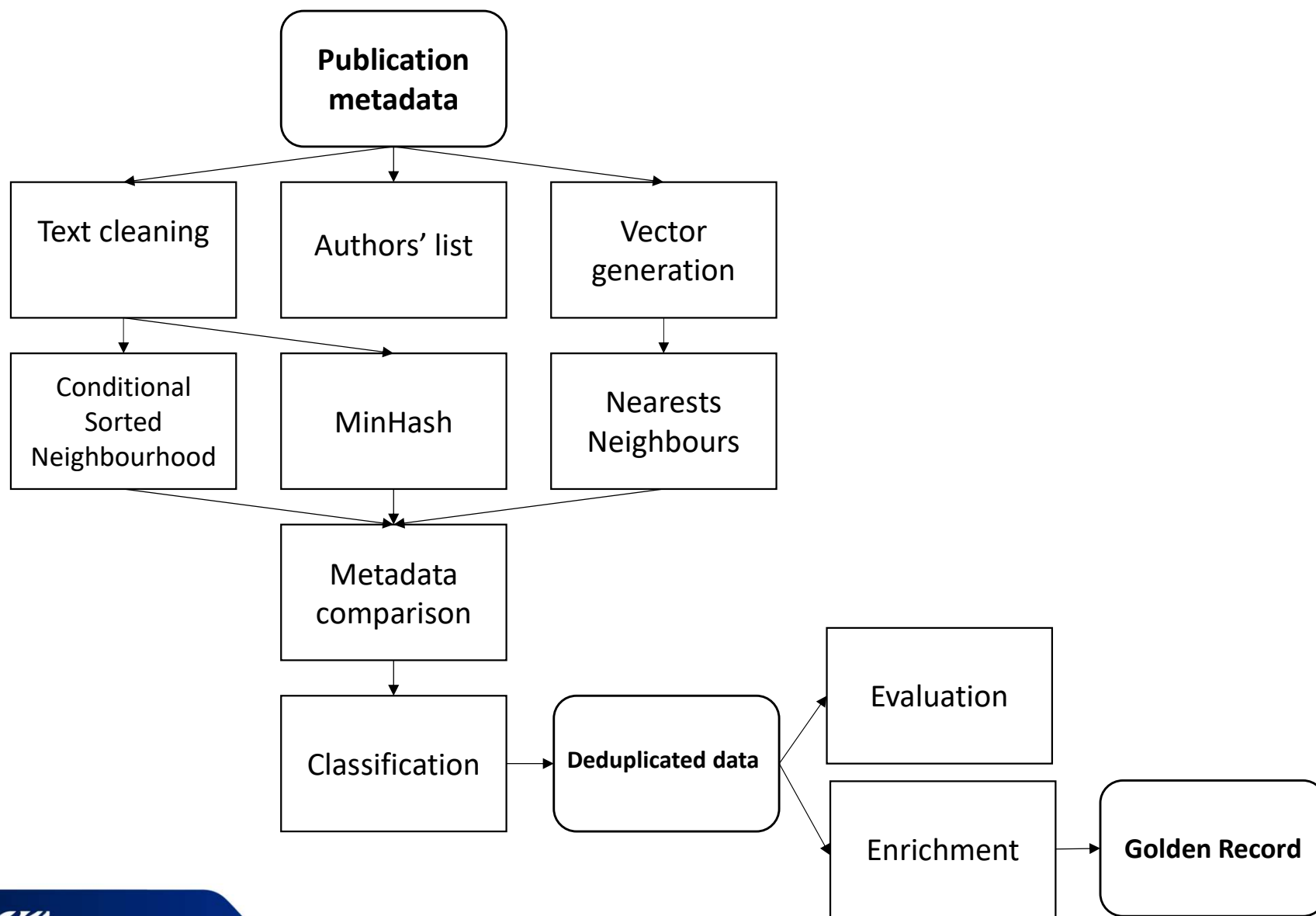
22 servers (each):

- 16 cores
- 90 GB RAM
- 2 TB disks

352 cores, 1.98TB RAM, 44TB disks

Hadoop + Spark Cluster







NATIONAL
INFORMATION
PROCESSING
INSTITUTE

Thank you for your attention!

Przemysław Korytkowski

przemyslaw.korytkowski@opi.org.pl

National Information Processing Institute
al. Niepodległości 188 B
00-608 Warsaw
Poland

phone: +48 22 570 14 00
faks: +48 22 825 33 19
e-mail: opi@opi.org.pl
www.opi.org.pl/en