Ludzie Nauki: A new CRIS for Poland

Przemysław Korytkowski

National Information Processing Institute – National Research Institute, Warsaw, Poland

The National Information Processing Institute (OPI PIB) has launched a new project using a national CRIS called *Ludzie Nauki* ('People of Science'), financed by the Ministry of Education and Science of Poland. The project has been running since September 2023 and the first phase will conclude at the end of 2024.

Ludzie Nauki is intended to constitute a central and reliable source of data on Polish scientists, which will be updated and verified continuously. The system will make available the public profiles of scientists and serve as a reliable source of information. *Ludzie Nauki* will integrate data from multiple existing science and higher education systems, including POL-on, PBN, and *Nauka Polska*. *Ludzie Nauki* will allow scientists to edit and extend data on themselves. The data will be made available through APIs, which enables its further processing and the provision of new services. The system will implement the principles of FAIR (findable, available, interoperable and reusable) data and the concept of Open Government Data.

The goals of *Ludzie Nauki* are:
- 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.

Data on scientists has been collected nationally in Poland since 1990 and has been partially available on the *Nauka Polska* service (https://nauka-polska.pl/) since 1999. The RAD-on service (https://radon.nauka.gov.pl/) has provided raw data, reports, and analysis on higher education and science since 2020. Currently, data from the various systems is unintegrated, which means that there is no single place to find information on all Polish scientists: data on degrees and employment can be found on *Nauka Polska*; data on publications on PBN and on RAD-on; data on doctorates before 2018 on *Nauka Polska*, and after 2018 on RAD-on.

Ludzie Nauki will provide data on all scientists and doctoral students currently employed at Polish research institutions and historical data that covers the last twenty years. The system

will contain information that pertains to approximately 200,000 individuals. Their profiles will include:
- scientific degrees obtained
- scientific disciplines/fields
- history of employment at Polish institutions
- scientific publications
- patents (national, EPO, PCT)
- projects
- doctoral supervisions
- dissertation reviews
- history of academic duties
- available research infrastructure
- research teams
- external identifiers (e.g. ORCID, WorldCat, ISNI, VIAF)

Ludzie Nauki will include a semantic search engine, which will allow users to search for researchers not only using keywords; the system will also employ deep learning and natural language processing approaches. It will be designed to locate potential dissertations or grant proposal reviewers, research project partners, experts in specific fields, and industrial consultants.

One of the more complex parts of *Ludzie Nauki* is the module that processes metadata on scientific publications. Since 2013, publications by Polish scientists (currently 700,000) have been collected on the Polish Scholarly Bibliography (PBN)—a system developed with research evaluation in mind. Because PBN focuses on scientific institutions, deduplication and data cleaning is necessary. We plan to enrich the publication dataset through integration with external systems, such as the Polish National Library, CrossRef, OpenAlex, OpenCitation, and WikiData. Our aim is to gather the most comprehensive information on publications by Polish scientists in one place.

The project will introduce the ability to import and export data in CERIF format. In addition, the data model will be based to some extent on the CERIF data model.

In subsequent phases, we plan to integrate *Ludzie Nauki* with the systems of national research funding agencies so that applications may be prefilled.