



# Ludzie Nauki\*: Data aspects of the Polish national CRIS

Emil Podwysocki<sup>1</sup>, Maria Bylina<sup>1†</sup>, Jacek Raczko<sup>1</sup>, Michał Ulaniuk<sup>1</sup>, Marek Michajłowicz<sup>1</sup> and Rafał Jendrzewski<sup>1</sup>

<sup>1</sup>The National Information Processing Institute, Warsaw, Poland  
emil.podwysocki@opi.org.pl, maria.bylina@opi.org.pl,  
jacek.raczko@opi.org.pl, michal.ulaniuk@opi.org.pl,  
marek.michajlowicz@opi.org.pl, rafal.jendrzewski@opi.org.pl

## Abstract

This article considers the intricacies of data management in the context of the implementation of the Polish Ministry of Science and Higher Education's latest initiative: the *Ludzie Nauki* platform. The article's primary objective is to underscore the importance of high-quality data. Supported and funded by the Polish government, *Ludzie Nauki* is linked closely with the national current research information system (CRIS). A critical aspect of this initiative is the management of the metadata associated with scientific publications, which integrates data from diverse sources to enrich the publication dataset. Business needs necessitate that the platform serve as a reliable source of information on Polish scientists, supporting scientific careers, and fostering collaboration between science and industry. *Ludzie Nauki* aligns with European Union data management trends, as well as adhering to FAIR principles, open science initiatives, and the Data Governance Act. It also interfaces with various CRIS systems, promoting interoperability and standardisation. The article also discusses challenges and solutions in data governance, technical implementation, and the awarding of golden records for scientific achievements, which demonstrates *Ludzie Nauki*'s commitment to enhancing research's visibility and impact with robust data management practices.

## 1 Introduction

This article discusses the data aspects of the implementation of the latest project of the Polish Ministry of Science and Higher Education, *Ludzie Nauki*. Its main goal is to describe the essence of

---

\* Polish Scholars

† <https://orcid.org/0000-0003-3551-4709>

possessing high quality data and how we achieved that. *Ludzie Nauki* is supported and financed by the Polish government.

Since 1990, information on scientists in Poland has been gathered systematically and nationally, and since 1999, some of this data has been accessible through the *Nauka Polska* platform (<https://nauka-polska.pl/>). The RAD-on platform (<https://radon.nauka.gov.pl/>), established in 2020, presents comprehensive data, reports, and analyses that relate higher education and scientific research. At present, however, these various data sources remain unconnected, which has resulted in a fragmented landscape in which information on Polish scientists is scattered across multiple platforms. For instance, academic qualifications and employment details are available on *Nauka Polska*, while publications are available on the PBN and RAD-on platforms. Moreover, data on doctoral degrees awarded before 2018 is stored on *Nauka Polska*, while information on those awarded after 2018 is stored on RAD-on (Korytkowski 2023).

*Ludzie Nauki* will contain information on all researchers and doctoral candidates presently engaged with Polish research establishments, along with historical records spanning the last two decades. The database will contain information on approximately 200,000 individuals.

One of the more intricate components of *Ludzie Nauki* is the module responsible for handling the metadata associated with scientific publications. Since 2013, publications authored by Polish scientists, totalling around 700,000, have been amassed on the Polish Scholarly Bibliography (PBN)—a system designed with research evaluation in mind. Given PBN's focus on scientific institutions, deduplication and data cleansing are imperative. Our strategy involves enhancing the publication dataset through integration with external systems, such as the Polish National Library, Crossref, OpenAlex, OpenCitation, and WikiData. The project's overarching goal is to aggregate the most comprehensive information on publications authored by Polish scientists on a single platform (Korytkowski 2023).

## 2 Business needs

*Ludzie Nauki* is the newest project of the Polish Ministry of Science and Higher Education, which aims to promote Polish scientists and their achievements. It is the largest, best-organised, and most reliable data collection of Polish scientists. The platform is the main source of the truth for all types of data connected with science in Poland, including publications, projects, and patents. It will also offer e-services to scientists, and facilitate cooperation between science and business.

*Ludzie Nauki* responds to several needs, including:

- the promotion of Polish science by presenting the achievements of Polish scientists and their profiles
- the facilitation of access to knowledge on Polish scientific output by collecting data on over 200.000 Polish scientists in one place
- support for the scientific careers of Polish researchers by offering e-services, such as help in the creation of scientific CVs, the generation of reports, grants applications, and participation in competitions and elections to specialist committees
- integration of the business and science communities by facilitating collaboration between scientists and entrepreneurs, public institutions, and nongovernmental organisations.

The implementation of *Ludzie Nauki* will enhance the social impact of research, as the public will enjoy access to wide-ranging, reliable, and trustworthy information on Polish researchers and their achievements. This will be accomplished with minimal effect on researchers' workload, as much of the

information presented on the platform is already contained elsewhere. *Ludzie Nauki*'s semantic search engine with deep learning and natural language algorithms will enable users to search using natural language. It will be designed to discover potential dissertations or grant proposal reviewers, research project partners, experts in specific fields, or industrial consultants.

### 3 Trends in data management and policies across the European Union

The European Commission, on behalf of the European Union (EU), advocates actively for openly accessible research (known as open science) and subsequently, for open access to research data (known as open data). As a result, the importance of strategies that aim to ensure the accessibility of research data has grown considerably, particularly in research projects funded by the EU. This emphasis on accessibility resonates with other research funding bodies and policymakers. Effective coordination and interoperability between local systems, infrastructures, policies, and EU counterparts are crucial to the advancement of research data processes at all levels<sup>‡</sup>.

*Ludzie Nauki* conforms to all of the data requirements of the European Commission, including FAIR data principles, open science, Horizon 2020, and the Data Governance Act. By adhering to these regulations and policies, the platform ensures the ethical and efficient handling of data, promoting transparency, accessibility, and reusability.

At present, the dominant approach to accessing researchers' data is open science. The development of science depends on free access to high-quality and reliable data.

The primary objective of the Data Governance Act is to establish a legal structure for data sharing across the European single market. It aims to ensure fair access to data, promote interoperability, and prevent lock-in effects (recital 2 Data Governance Act). The European Parliament adopted the Data Governance Act on 6 April, 2022.

FAIR principles are essential in effective data management. FAIR data must be findable, accessible, interoperable, and reusable. This means that data requires robust metadata, unique identifiers, accessibility to humans and machines, adherence to common structures, and compliance with usage licenses and community standards. Embracing FAIR principles fosters transparency, efficiency, collaboration, and innovation in science.

### 4 Landscape of CRIS systems

In research information management, an array of systems akin to current research information system (CRISs) have proliferated globally, each catering to the diverse needs and contexts of research institutions. Examples include the Finnish Research Information Hub, IEEE Xplore, academia.edu, and ResearchGate. These platforms serve as vital hubs for researchers, facilitating the dissemination of academic output, networking opportunities, and access to research resources. Recognising the significance of interoperability and standardisation in this landscape, *Ludzie Nauki* will endeavour to enhance its capabilities by enabling the import and export of data in the Common European Research Information Format (CERIF). Moreover, the platform's data model will draw upon the principles of the CERIF data model, ensuring compatibility and fostering seamless integration with other CRIS systems and research infrastructures. This strategic alignment underscores the creators' commitment to promoting openness, collaboration, and efficiency in research information management practices globally.

---

<sup>‡</sup> <https://www.umu.se/en/library/research-data/specialised-topics/open-data-and-the-eu/>

The Finnish Research Information Hub is a national initiative that collects and connects information on Finnish science from various sources and releases it publicly on the research.fi platform. The platform contains information on researchers, publications, projects, data, infrastructures, and other research activities, as well as statistics and news on the Finnish research system and its impact. It aims to support the visibility and reuse of research information, collaboration between science and business, and evidence-based policy-making on science. It also offers a researcher's profile tool, which allows researchers to create profiles on the platform and select what information they want to present, using the ORCID and MyData approaches. The Finnish Research Information Hub is owned and financed by the Ministry of Education and Culture of Finland, and is implemented and coordinated by CSC - IT Center for Science Ltd.

Academia.edu is a social network for scientists, which enables the sharing and downloading of scientific publications. Users can create profiles on which they present their data, interests, affiliations, and publications. They can also follow other scientists, join thematic groups, and comment on and cite publications. The service aims to promote open access to knowledge and international cooperation in science.

ResearchGate is a social network for scientists and researchers, on which they can share and access scientific publications, data, and expertise. Users can create profiles, and add their skills and interests. The search engine facilitates the discovery of research.

IEEE Xplore is a digital research library that offers access to articles, conference proceedings, technical standards, and other resources that concern engineering, computer science, and related fields. The platform contains content from approximately 200 journals and magazines of IEEE, over 1,800 annual conferences, over 1,200 active technical standards, 400 e-learning courses, and 4,000 e-books. IEEE Xplore enables the searching, filtering, downloading, and sharing of scientific information, as well as the creation and viewing of visualisations and analyses of data.

The Indian Research Information Network System (IRINS) is a research information management system implemented at Indian higher education institutions, which aims to aggregate and showcase research-related information to the academic community. It operates as a "software as a service" model, and supports academic and R&D organisations across India. Its key functionalities include CV preparation, publication management, research productivity visualisation, coauthor network analysis, and federated search. Integrated with internal and external data sources, IRINS facilitates comprehensive research information. As of June 2022, the platform has onboarded 534 institutions and interconnected over 83,000 faculty members/scientists, accumulating metadata on 1.41 million publications. IRINS plays a crucial role in organising academic communication activities, assisting research administrators, and enhancing research visibility. The system continues to evolve with additional modules and aims to reach more international audiences as an open-source CRIS.

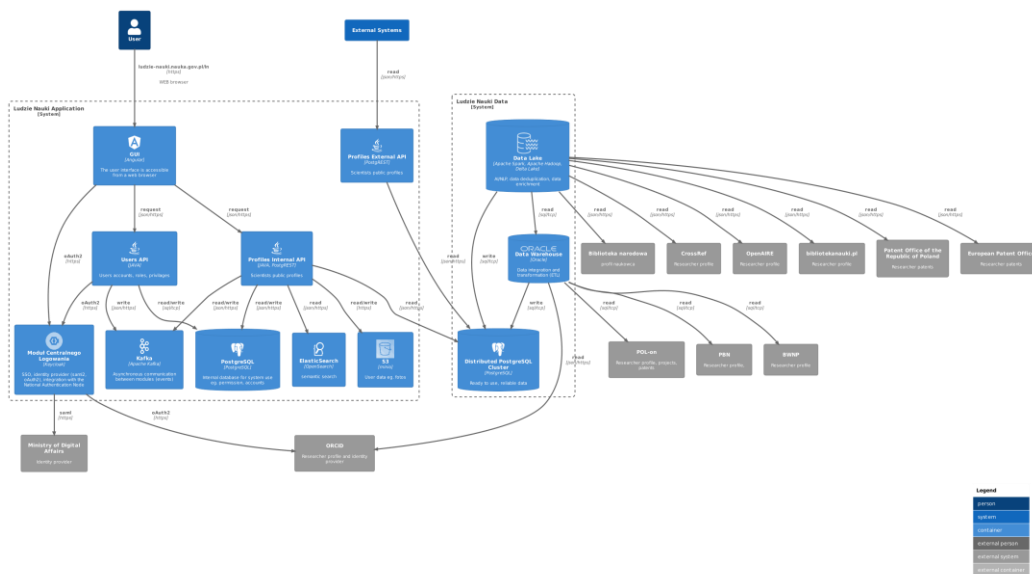
## 5 Technical implementation of the data processing ecosystem

*Ludzie Nauki* is a distributed system based on a microservices architecture. The architecture can be divided into two main parts: application and data. This article focuses on the data aspects of *Ludzie Nauki*'s CRIS implementation, and they are described in detail below.

*Ludzie Nauki* utilises a combination of open-source and commercial technologies. OPI PIB selected trusted and reliable technologies like Oracle, PostgreSQL, Apache (including HDFS, Spark, and Kafka), OpenSearch, Delta Lake, and Amazon S3 (min.io open-source implementation) for the project's implementation. An essential component of the architecture is PostgRest<sup>§</sup>, a standalone web server that seamlessly transforms PostgreSQL databases into RESTful APIs. The API endpoints and operations are determined by the structural constraints and permissions configured in the database.

---

<sup>§</sup> <https://postgrest.org/en/stable/>



**Figure 1.** Architecture overview

A key function of *Ludzie Nauki* is the provision of reliable data collected from trusted sources. After conducting research, OPI PIB selected trusted data sources that should be harvested by *Ludzie Nauki*:

- POL-on – the Information System of Science and Higher Education: an integrated and centralised information system that supports the Polish Ministry of Science and Higher Education, as well as other ministries and institutions related to science and higher education (Michajłowicz et al. 2018; Protasiewicz et al. 2019; Michajłowicz et al. 2022)
- PBN – the Polish Scholarly Bibliography: a database of the achievements of Polish scientists; a system designed with research evaluation in mind (Michajłowicz et al. 2022)
- BWNP: a legacy database of Polish scientists
- UPRP: The Patent Office of the Republic of Poland
- National Library of Poland
- Crossref: a nonprofit digital infrastructure organisation that serves the worldwide academic research community
- OpenAIRE: a European initiative that advocates for and supports open science. It comprises a network of specialised open science professionals who promote and offer training on open science principles. Additionally, OpenAIRE operates as a technical framework that collects research outputs from affiliated data sources
- Open ALEX: developed by OurResearch, a nonprofit dedicated to making research open. It is a free global index of researchers, research institutions, academic journals, and citations
- ORCID: a worldwide, nonprofit organisation that assigns persistent digital researcher identifiers

We identified two integration points, designed for different use cases:

- the data warehouse integrates data from internal ministry systems, such as POL-on and PBN
- the data lake processes external data from sources such as Crossref or OpenAIRE.

The data warehouse plays an important role in the integration process. Implemented in 2019, a central database collects data from internal systems governed by the Polish Ministry of Science and Higher Education. The Polish information ecosystem on science and higher education has a highly heterogeneous data structure. Its ecosystem includes Oracle, MongoDB, PostgreSQL, Elasticsearch, and other technologies. (Podwysocki et al. 2019). The data stored in the data warehouse is reliable and is ideally structured for further processing using artificial intelligence and natural language processing algorithms. The data warehouse is implemented using the Oracle platform, based on several trustworthy tools like Oracle Golden Gate and Oracle Data Integrator, to deliver the highest data quality for analytics, reports, and third-party systems such as *Ludzie Nauki*.

Gartner\*\* defines a data lake as “...a concept consisting of a collection of storage instances of various data assets. These assets are stored in a near-exact, or even exact, copy of the source format and are in addition to the originating data stores.” In simple terms, a data lake is a storage system designed to accommodate all types of data, whether structured, semi-structured, or unstructured, from various sources. *Ludzie Nauki*'s data lake is a combination of Hadoop, Delta Lake, and Spark.

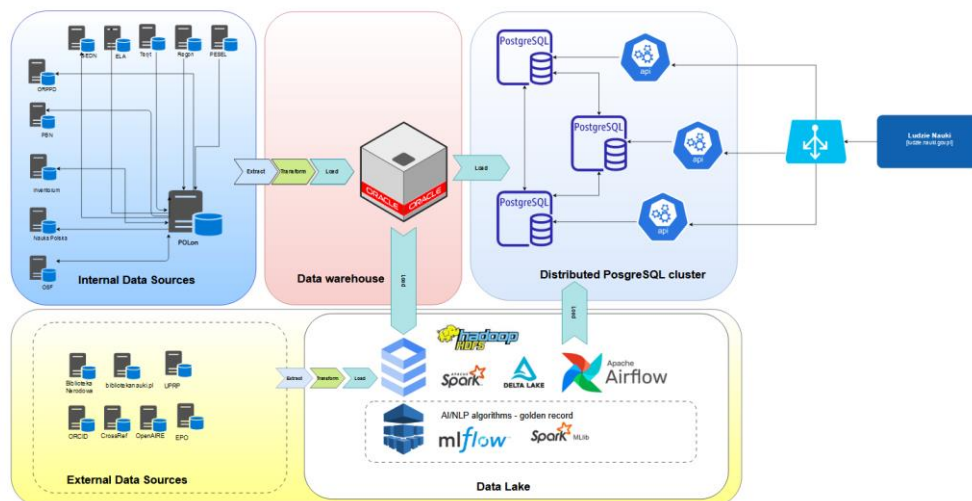


Figure 2. Data flow

Delta Lake is implemented as a storage format and as a set of access protocols for clients, making it simple to operate and highly accessible, as well as granting clients direct, high-bandwidth access to the object store. Delta Lake is used at thousands of organisations to processes exabytes of data per day, often replacing more complex architectures that involved multiple data management systems. It is available open source under an Apache 2 license at <https://delta.io>. (Armbrust et al., 2020).

Delta Lake ensures reliability in the data lakes. This is enhanced by its provision of atomicity, consistency, integrity, and durability (ACID) properties, scalable metadata management, and integration of streaming and batch data processing.

\*\* <https://www.gartner.com/en/information-technology/glossary/data-lake>

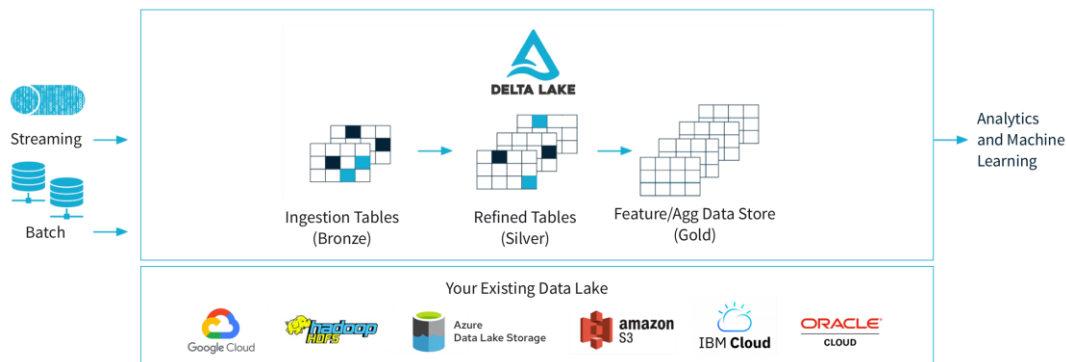


Figure 3. The Delta Lake zones architecture (<https://delta.io>)

The Delta Lake architecture is organised into three zones:

- Bronze: data is collected in its raw form from various sources, often containing inconsistencies and imperfections
- Silver: data is cleaned and refined through various functions, filters, and queries, gradually enhancing its quality
- Gold: data undergoes additional refinement and rigorous testing.

This is crucial, as end consumers—including machine learning algorithms and data analysis tools—have high standards of data quality and cannot accept data impurities.

## 6 Data challenges and solutions

### 6.1 Unified scientist identifier

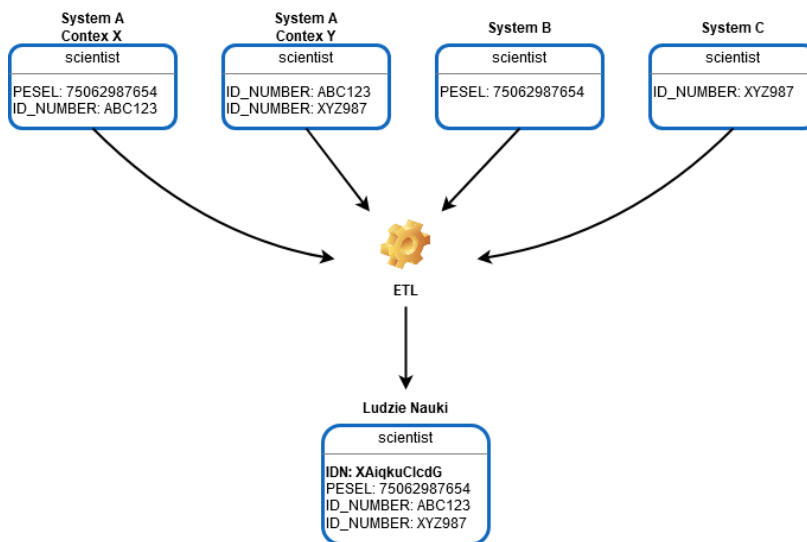


Figure 4. Identifiers unification process

One of the system's functions is its provision of unified scientist identifiers (IDNs), which enables the unambiguous identification of individuals in the context of any POL-on network system or related system. The architecture of the systems that constitute the data source for *Ludzie Nauki* is based on the concept of contextual data, i.e. each entity of the higher education and science system has its own data context, including a set of personal and identification data. Additionally, one scientist may appear in various components in the POL-on network systems. As part of advanced ETL processes maintained on the side of the data warehouse—which contains raw data from all domain systems and external sources—processes were developed and implemented to process data from individual sources and combine many logical objects that are considered a representation of a single physical object.

Due to the fact that the scientists' metadata set in system A might not enable the data to be combined with system B in a single operation, OPI PIB deployed a cross-merging algorithm to compare the metadata sets of individual entities. They were compared both at the levels of raw data and processed data.

To ensure the highest quality and consistency of *Ludzie Nauki*'s data, validation processes are performed on the final objects to verify the accuracy of the merge operation by comparing standardised features of individual scientists, such as year of birth, gender, or lemmatised patterns of names and surnames. In the global context, a set of attributes for each scientist is calculated, which comprises data from many contexts and systems selected by algorithms based on defined business rules.

## 6.2 CERIF implementation

CERIF serves as the standard for documenting and exchanging research information across the European Union. It facilitates the recording of metadata for research collections, and enables the representation of metadata that pertains to research units, their interconnections, and their scientific activities and outcomes in that domain. CERIF enhances the interpretability of data from IT systems developed and maintained across Europe, and facilitates the management of intricate relationships between various units in the research ecosystem, owing to its flexible data model.

*Ludzie Nauki*, acting as a central hub for information on Polish scientists, is tasked with supporting data sharing in a manner that is conducive to its interpretability both nationally and internationally. The decision to construct a data model based on CERIF, which encompasses application data structures and publicly accessible REST API interfaces, is a pivotal element in ensuring the consistency and efficiency of information exchange in the research domain. The individual systems of the POL-on network, which serves as the data source for the *Ludzie Nauki* platform, comply with legislative acts and regulations. Adapting the data collected in *Ludzie Nauki* to CERIF standard constituted a

```
{
  "cfResPublId": 38047368,
  "cfPublicationType": "book",
  "cfTitle": "A monograph of the Afrotropical
  Cassidinae (Coleoptera: Chrysomelidae). Part 4.
  Revision of the genus Chiridopsis Spaeth.",
  "cfPersAuthors": [
    {
      "cfPersFirstName": "Jolanta",
      "cfPersLastName": "Świętojańska",
      "cfOrgUnit": "Uniwersytet Wrocławski"
    },
    {
      "cfPersFirstName": "Lech",
      "cfPersLastName": "Borowiec",
      "cfOrgUnit": "Uniwersytet Wrocławski"
    }
  ],
  "cfResPublYear": "2017",
  "cfJournal": null,
  "cfPublisher": {
    "cfPublisherName": "Magnolia Press"
  },
  "cfISBN": "9781776702176",
  "cfISSN": "1175-5334",
  "cfTotalPages": "85",
  "cfSeries": "Zootaxa, Monograph",
  "cfDOI": "10.11646/zootaxa.4316.1.1",
  "cfURI": "https://doi.org/10.11646/zootaxa.4316.1.1",
  "cfLanguage": "en",
  "cfKeywords": [
    "Coleoptera",
    "monograph",
    "Chrysomelidae",
    "Cassidinae",
    "Aethiopocassis",
    "Afrotropical region"
  ]
}
```

Figure 5. Example of a REST API response for the core publication object



considerable challenge in the data transformation.

Mapping existing data to align with the structure and requirements defined by CERIF has facilitated the identification of areas that require changes or additions at the data model level. Modifications to the base data model have focused on three key areas:

- Data that conforms directly to CERIF standards
- Data that requires transformation to conform to CERIF standards
- Data not currently addressed by CERIF.

The implementation of data management practices and the conducting of an analysis of *Ludzie Nauki*'s scope and data compliance with CERIF enabled OPI PIB to design and implement a data storage system in the *Ludzie Nauki* platform that complies with the CERIF standard. The institute achieved this while maintaining the specificity of the data collected in the Polish ecosystem.

The CERIF standard has also been applied to the RESTful API services, which provide data on Polish scientists, as well as their scientific and artistic achievements.

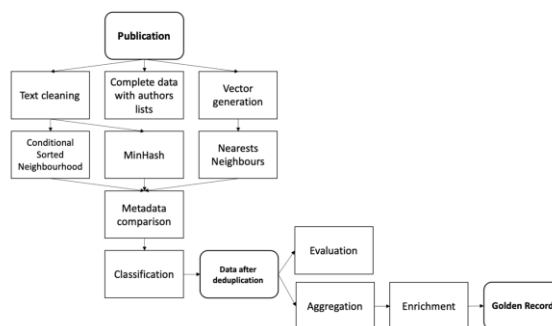
This enables the integration of external systems with the services provided on *Ludzie Nauki* while maintaining a consistent and interpretable data model.

### 6.3 Golden record

According to Data Clarity<sup>††</sup>, “The golden record is the ultimate prize in the data world. A fundamental concept within master data management defined as the single source of truth; one data point that captures all the necessary information we need to know about a member, a resource, or an item in our catalog—assumed to be 100% accurate.” *Ludzie Nauki* implements a set of algorithms to deliver reliable, coherent, and comprehensive datasets that are intended to serve as a reference for scientific achievements, such as publications, projects, or patents (outputs in CERIF nomenclature). The process involves deduplication and data enrichment using external data sources, as outlined in the following schematic for the publication object.

The basic preparation of a golden record encompasses six stages of data deduplication: initial text processing, candidate determination, comparison, classification, evaluation, and enrichment.

The process begins with text preprocessing, which aims to cleanse the data and prepare it for further analysis. This involves the removal of XML and HTML tags, and the conversion of unencoded characters (typically, language-specific ones) into the Basic Latin Unicode block—for example “é” into “e”. Next, the list of publication authors is supplemented by external sources. A character bigram vector is generated for the publication title, which enables the representation of publications in vector space. Duplicate candidates are subsequently identified. A conditional sorted neighbourhood algorithm identifies potential duplicates by sifting through the sort table and determining the neighbours of each record. Methods such as ClosestNeighbors (which, based on the character bigram vector from a publication's title, identify other publications with similar titles) and the MinHash algorithm (which is based on a publication's title) are used for this purpose. After determining



**Figure 6.** Golden record processing, transformation into CERIF standard

<sup>††</sup> <https://www.dataclarity.uk.com/2022/11/29/what-is-golden-record-management-and-why-it-is-important/>

the candidates, their metadata (including DOI, publication year, or author list) is compared to classify publication pairs as duplicates or nonduplicates. Evaluation is performed on a manually classified dataset, which enables verification of the operations' correctness. Then, based on clustered duplicates, the most complete record is determined using data from the source system. In the final stage, a publication's metadata is enriched based on data from external sources, such as Crossref.

## 7 Open science for everyone

Open science requires consideration of all users, regardless of their health problems. Aging populations are particularly vulnerable to exclusion due to their health condition. In 2021 in Poland, there were 5.4 million people with disabilities. 3.4 million people had a certificate of disability<sup>§§</sup>.

The level of the accessibility standard for the public sector is set by an EU directive of the European Parliament and of the Council (EU)<sup>§§</sup>. To prepare the *Ludzie Nauki* platform, the ETSI EN 301 549 V3.2.1:2021 standard specified in the directive is used. According to the standard, the *Ludzie Nauki* platform meets WCAG 2.1 at level AA<sup>\*\*\*</sup>.

The website has many improvements resulting from WCAG, e.g.:

- content compatible with screen readers. The tests included the NVDA screen reader, which is constantly being developed and is free for blind people
- appropriate contrast and improved focus visibility and support for content magnification. This is especially helpful for low vision people
- all functionalities are available to people using only a keyboard (physical and virtual, e.g. paralyzed people)
- care was taken to ensure that the instructions were not based on only one sense
- the website has code that works with assistive technologies, etc.

The project is on higher level than required by the standard, e.g.:

- the website did not have animations that could negatively affect people with vestibular disorders. This is especially important so as not to cause headaches with interaction-based animations
- there are no flashing elements to minimize the chances of triggering an on-screen epileptic fit
- the texts have been prepared to increase their understandability. Good practices from the "Prosty Język" project of the Polish Civil Service were used here<sup>†††</sup>.

Accessibility for people with disabilities is an important pillar in National Information Processing Institute software development. A WCAG specialist participates in all software life cycles. At the design and implementation stage of the code, it supports adaptation to the needs of people with disabilities. During implementation, it tests compliance with WCAG, and during the maintenance stage, it controls accessibility for people with disabilities.

---

<sup>§§</sup> <https://stat.gov.pl/en/national-census/national-population-and-housing-census-2021/final-results-of-the-national-population-and-housing-census-2021/population-ageing-in-poland-population-in-the-light-of-the-results-of-the-2021-census,2,1.html>

<sup>§§</sup> Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies.

<sup>\*\*\*</sup> <https://www.w3.org/TR/WCAG21/>

<sup>†††</sup> <https://www.gov.pl/web/sluzbacywilna/prosty-jezyk>

Updates and patches may have a negative impact on availability, so in this project the work of a WCAG specialist does not end with implementation. The project has the ability to report WCAG errors by phone/text/e-mail<sup>\*\*\*</sup>. Communication channels are published in the Accessibility Declaration, which has a link in the footer of the page. The project will additionally be covered by automatic tests supporting the detection of WCAG errors<sup>§§§</sup>.

Recent usability studies have confirmed the potential value of the newly developed portal for its users, highlighting its credibility in presenting data as a key advantage. Researchers, who are the primary audience for the portal, emphasized the importance of incorporating diverse data sets that are crucial in the scientific community. Moreover, they expressed that the portal could play a significant role in disseminating knowledge about Polish scientists, thereby enhancing their visibility in the global research landscape. Moreover, it is possible that it will help establish cooperation with science and business in the domestic and international arena.

## 8 Conclusion

The *Ludzie Nauki* initiative by the Polish Ministry of Science and Higher Education represents a significant leap in the consolidation and management of data on scientific research in Poland. By addressing the fragmentation of existing data sources and leveraging advanced data management techniques, *Ludzie Nauki* aims to serve as a comprehensive and reliable platform for showcasing the achievements of Polish researchers.

Through meticulous attention to data quality and adherence to international standards, such as the FAIR principles and open science initiatives, *Ludzie Nauki* aligns with emerging trends in data management across the European Union. By promoting transparency, accessibility, and interoperability, the platform not only facilitates collaboration in the scientific community, but also enhances its engagement with industry and policymakers.

The implementation of advanced data governance practices, technical solutions, and the pursuit of a golden record for scientific achievements underscore the commitment of *Ludzie Nauki* and its creators to excellence in data management. These efforts not only enhance the visibility and impact of Polish research, but also contribute to the broader objectives of promoting innovation and advancing knowledge.

As *Ludzie Nauki* continues to evolve, it serves as a beacon of best practices in research data management, demonstrating the transformative potential of comprehensive and well-executed data initiatives. *Ludzie Nauki* embodies the principles of openness, collaboration, and excellence, positioning Poland as a leader in research data management on the European research landscape.

## 9 References/Citations

Armbrust, M., Das, T., Sun, L., Yavuz, B., Zhu, S., Murthy, M., Torres, J., van Hovell, H., Ionescu, A., Łuszczak, A., Świtakowski, M., Szafranski, M., Li, X., Ueshin, T., Mokhtar, M., Boncz, P., Ghodsi, A., Paranjpye, S., Senster, P., ... Zaharia, M. (2020). Delta lake: High-performance ACID table storage over cloud object stores. *Proceedings of the VLDB Endowment*, 13(12), 3411–3424. <https://doi.org/10.14778/3415478.3415560>.

---

<sup>\*\*\*</sup> <https://ludzie.nauka.gov.pl/wp/deklaracja-dostepnosci-serwisu/>

<sup>§§§</sup> <https://playwright.dev/docs/accessibility-testing>

Bylina, M., Podwysocki, E., & Michajłowicz, M. (2023). *Data Governance at the National Information Processing Institute in Poland*. Paper presented at the EPIc Series in Computing, 95, 79–88. <https://doi.org/10.29007/1rp9>.

Jörg, B. (2010). *CERIF: The Common European Research Information Format Model*. *Data Science Journal*, 9, CRIS24–CRIS31. <https://doi.org/10.2481/dsj.CRIS4>.

Korytkowski, P. (2023). *Ludzie Nauki ('People of Science'): A new CRIS for Poland | EuroCRIS*. Home | EuroCRIS. Retrieved February 9, 2024, from: <https://dspacecris.eurocris.org/handle/11366/2505>.

Michajłowicz, M., Idzi, A., Kierzkowski, J., Kucharska, I., Paszkowska, M., Podwysocki, E., Rodzik, P., Sadłowski, A., Sobkowicz, A. (2021). *Systemy informatyczne wspierające naukę i szkolnictwo wyższe : POL-on : System Informacji o Nauce i Szkolnictwie Wyższym : TOM II*. Retrieved February 2, 2024, from: <https://opi.org.pl/wp-content/uploads/2021/10/POL-on.TomII-.pdf>.

Michajłowicz, M., Kozłowski, M., Kowalski, M., Fijałkowski, S., Drogosz, J., Bylina, M., & Furmankowska-Podnieśńska, A. (2022). *Scientific activity evaluation in Poland: the IT ecosystem and the optimal selection of achievements*. Paper presented at the EPIc Series in Computing, , 86 56–65 from: <https://easychair.org/publications/paper/Rf1d>.

Michajłowicz, M., Niemczyk, M., Protasiewicz, J., & Mroczkowska, K. (2018). *POL-on: The Information System of Science and Higher Education in Poland*. *EUNIS 2018 Congress Book of Proceedings*. Retrieved February 6, 2023, from: [https://www.eunis.org/download/2018/EUNIS\\_2018\\_paper\\_70.pdf](https://www.eunis.org/download/2018/EUNIS_2018_paper_70.pdf).

Nikkanen, J., & Puuska, H.-M. (2022). Researchers' profiles in the Finnish Research Information Hub. *Procedia Computer Science*, 211, 206–210. <https://doi.org/10.1016/j.procs.2022.10.193>.

Palavesm, K., & Joorel, J. S. (2022). IRINS: Implementing a Research Information Management System in Indian Higher Education Institutions. *Procedia Computer Science*, 211, 238–245. <https://doi.org/10.1016/j.procs.2022.10.197>.

Podwysocki, E., Błaszczuk, Ł., Niemczyk, M., Protasiewicz, J., Michajłowicz, M., Rosiak, S., Kucharska, I. (2019). Distributed services and a warehouse as an ecosystem on science and higher education. *EUNIS 2019 Congress Book of Proceedings*. Retrieved February 10, 2024, from: [https://www.eunis.org/download/2019/EUNIS\\_proceedings\\_2019.pdf](https://www.eunis.org/download/2019/EUNIS_proceedings_2019.pdf).

Protasiewicz, J., Rosiak, S., Kucharska, I., Podwysocki, E., Niemczyk, M., Błaszczuk, Ł., & Michajłowicz, M. (2019). *RAD-on: An integrated System of Services for Science - Online Elections for the Council of Scientific Excellence in Poland*. *EUNIS 2019 Congress Book of Proceedings*. Retrieved February 4, 2024, from: [https://www.eunis.org/download/2019/EUNIS\\_proceedings\\_2019.pdf](https://www.eunis.org/download/2019/EUNIS_proceedings_2019.pdf).

Rozkosz, E. (2014). *Serwisy społecznościowe dla naukowców (SSN) na przykładzie Researchgate i Academia.Edu*. *Forum Bibliotek Medycznych*, 7(2), 16–24. Retrieved February 4, 2024, from: <https://fbc.pionier.net.pl/details/nmm91ml>.

Zaharia, M.A., Ghodsi, A., Xin, R., & Armbrust, M. (2021). Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. *Conference on Innovative Data Systems Research*. Retrieved February 9, 2024, from: [http://cidrdb.org/cidr2021/papers/cidr2021\\_paper17.pdf](http://cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf).

## 10 Author biographies



**Emil Podwysocki (MSc.)** was awarded a master's degree in telecommunications systems at the Technical University of Lodz in 2009. He has over ten years' professional experience related to ETL/ELT, data warehousing, and business intelligence. His areas of interest include Oracle technology, big data, business intelligence, low-code/no-code technology, and data visualisation. He is a certified Business Intelligence and APEX specialist. He currently serves as the head of the Laboratory of Databases and Business Analytical Systems at the National Information Processing Institute.

Email: [emil.podwysocki@opi.org.pl](mailto:emil.podwysocki@opi.org.pl)

LinkedIn: [www.linkedin.com/in/emil-podwysocki](https://www.linkedin.com/in/emil-podwysocki)



**Maria Bylina (MSc.)** is a software designer and business analyst at the National Information Processing Institute. She was awarded a master's degree at the Faculty of Geodesy and Cartography at Warsaw University of Technology in 2018. She also has several years' experience as a GIS analyst and service maker. Her areas of interest include agile project management, data governance, ETL, and software design. She holds Professional Scrum Product Owner, Professional Scrum with Kanban, and UML Professional 2 Foundation Level certificates.

Email: [maria.bylina@opi.org.pl](mailto:maria.bylina@opi.org.pl)

LinkedIn: [www.linkedin.com/in/mariabylina](https://www.linkedin.com/in/mariabylina)



**Jacek Raczko (MSc.)** is deputy head of the Laboratory of Databases and Business Analytics Systems at the National Information Processing Institute. He has worked with data for more than ten years in areas as diverse as data integration and transformation, the construction of data warehouses, business intelligence, and management reporting. He is a certified Oracle solutions specialist.

Email: [jacek.raczko@opi.org.pl](mailto:jacek.raczko@opi.org.pl)



**Michał Ulaniuk (MSc.)** was awarded a master's degree in computer science (specialising in artificial intelligence) at Adam Mickiewicz University in Poznań in 2023. He is currently pursuing his PhD at the Systems Research Institute of the Polish Academy of Sciences. His professional focus encompasses computer vision, natural language processing, and document understanding. He currently works on the data science team at the National Information Processing Institute.

Email: [michal.ulaniuk@opi.org.pl](mailto:michal.ulaniuk@opi.org.pl)

LinkedIn: <https://www.linkedin.com/in/ulaniuk>



**Marek Michajłowicz (MSc.)** is the deputy head of the National Information Processing Institute. He holds a master's degree in sociology from Cardinal Stefan Wyszyński University in Warsaw, a Bachelor of Engineering degree in computer science from Warsaw School of Information Technology under the auspices of the Polish Academy of Sciences, and an MBA from Warsaw University of Technology. He has many years' professional experience in business and systems analysis. His areas of interest include agile project management, software design and development, big data, and warehousing.

Email: [marek.michajlowicz@opi.org.pl](mailto:marek.michajlowicz@opi.org.pl)



**Rafał Jendrzejewski (MSc.)** has been dealing with accessibility for people with disabilities since 2011. He worked on projects of: the Ministry of Funds and Regional Policy, the Ministry of Digitization, the Ministry of National Education, the Ministry of Science and Higher Education. He currently works on the Application Design and Implementation Team at the National Information Processing Institute.

E-mail: [rafal.jendrzejewski@opi.org.pl](mailto:rafal.jendrzejewski@opi.org.pl)