



EPIc Series in Computing

Volume 107, 2025, Pages 247–255

Proceedings of EUNIS 2025 annual congress in Belfast



The Impact of Data Quality and the Implementation of Data Governance on the Usability of the *Ludzie Nauki* CRIS System: Lessons learned

Maria Bylina^{1*}, Emil Podwysocki¹, and Marek Michajłowicz^{1†}

¹The National Information Processing Institute, Poland

maria.bylina@opi.org.pl, emil.podwysocki@opi.org.pl,

marek.michalowicz@opi.org.pl

Abstract

This article discusses the implementation and challenges of the Ludzie Nauki system, which promotes Polish science and manages research data. It highlights the legal and technical complexities faced during the system's development, particularly in compliance with Poland's Higher Education and Science Act and the GDPR. The system's integration with various data sources and its role in ensuring the accuracy and accessibility of scientific data are emphasized. Key issues, such as user dissatisfaction with the inability to edit published information, are also addressed alongside planned improvements to enhance usability, including AI-driven semantic search and personalized researcher profiles. Moreover, this article explores the significance of data governance in research information systems, discussing its role in improving data quality, integration, and the system's overall effectiveness. The experience we have gained from implementing the system provides insights into project management methodologies, including the use of the SCRUM, KANBAN, and waterfall approaches. This article concludes by emphasizing the need for continuous development and the provision of services based on the data collected in Ludzie Nauki, to maximize the potential of the platform and to foster innovation in the scientific community.

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

† <https://orcid.org/0000-0003-2096-5005>

1 Introduction

Ludzie Nauki is the newest project of the Polish Ministry of Science and Higher Education. It aims to promote Polish scientists and their achievements. The system is the largest, best-organized, and most reliable data collection of Polish scientists. It is the main source of 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 (Podwysocki et al. 2024).

The *Ludzie Nauki* initiative marks a major advancement in the integration and management of scientific research data in Poland. By tackling the fragmentation of existing data sources and utilizing advanced data management strategies, *Ludzie Nauki* seeks to become a comprehensive and trustworthy platform for highlighting the accomplishments of Polish researchers.

Did the project manage to meet its initial goals and expectations? How did the implementation of the data governance framework contribute to the successful realization implementation of the project, and is it worth continuing to develop this process further? How was the work organized within the project, and are there areas wherein which areas could improvements could have been made? Did the application of the “golden record” approach yield the expected results? Is the portal truly user-friendly and beneficial for its intended audience? These questions invite further exploration of the project's outcomes and the lessons learned in the sections below.

2 Post-mortem analysis-based conclusions

The implementation of the *Ludzie Nauki* system required compliance with a number of legal requirements, particularly in the context of Poland's Higher Education and Science Act (*Ustawa z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce*) and regulations that relate to personal data protection (namely the GDPR). The system had to ensure compliance with the legal provisions regarding the sharing of scientific data. The Higher Education and Science Act regulates the publication of data on researchers, their achievements, their publications, and their research projects. A considerable challenge arose from the interpretations of regulations issued by various departments of the Polish Ministry of Science and Higher Education regarding the exclusion of certain data from publication.

Ludzie Nauki also needed to adapt to the GDPR regulations, particularly those that concern the processing of the personal data of researchers. Some data, including personal information on researchers, was based on information from the source system, POL-on. Suitable access control was necessary to ensure that users could manage their data in line with privacy protection requirements.

The concept of the system was to prevent users from editing data that originates from reference sources to ensure the reliability of the information. Upon logging in, users have access to their profiles and can decide which data will be available to the public, with the exception of data that must be publicly accessible according to legal regulations. To correct data sourced from the source systems, users must contact the institutions responsible for registering that data.

Despite usability studies being conducted, *Ludzie Nauki* did not always meet the expectations of different user groups, who expressed dissatisfaction with the system's inability to update the published data immediately.

Integration with Poland's national identification number (PESEL) database could help automatically mark the status of researchers, such as indicating that a researcher is deceased, which would enable more precise data management. Such integration with central databases should have been considered during the design phase, although it is also relates closely to current legislation.

3 How can the implementation of data governance provide opportunities for a the CRIS system?

Effective data governance plays a crucial role in improving the quality, accessibility, and security of research information systems. As OPI PIB collects and manages vast amounts of data from Polish universities and research institutions, the need for standardized policies, clear ownership, and integrated metadata management has become abundantly clear.

During the implementation of the data governance framework, we gained valuable insights into the business areas that relate to the collected data. This process enabled us to inventory subareas and data entities, providing a comprehensive understanding of the structure and relationships in the dataset (Bylina et al. 2023).

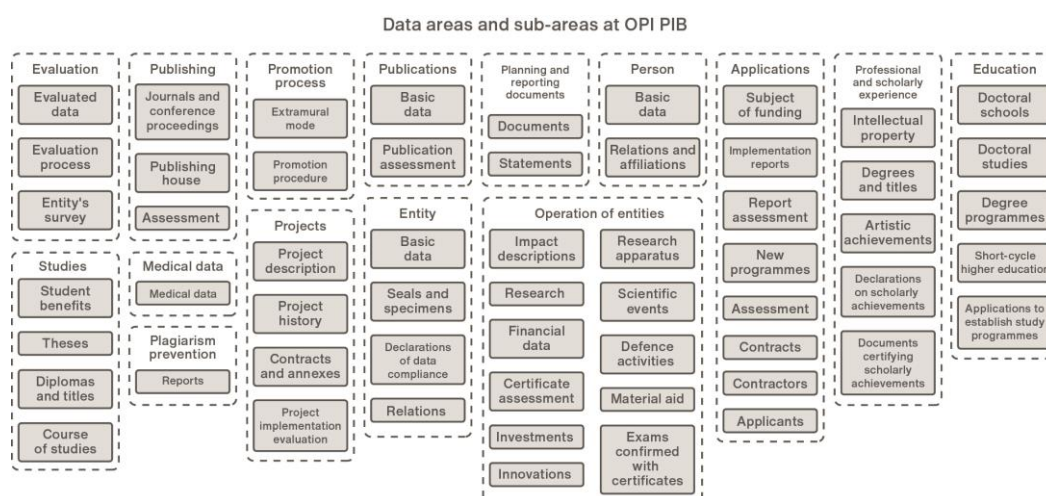


Figure 1: Business areas and sub-areas at OPI PIB

OPI PIB is currently in the process of implementing a data catalog and integrating it with the a business glossary to establish seamless connections between business terminology and physical data. While this phase presents its own challenges, it is expected to deliver considerable benefits, including improved metadata management, enhanced searchability, and greater overall efficiency in data governance. The data catalog will serve as a centralized tool that facilitates easy access to data descriptions, enabling users to quickly identify available datasets, as well as understanding their context and security rules. Additionally, it will simplify integration with external systems, as data owners can determine existing data assets and map them to external sources efficiently. Integration with external systems enable the easier sharing of research outcomes, the support of international collaboration, and the facilitation of the monitoring and reporting of scientific achievements.

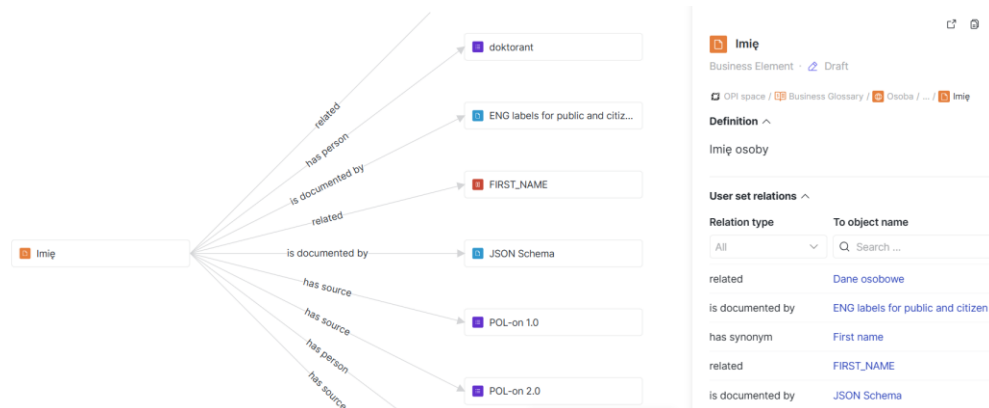


Figure 2 Data catalog - Business Term relations

Moreover, by developing data quality indicators, OPI PIB intends to establish a foundation for the continuous monitoring and improvement of data accuracy and consistency. This proactive approach ensures that research data remains reliable, which supports evidence-based decision-making.

It is worth emphasizing that the launch of the *Ludzie Nauki* project was accompanied by the start of the implementation of a Data Governance framework. In retrospect, it can be said that earlier identification of data resources and all their potential sources in reference data could have improved the process of improving the quality of source data, facilitating data integration and increasing the efficiency of data use. This confirms that it is essential to continuously take care of data in the organization.

4 Project management

The development of the Polish national CRIS system, *Ludzie Nauki*, posed a unique challenge from the perspective of OPI's experience in implementing IT projects. Previous projects executed at the institute were organized primarily on the basis of the SCRUM methodology, under which the key stages of the development process included analysis, UX design, programming, testing, and the iterative deployment of individual functionalities based on CI/CD pipelines. The implementation of systems such as the Integrated Science and Higher Education System (POL-on) or the Polish Scientific Bibliography (PBN) was founded on decomposing system components into autonomous microservices (Michajłowicz et al. 2018). These microservices could be independently developed by SCRUM teams following methodological principles, such as a dedicated product backlog, self-sufficient development teams consisting of 5-9 five to nine members, and two-week iterations beginning with planning and concluding with a demonstration of the functioning product.

However, this project introduced new challenges, including the pivotal role of a separate team responsible for data integration (data warehouse developers and data analysts) and an architecture that did not allow for system decomposition into smaller components (products). Due to these factors—architectural complexity and an expanded project team working jointly on a single product—a decision was made to deviate from the SCRUM approach in the initial phase. Instead, the project was managed using a KANBAN approach at the product development level combined with traditional waterfall model for strategic project management. The continuous product development method based on the principles of the LEAN KANBAN methodology enabled better identification of bottlenecks at various stages of the development process and facilitated the expansion of the development teams.

The experience OPI PIB gained from integrating development teams that work on products built with technologies characteristic of transactional systems—such as a single-page application frontends, microservices, containerized applications, and persistence based on JSON file storage—will be analyzed further and adapted to the realities of implementing a project architecture that relies strictly on data integration processes (ETL) and centralized storage, such as a data lake.

One possible area for improvement in the project management of the *Ludzie Nauki* system is a more strategic alignment between data integration efforts and software development practices from the outset. The early identification of architectural constraints, particularly the lack of system decomposition into smaller components, might have enabled a hybrid approach that combined the flexibility of SCRUM for certain aspects of development with a more structured approach for data integration. Establishing clearer coordination mechanisms between the software development teams and the data integration team could have mitigated some of the initial inefficiencies caused by working with separate methodologies.

Additionally, adopting a phased approach to integrating the data warehouse team into the development cycle, rather than operating in parallel with different methodologies, could have improved communication and synchronization between the teams. The introduction of cross-functional teams, comprising both software developers and data analysts, might have facilitated a more agile response to emerging technical challenges, reducing dependencies and bottlenecks in the workflow.

Finally, while the use of KANBAN helped in identifying workflow bottlenecks—introducing more structured sprint reviews and retrospectives, similar to SCRUM practices—it might also have fostered a more iterative feedback loop, leading to incremental improvements in both the product and the management process. Such refinements might have contributed to greater efficiency, improved adaptability, and a more seamless integration of software and data components in such large and complex projects in the future.

5 “Golden records” and data quality

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, which is outlined in the following schematic for the publication object. The central goal was to create single source of truth for the scientific data contained in combined national data sources, such as POL-on[‡], PBN[§], or BWNP^{**} and international ones, such as Crossref^{††}, OpenAIRE^{‡‡}, or Open ALEX^{§§} (Podwysocki et al. 2024).

Ensuring of high data quality in open data repositories remains a critical challenge in the scientific community. As data-driven research becomes increasingly prevalent, the integrity, consistency, and reliability of datasets are paramount. The initial iteration of the “golden record” algorithm was not without limitations; however, its implementation yielded valuable insights into data quality issues,

[‡] 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

^{††} 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

inconsistencies, and conflicts. Specifically, it facilitated the identification of anomalies and discrepancies that could compromise the validity of research findings.

One of the most complex aspects of this process was the creation of the data source hierarchy needed to successfully develop a hierarchical framework for data sources, which proved essential for the effective resolution of data conflicts. The establishment of such a hierarchy required a systematic approach to the ranking of sources based on credibility, accuracy, and domain relevance. Future research should focus on refining these methodologies to enhance automated conflict resolution mechanisms and to ensure the robustness of data governance frameworks in open science ecosystems.

The scientific community in Poland was inadequately prepared for this approach to data management. The concept of the “golden record” was poorly understood, which led to considerable misconceptions. Researchers emphasized the need for access to raw data without enrichments or modifications. The high volume of support tickets submitted to OPI PIB’s helpdesk highlighted the scale of the issue. This experience demonstrated the necessity of launching an educational campaign before introducing the system.

Collecting, integrating, storing, and analyzing data in a database system is a well-established concept. The introduction of the CRIS means that scientific institutions must provide the required information on their research activities and research results at a high quality. A one-time cleanup is insufficient; the data must be curated and maintained continuously. Some data errors (such as missing values, spelling errors, inaccurate data, incorrect formatting, or inconsistencies) can be traced across different data sources and are difficult to find. Minor mistakes can render the data unusable, and corrupted data can have serious consequences. The sooner quality issues are identified and remedied, the better (Azeuorual O et al. 2019).

6 User impressions and the perception of the project in the scientific world

Usability tests of the *Ludzie Nauki* showed that its overall aesthetics and transparency were rated positively, and key features such as the search engine for scientists has great potential, although they need refinement.

The system is monitored continuously to ensure its performance and effectiveness. Usage logs, particularly from the search engine and filtering tools, indicate that users opt predominantly for the main search feature. This insight has led to the idea of enhancing the search functionality by the incorporation of AI algorithms and the implementation of a semantic search engine, which would improve the precision and relevance of search results, further increasing the system’s usability.

Users expect visual consistency and intuitive navigation mechanisms, especially for profiles, which should contain more complete and up-to-date information. Criticism has been leveled at the illegible labeling of scientific fields, the problematic error reporting system and mismatched graphic elements. Users frequently submit requests for corrections to their data; this highlights a significant issue. The responsibility for these updates, however, falls on the higher education institution (HEI) where the researcher is currently employed. This structure emphasizes the importance of institutional accountability in maintaining the accuracy of data within the system, which underlines the need for clear communication and coordination between the platform and educational institutions HEIs.

One of the main concerns expressed by users is the inability to edit the published data. The initial concept of the system was based on the publishing of data sourced from reference systems. After multiple consultations and feedback from users, OPI PIB introduced a feature that enables users to create a personalized version of their researcher profiles that are displayed on the portal. Although users remain limited to selecting from reference data, they now have the option to hide unwanted information, such as particular publications, which grants them greater control over their public profiles.

Ludzie Nauki should ensure full transparency in reviewer selection processes— whether for theses, promotion proceedings, or grant funding. The system can also be used for evaluation, grant calculation and all kinds of election processes in the scientific community. Through *Ludzie Nauki*, scientists could also receive notifications of new competitions and opportunities to obtain research funding.

OPI PIB plans to introduce additional services in *Ludzie Nauki* for users from both the academic and business sectors. This expansion aims not only to promote the achievements of Polish researchers but also to foster networking opportunities and provide avenues for professional growth. By extending the system's functionalities, it is hoped that *Ludzie Nauki* will become a valuable tool for connecting academia with the business world; one that enables, both parties to collaborate and innovate more effectively. It is also necessary to raise awareness of the portal through more effective promotion and clear communication of its functions.

The ultimate goal of research information systems is “to provide the best evidence base on which to support high-quality decision-making” (Haak et al. 2012).

7 Conclusion

Despite numerous challenges and issues encountered along the way, the implementation of *Ludzie Nauki* has been successful and is generating considerable interest in the scientific community. Ensuring data quality in the source systems and effectively integrating with other data sources are crucial. Cooperation between OPI PIB and HEIs plays a key role in maintaining high data quality on the platform.

Communication with users must be improved, effective helpdesk support must be provided, and the needs of end users must be met while maintaining transparency and usability in the system's UX design. To keep the portal dynamic, it should be promoted widely, particularly in the business sector, and it should be developed alongside e-services, such as an expert recommendation system.

The implementation of *Ludzie Nauki* faced several challenges that relate to legal regulations, data protection, and user needs. The system's adaptation to everchanging legal frameworks is essential, as is a deep understanding of the diverse requirements of various user groups to ensure the system's effectiveness and full functionality. Moreover, the introduction of mechanisms that facilitate data management, with consideration for both legal obligations and user comfort, is crucial for its continued success.

The idea of a “golden record” for scientific achievements underscores *Ludzie Nauki's* and its developers' dedication to maintaining high standards in data management. These efforts enhance the visibility and influence of Polish research while supporting broader objectives, such as the fostering of innovation and the advancement of knowledge.

As *Ludzie Nauki* continues to develop, it stands as a prime example of best practices in research data management; one that showcases the transformative power of well-structured and thoughtfully implemented data initiatives. The system reflects the values of openness, collaboration, and excellence. It establishes Poland as a frontrunner in research data management and the implementation of data governance at the central level across the European research community.

8 References/Citations

Azeroual, O., & Schöpfel, J. (2019). Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries. *Publications*, 7(1), 14. <https://doi.org/10.3390/publications7010014>.

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, 7988. <https://doi.org/10.29007/1rp9>.

Haak, L. L., Baker, D., & Probus, M. A. (2012). *Creating a data infrastructure for tracking knowledge flow*. 113–118. Scopus. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84898906486&partnerID=40&md5=381e8d87c25714d96f34cb075e2ca3e3>.

Jim, C. K., & Chang, H.-C. (2018). *The current state of data governance in higher education*. *Proceedings of the Association for Information Science and Technology*, 55(1), 198–206. <https://doi.org/10.1002/pr2.2018.14505501022>.

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 17, 2023, from: <https://opi.org.pl/wp-content/uploads/2021/10/POL-on.TomII-.pdf>.

Michajłowicz, M., Kozłowski, M., Kowalski, M., Fijałkowski, S., Drogoz, J., Bylina, M., & Furmankowska-Podniesińska, 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.

Podwysocki, E., Bylina, M., Raczko, J., Ulaniuk, M., Michajłowicz, M., & Jendrzewski, R. (2024). *Ludzie Nauki: Data aspects of the Polish national CRIS*. 94–79. <https://doi.org/10.29007/23qb>.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>.

9 Author biographies



Maria Bylina (MSc) is a software designer, business analyst and Data Governance Lead at the National Information Processing Institute. She obtained a master's degree at the Faculty of Geodesy and Cartography at Warsaw University of Technology. She has also 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

LinkedIn: www.linkedin.com/in/mariabylina



Emil Podwysocki (MSc) holds a Master's degree in Telecommunications Systems on Lodz University of Technology. He also completed postgraduate MBA studies for the IT industry at the Polish-Japanese Academy of Information Technology. He has been affiliated with the National Information Processing Institute – National Research Institute since 2017. For nearly five years, he led the Laboratory of Databases and Business Analytics Systems, and later headed the Data Warehouse Department. He has over a dozen years of professional experience and holds numerous certifications in the areas of enterprise

architecture, business intelligence systems, and software development using low-code technologies. His professional interests include data management, cybersecurity, DevOps, and the practical application of low-code technologies in optimizing business processes. He also serves as the Plenipotentiary of the Minister of Science and Higher Education for Open Data.

Email: emil.podwysocki@opi.org.pl

LinkedIn: www.linkedin.com/in/emil-podwysocki



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 (B.E.) 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

LinkedIn: <https://www.linkedin.com/in/marek-michajlowicz/>