NATIONAL INFORMATION PROCESSING INSTITUTE

POL-on

A central information system for science and higher education in Poland

Scientific editor
Marek Michajłowicz

Information technology systems that support science and higher education

POL-on: A central information system for science and higher education in Poland

Scientific editor:
Marek Michajłowicz

NATIONAL INFORMATION PROCESSING INSTITUTE

Information technology systems that support science and higher education
POL-on: A central information system for science and higher education in Poland

Scientific editor:
Marek Michajłowicz

Authors:
Dr Jacek Bieliński: Chapter 1
Dr Jakub Kierzkowski: Chapter 3
Maciej Kowalski: Chapter 1
Iwona Kucharska: Chapters 1 and 2
Dr Anna Łobodzińska: Chapter 2
Marek Michajłowicz: Chapters 1, 2 and 3
Marek Nurzyński: Chapter 1
Dr Piotr Rodzik: Chapters 1 and 2
Szymon Roziewski: Chapter 3
Sławomir Rybka: Chapter 2

Copyeditor:
Kevin McRobb

Reviewer:
Dr Zenon A. Sosnowski, Associate Professor at the Bialystok University of Technology

Published by:
National Information Processing Institute
al. Niepodległości 188b
00-608 Warsaw
email: opi@opi.org.pl
www.opi.org.pl

ISBN 978-83-63060-29-9

Cover and graphic design:
Anna Kozłowska, Monika Klimowska

Typesetting in LaTeX:
Dr Jakub Kierzkowski

Printed by:
Drukarnia DSS Szczepan Szymański
ul. Wolska 108, 05-119 Wola Aleksandra

CONTENTS

INTRODUCTION

The effective management of information is a critical component of modern science and higher education systems. As the volume and complexity of data continue to grow, national-level information systems have become essential tools for ensuring transparency, supporting evidence-based policymaking, and enabling efficient administration. In this context, the Polish Integrated System of Information on Science and Higher Education (POL-on) was developed to address the specific needs of Poland's academic and research landscape. POL-on is a comprehensive, legally mandated platform that consolidates data from all higher education and research institutions in Poland. Its architecture and operational model are designed to ensure consistency, interoperability, and alignment with national regulations. Unlike many comparable systems, POL-on integrates a wide range of data domains—spanning students, academic staff, research outputs, institutional infrastructure, and funding—into a unified framework. This level of integration supports a high degree of standardisation and central oversight, while also enabling institutions to manage their own data effectively. This publication provides a detailed examination of the POL-on system, focusing on its design principles, implementation strategies, and practical applications. It situates POL-on within the broader European and global context of research information systems, offering comparative insights and highlighting distinctive features of the Polish approach. Particular attention is given to the technical aspects of system architecture, data governance, and the transition from a monolithic to a microservice-based infrastructure. The aim of this work is to contribute to the ongoing discourse on the development and optimisation of national information systems in science and higher education. By documenting the experiences and lessons learned from the implementation of POL-on, the publication seeks to inform future initiatives in this domain and support the exchange of best practices among stakeholders involved in academic digital transformation.

CHAPTER 1

INFORMATION SYSTEMS IN SCIENCE AND HIGHER EDUCATION

Dr Jacek Bieliński
Maciej Kowalski
Iwona Kucharska
Marek Michajłowicz
Marek Nurzyński
Dr Piotr Rodzik

In the twenty-first century, information systems have become critical infrastructure in nearly every sector. In industries as diverse as finance, logistics, and healthcare, integrated IT systems ensure real-time decision-making, regulatory compliance, performance monitoring, and long-term strategic planning. Their effectiveness lies in their capacity to consolidate complex data into actionable insights, to eliminate manual inefficiencies, and to enforce standardised practices across distributed organisations.

Universities and scientific institutions, although typically regarded to be more autonomous and diverse than corporations, are increasingly subject to the same pressures for digital transformation. They must respond to growing public expectations for transparency, evidence-based funding, international competitiveness, and operational efficiency. As in other sectors, information systems in higher education and science now play a pivotal role in the management of critical processes—ranging from student recruitment, research output tracking, and quality assurance, to personnel management and resource allocation.

Institutionally, this transformation has been marked by the widespread adoption of student information systems (SISs), current research information systems (CRISs), and enterprise-level platforms that integrate human resources, finance, and academic operations. However, **national information systems**—which serve as centralised platforms for *coordinating, supervising, and shaping* the wider science and higher education ecosystem within a country—are equally important.

Such systems typically serve three overarching purposes: *regulatory and administrative oversight*, *strategic planning and funding allocation*, and *scientific evaluation*. In their regulatory role, they ensure compliance with legal and policy requirements by collecting standardised and verified data from all academic institutions. For strategic purposes, they enable policymakers to analyse trends, design evidence-based interventions, and to allocate resources efficiently. Finally, in the context of scientific evaluation, they provide the foundational data infrastructure required to assess research performance nationally—supporting transparent, consistent, and data-driven review processes.

This broad scope of utilisation makes such systems not only tools of information management, but also key instruments in steering the development of science and higher education in alignment with national priorities and international standards.

These national systems—such as POL-on in Poland, CRIStin in Norway, FRIS in Flanders, and ETIS in Estonia—have been developed to address several strategic objectives:

- Ensure consistent and legally compliant data collection from all accredited institutions
- Support public governance, including policy evaluation, strategic funding, and institutional oversight
- Enhance transparency and trust in research and education by offering verified public datasets
- Facilitate international benchmarking and interoperability, often through compliance with standards like CERIF
- Streamline administrative processes, reducing redundancies and improving quality control at a national scale.

These platforms consolidate data on students, academic staff, research projects, publications, funding flows, and institutional performance. Their design reflects the need to balance institutional autonomy with public accountability.

This chapter examines the landscape of information systems in science and higher education, paying particular attention to those developed and operated at a national level. The chapter opens with a discussion on international trends and reference systems used in research policy and university management. It then explores the rationale behind centralised systems, their typical architecture, and the societal challenges they aim to address—setting the stage for a detailed analysis of the POL-on system in subsequent chapters.

1.1. The diversity of national information systems

No single model can be used to build national-level systems for the tracking of scientific and educational achievements. In practice, not all national bodies possess—or wish to possess—such databases. Scientific institutions typically enjoy high degrees of autonomy regarding how they conduct research and teaching. This often makes deep centralisation and unification difficult or even impossible. Nevertheless, central coordination is necessary for coordinated activities to be conducted at a higher level. What truly differentiates models is their scale and form of implementation.

In simplified terms, we can distinguish models based on distributed network solutions (e.g. CRIS platforms, repositories), where repositories provide infrastructure and communication interfaces for collaborating research institutions. Examples include the DABAR digital archives and academic repositories in Croatia [11] and library management systems that use the OAI-PMH protocol: the standard developed for harvesting repository metadata [51].

Norway utilises a model similar to Poland's, under which the country's Ministry of Science delegates the development of IT solutions to the University Centre for Information Technology in Oslo, backed by appropriate infrastructure and know-how. This has resulted in the creation of a central, nationally deployed CRIS repository called CRISTIN[1].

Another type of architecture is built on cloud infrastructure, in which collaborating institutions integrate their infrastructure with a shared software platform. Examples of this approach include Sciebo, the campus cloud for North Rhine-Westphalia in Germany [76] and the Higher Education Institutions Closer interoperability platform in Portugal [54].

At the other end of the spectrum are centralised systems, such as POL-on [39] or the Norwegian Common Student System[2].

We must recognise that stakeholders of such systems include not only decision-makers, but also research managers, administrators, research councils, technology transfer organisations, the media, and the public. All require easy and open access to comprehensive scientific information, which can be aggregated in single, central systems [4]. Such systems can also address the increased reporting demands on researchers, who are often required to submit the same data multiple times.

When conducting a comparative analysis, the dimensions under investigation must be defined clearly. One such dimension is the reporting and evaluation of scientific achievements. The corresponding systems' structure is shaped primarily by the type of evaluation employed. In systems that use parametric (quantitative) evaluation, a crucial element is data structuring and quality. These requirements are considerably higher than in systems based on expert assessment, where the dominant criterion is qualitative analysis.

In the context of central systems that collect data on students and educational processes, the key consideration is whether such a system is necessary—for a given country—in the implementation of coordinated, fact-based policy. Data can be collected in various ways; most often it is sufficient to gather reports, aggregated data, and statements directly from universities. A country's population size and the rate of enrolment

[1] cristin.no (accessed 17 November 2025)
[2] fellesstudentsystem.no (accessed 17 November 2025)

(‘scholarisation index’)[3] also play key roles. In countries with relatively small populations and few universities, the need for the establishment of dedicated central registries appears to be lower.

1.2. CRIS-type systems

The growing amount of data collected by academic and higher education institutions demands the use of IT systems to manage information, as well as generating knowledge and sharing it with stakeholders in local, national, and international academic systems. In the twenty-first century, a growing need to use large datasets to generate knowledge has emerged. For the first time in history, advancing digitalisation enables us to transform descriptions of reality (i.e. data) into multicontextual information, and information into valuable knowledge. It has become clear that the sheer volume of data increases its potential to be transformed into reliable knowledge, as it can mask aberrations and present an accurate picture of the circumstances in science and higher education systems.

The digitalisation of academia increases the need for transparent, evaluable, and open research data, in addition to metadata on conducted research, teaching, and science funding. In such an environment, CRISs have become a key tool in the management of academic activity. Such systems enable the collection, processing, analysis, and sharing of contextualised metadata on processes and the results of academic research and teaching. CRISs support operational as well as strategic monitoring, decision-making, and policy building at various levels of academic life.

1.2.1. The definition of a CRIS

The concept of unified research information systems dates back to the 1970s, when the first international initiatives aimed at developing tools for the collection of data on the scientific and technological potential of individual countries emerged [1]. Their aim was to use quantitative and quantifiable data to shape scientific policy. A landmark document was the study of a global scientific information system published in 1971 by UNESCO [69].

The early research information management systems developed in the 1970s were marked by the use of batch processing systems primarily employed by research funding organisations. They gathered data on research grants awarded, together with the principal investigators’ credentials, their institutional affiliations, and associated financial transactions [28, p. 335].

[3] This is a statistical term defined as the ratio of people in education (as of the beginning of a school year) at a given level, to the population (as of 31 December) of the age group that corresponds to that level

In the 1980s and 1990s, growing awareness of the strategic value of scientific information led to the establishment of several European initiatives—including projects supported by the European Commission, such as IDEAS and EXIRPTS—that aimed to develop interoperable research management systems. These efforts resulted in the preliminary Common European Research Information Format (CERIF) 91 standard, which was later extended into the more comprehensive and semantically richer CERIF2000 standard [28, p. 335].

Presently, academic institutions, regional and national public administration entities, and supranational organisations use IT systems to collect and share information on ongoing scientific projects and to shape knowledge-based science policy. These systems are referred to jointly as research information systems (RISs).

Information systems that support science and higher education management can be divided into four main classes: research information management systems (RIMSs), current research information systems (CRISs), research data or publications repositories, and student management systems (SMSs). Each of these types of system performs separate functions and is based on a different data model. RIMSs focus on the management and presentation of information on researchers' scientific activities—in particular publications, grants, and researcher profiles—often integrating data from multiple sources and supporting mechanisms for the evaluation of scientific achievements and visibility. While such solutions are relatively popular at the level of individual universities or their associations, few countries have RIMSs at the central level [67, p. 16].

CRISs, although they overlap with some of the functions of RIMSs, have a broader scope that covers not only publications, but also patents and other products of scientific activity, projects, institutions, research infrastructure, funding, and links between these elements. CRISs are often based on the formal CERIF model, which enables high interoperability and the precise mapping of complex relationships in the science ecosystem [12].

Repositories are systems used primarily for the long-term storage and sharing of research results—particularly publications, doctoral dissertations, and, increasingly, research data—as part of an open access policy. They are typically based on simpler metadata schemas (e.g. Dublin Core Metadata Element Set, DCMES[4]) and do not have the functionality typical of CRISs. Research data repositories can be divided into four main types: thematic, discipline-specific, institutional, and general-purpose. Thematic repositories, such as the Protein Data Bank [80], store highly structured datasets in narrow research domains. Discipline-specific repositories, such as the Social Data Repository (RDS[5]), Dryad[6] for medical and natural sciences, or Archaeology Data Services[7], collect datasets from broader academic fields, improving data discoverability for re-

[4] dublincore.org (accessed 17 November 2025)
[5] rds.icm.edu.pl (accessed 17 November 2025)
[6] datadryad.org (accessed 17 November 2025)
[7] archaeologydataservice.ac.uk (accessed 17 November 2025)

searchers in a given discipline. Institutional repositories like DBUW[8]) store research data generated by members of a particular institution, often reflecting its disciplinary focus. General-purpose repositories—such as preprints repository arxiv.org[9], RepOD[10], or Zenodo[11]—are open to publications and data from any field or institution, and are suitable for interdisciplinary or project-specific results of scientific activity [25].

Finally, SMSs constitute a separate category of systems used to manage the teaching process. They include data on students, subjects, course registration, credits, and diplomas. Their main purpose is to support academic administration in education management, not research.

Despite some points of contact (e.g. the need to report to common national systems), each of the types of system described above operates on a different set of data, responding to the varying needs of students and staff in the higher education and science systems, as well as scientific and educational institutions [14, p. 2].

CRISs, for that reason, are the systems with the most complex architectures. They combine data from various sources—including institutional RIMS, SMS, and repositories of publications and research data—into a single unified ecosystem. In modern CRISs, data on objects and entities that operate in the science and higher education sector are used to support business processes that meet the needs of various stakeholders through microservices [46]. These systems are updated frequently and provide the right contextual information to the right user at the right time [29, 30].

1.2.2. Key features of CRISs

CRISs are specialised IT systems that are used to collect, manage, process, and share contextual data on scientific and research activities. The key functions of CRIS systems include: 1) recording data on scientific activity—including academic staff, academic institutions, research projects, publications, patents, artistic achievements, and infrastructure; 2) supporting reporting and evaluation processes—CRIS systems provide data for institutional reporting, internal and external evaluations, including support for the allocation of funds for scientific activities and strategic analyses; 3) integrating with other systems, such as repositories, financial systems, bibliographic databases (e.g. Scopus, Web of Science), and persistent identifiers (e.g. ORCID[12], ROR[13]); 4) managing policy compliance, e.g. monitoring compliance with Open Access (OA) requirements, Plan S, RFOs policies; 5) disseminating and promoting scientific output through public institutional portals and the aggregation of data into national or international systems.

[8] danebadawcze.uw.edu.pl (accessed 17 November 2025)
[9] arxiv.org (accessed 17 November 2025)
[10] repod.icm.edu.pl (accessed 17 November 2025)
[11] zenodo.org (accessed 17 November 2025)
[12] orcid.org (accessed 17 November 2025)
[13] ror.org (accessed 17 November 2025)

1.2.3. CRIS-type systems

Although CRIS and RIMS systems can be classified according to various criteria, due to their operating model, the services they provide, and the business processes they support, the following criteria are of key importance: 1) ownership and licensing; 2) compliance with international standards and ensuring interoperability; 3) the scope of science and higher education systems supported; and 4) the development and financing model.

Based on ownership and licensing criteria, CRIS systems can be divided into commercial, and therefore closed-source and open-source systems developed on a non-commercial basis. Elsevier Pure[14] and Clarivate Converis[15] are examples of commercial RIMS/CRIS systems used for the comprehensive management of information on scientific activity. Both solutions support the collection, integration, organisation, and presentation of data on scientific achievements, research projects, institutional activity, and related information (e.g. affiliations, funding sources, bibliometric metrics).

Elsevier Pure is integrated with the Scopus database of publications, abstracts, and citations, as well as the SciVal analytics platform. It offers a ready-made data structure based on the CERIF model, the automatic synchronisation of publications and metrics, and a public portal that presents the profiles of researchers and institutions. The system is highly automated, but offers less flexibility in adapting the data model to the specific needs of academic institutions.

Clarivate Converis stands out for its greater configuration flexibility and extensive modules that support the research project lifecycle (pre-award, postaward), as well as internal process management. The system is integrated with the Web of Science, InCites, and ORCID databases, and its open architecture enables it to be adapted to different organisational models and reporting systems at the level of scientific institutions.

Although both systems are CERIF-compliant, support euroCRIS, and are recommended by that organisation, they also differ in their development strategy: Pure focuses on interoperability in the Elsevier ecosystem, while Converis focuses on configurability and extensive integration capabilities with external institutional and national systems.

Another example of a commercial CRIS solution is Symplectic Elements[16], a flexible and automated CRIS system. Its key advantage is its extensive integration with external data sources, including PubMed, Scopus, Crossref, Web of Science, and ORCID. The system enables the easy collection, deduplication, and organisation of data on publications, projects, grants, patents, research data, activities, and researchers' affiliations.

[14] elsevier.com/products/pure (accessed 17 November 2025)
[15] clarivate.com/academia-government/scientific-and-academic-research/research-funding-analytics/converis (accessed 17 November 2025)
[16] symplectic.co.uk/theelementsplatform (accessed 17 November 2025)

Users can manage their profiles independently and institutions can generate reports in line with evaluation and management requirements. Elements is highly interoperable and focuses on automation and end-user ergonomics.

Among the CRIS systems developed in the open-source model, DSpace-CRIS[17] and VIVO[18] deserve special attention, as they offer advanced functionalities for the management of information on scientific and research activities. DSpace-CRIS, developed by 4Science as an extension of the popular open-source repository system DSpace, integrates the functions of a classic repository with the ability to model complex data structures related to projects, people, organisational units, and research results. The system is based on a relational data model and supports standards such as CERIF and the OpenAIRE Guidelines, which makes it a flexible tool in institutional reporting and the presentation of scientific achievements. VIVO, developed by an international consortium including Cornell University and DuraSpace, represents a semantic approach in which research data is described using ontologies. VIVO enables the creation of networks of connections between researchers, projects, publications, and institutions, which promotes knowledge exploration and integration with global information ecosystems. An example of the large-scale implementation of the VIVO system is the Brazilian BrCRIS[19], in which data collected from various national and international sources is aggregated and exported into VIVO. Although different in architecture and design, both systems serve as mature, open alternatives to commercial CRIS solutions. They support openness, interoperability, and the transparency of research data and metadata.

The Directory of Research Information Systems (DRIS)[20] maintained by euroCRIS lists 1,455 active CRIS implementations, a large portion of which are in-house-built solutions; in other words, they were developed internally by institutions. In-house solutions are gaining importance for institutions that seek to maintain comprehensive control over their architecture, processes, and data models—particularly in the context of adherence to local legislation, scientific policies, and reporting requirements. While commercial and open-source systems occupy strong positions in this field, in-house solutions are particularly salient in instances in which implementations encompass all institutions in a nation or institutions with particular interoperability requirements. Such models are frequently developed in the context of publicly funded initiatives or consortia. Their adaptability enables the comprehensive customisation of the CERIF data model, adherence to local standards, and the automated public dissemination of data via API. The most prominent examples of such platforms are those developed nationally, including CRIStin (to be replaced in the future by the Norwegian Research Information Repository (NVA))[21], the Italian Institutional Research Information System (IRIS), the Indian Research Infor-

[17] 4science.com/dspace-cris (accessed 17 November 2025)
[18] vivoweb.org (accessed 17 November 2025)
[19] brcris.ibict.br/en (accessed 17 November 2025)
[20] dspacecris.eurocris.org (accessed 17 November 2025)
[21] nva.sikt.no (accessed 17 November 2025)

mation System (IRINS), and the Finnish Research.fi system. This class of system also includes POL-on, which covers the entire science and higher education sector in Poland, and the OMEGA-PSIR[22] system, developed at the Warsaw University of Technology and dedicated to academic institutions.

Based on data collected in the DRIS catalogue, CRIS systems can also be classified according to the scope of their operation. The vast majority are institutional systems; as many as 1,431 of the reported CRISs operate at the level of individual universities, research institutes, or other scientific units. Less numerous are systems with wider scopes: national (26), regional (8), and international (1). Another key category is systems implemented by research funding institutions (13), as well as those that aggregate data from multiple sources (5). Occasionally, systems are also used at the level of faculties or internal units of scientific institutions (2). This data demonstrates the dominance of local solutions that respond to the needs of specific organisations, as well as the growing importance of systems that support the coordination and monitoring of scientific policies at higher levels of management.

1.2.4. CERIF

Contemporary science operates in an increasingly complex information environment, in which interoperable data exchange between IT systems—such as publication repositories, project databases, evaluation systems, grant registers, and research data management systems—is becoming a prerequisite for the effective management of information, the transparency of funding, and the dissemination of research results. In response to this need, CERIF, a formal, relational data model developed and maintained by the euroCRIS organisation, was developed. CERIF is a standard for describing and exchanging data in CRISs, which enables the structured mapping of scientific institutions, individuals, projects, publications, implementation results, funding, and the relationships between entities, with consideration for their dynamics over time. Key feature of the CERIF model are its flexibility and semantic precision, which make it the foundation for the technical and semantic interoperability of information systems in science, both nationally and internationally. It not only enables the automatic processing and aggregation of data from various sources, but also supports open science policies, research performance evaluation, and the development strategies of scientific institutions based on harmonised and reliable information.

Although many contemporary CRIS systems claim to be interoperable with CERIF, only some of them implement the model fully and natively. Among the systems that are fully compatible with CERIF are Converis, Pure, and DSpace-CRIS, all of which use the CERIF structure in their internal databases and enable data export in the CERIF-XML format in accordance with the requirements of euroCRIS.

[22] omegapsir.io/pl (accessed 17 November 2025)

Both Pure and Converis, although based on the CERIF model, enable it to be extended with additional elements and relationships that correspond to the specific needs of research institutions. In the case of Converis, the user can define their own entity types and attributes while maintaining the structure and logic of CERIF. Pure, on the other hand, enables the data model to be expanded through the configuration of classification schemes and semantic extensions, which enables institutions to enrich the descriptions of their research units without compromising consistency with the CERIF model.

Other solutions, although they often enable data to be exported in the CERIF-XML standard, are not built on the native CERIF model. One example is the Indian IRINS, which is based on the DSpace-CRIS architecture but operates internally on the VIDWAN data model. Data mapping to CERIF is only conducted at the stage of integration with external registers [44]. A similar situation applies to numerous in-house systems developed locally by academic institutions (e.g. in Poland, Germany, Norway, Czechia, and India), which often implement CRISs without full compliance with CERIF, using their own data models and APIs, and only partially implementing the elements required by DRIS-type catalogues. The same applies to the POL-on system, which is incompatible with CERIF—although internally it describes entities according to a domain-specific ontology as people (students, doctoral students, and researchers), institutions, research projects, publications, research data, patents, funding, and relationships between them.

The lack of full compliance of CRIS systems with the CERIF data model is primarily due to the high degree of complexity of the standard and the significant implementation costs involved in its full adaptation. CERIF, as a highly standardised relational model focused on representing contextual relationships between scientific information objects, requires advanced database architecture and specialist technical knowledge. In many cases, academic institutions or national research organisations opt to develop systems based on local data models that are better suited to applicable legal regulations, reporting schemes, and administrative needs. For example, in India, the IRINS system is based on the national VIDWAN model; in Germany, many institutions use internal solutions built on the HISinOne system [26]. Where there is no formal requirement to report in CERIF format (e.g. from research funding agencies or national evaluation systems), institutions typically do not see compliance with the CERIF standard as a priority. Local CRISs are often created to support narrowly defined functions—such as publication output recording, grant process management, or periodic staff evaluation—in which the full semantic expressiveness of the CERIF model is unnecessary and its implementation could even complicate simpler procedures.

The consequence of noncompliance with CERIF is limited international interoperability and reduced usefulness in the context of cross-border research data exchange, integration with European institutional infrastructure and registries (such as OpenAIRE or the DRIS catalogue), and inclusion in international evaluation systems. The need to map local data models to the CERIF structure often results in the loss of semantic information, difficulties in automating data export, and a greater risk of inconsistency. As a result, such systems have limited ability to support open science strategies, and hinder the implementation of transparency and accountability policies in scientific research.

For this reason, the development of CERIF-compliant systems—or, at least, systems that partially implement the model—is seen as a strategic step towards increasing the interoperability of scientific information systems at the European and global levels.

1.2.5. CRIS stakeholders

CRISs serve a diverse group of stakeholders, and this is reflected in their architecture, functionalities, and role in the science ecosystem. One key group of users is researchers, who use CRISs to document and present their scientific achievements, manage their research portfolios, and fulfil their reporting and grant-application obligations. The second key group is institutional administrators—the people who are responsible for coordinating scientific activities at the level of universities, faculties, or institutes—for whom CRISs serve as tools for reporting, evaluation, strategic planning, and generating bibliometric analyses. Research funding institutions also play an important role, using CRIS data to monitor the use of funds and to evaluate the results of scientific projects. From the perspective of research policy, the data collected in CRISs are also used by public administration bodies, such as ministries and evaluation agencies, to monitor and shape research and development strategies and to allocate funds. Other stakeholders include academic libraries [8] and IT departments, which are responsible for maintaining the systems, as well as integrating them with repositories, project databases, and human resources systems, and ensuring data quality and interoperability. In the context of open science, stakeholders in CRISs also include the public and the media, for whom data from such systems can provide information on research activities, increasing their transparency and facilitating communication between science and society. CRISs are also used by members of the scientific community, such as those who publish journals (in the selection of reviewers and editors) and information brokers who facilitate the flow of knowledge and technology from science to business. From this perspective, CRISs can be seen as platforms for the transfer of knowledge from the science system to society, while ensuring the efficient management and accountability of science [30, p. 25].

Students and prospective students are an increasingly important, though often overlooked, group of stakeholders in CRISs. Their role is not limited to being passive recipients of knowledge, but extends to active participation in the research environment—particularly at the master's and doctoral levels. Access to information collected and made available by CRISs—including on current research projects, researcher profiles, publications, research infrastructure, and international partnerships—can influence educational and academic decisions heavily. Prospective students can use publicly available CRIS systems or rankings based on data from them. For example, the Polish *Perspektywy* University Ranking [52] makes extensive use of data from the national POL-on system, which accounts for approximately 55% of the overall ranking position of a university. Moreover, prospective students may use CRISs to compare universities and departments in terms of scientific achievements, involvement in international projects, or research topics studied at doctoral schools. In turn, students already enrolled at universities can resort to CRISs as navigation tools in the institutional and substantive structure of research, facilitating, for example, the selection of a supervisor, thesis topics, or involvement in research activities. From the perspective of higher education institutions, this means that the presentation layer of CRISs must be adapted to the needs of users who are not experts in scientific research.

1.2.6. Selected national-level CRISs

Analysis of the reasons for the creation and the review of the objectives of CRIS-type systems in Europe reveals a picture of the common problems they are intended to solve and the questions they are intended to answer. This approach seems a useful starting point for comparing different CRIS-type systems. Although many CRIS-type systems operate at the university or regional levels in Europe, the list will also include such systems that operate at the national level (such systems are analogous to POL-on).

One of the major goals of CRIS-type systems is to become effective and efficient tools that provide answers to the questions posed by their founders. These answers are often hidden in denormalised data scattered across multiple institutions. Gathering this data for national-level funders requires a commonality of language at the national level. Collecting data for the needs of funders at the pan-European level requires a common language at the international level, such as CERIF.

Adaptation of the centrally developed (at the European level) data model is a challenge for system developers. In 2014, only seven European systems were deemed compatible with the CERIF model [45]. One of them was the Flanders Research Information Space (FRIS)[23], a regional portal that collects information on Flanders scientists and research. The Flemish government wanted to give its scientists the opportunity to increase the visibility of their work. Adopting an international data model makes the data easier for a wider audience to understand. FRIS shares its data openly using, CERIF SOAP API[24].

Due to the benefits of a common language, more systems are joining the group of CERIF model implementations. In 2019, the National Academic Research and Collaborations Information System (NARCIS)[25] was in an advanced phase of mapping its model to CERIF [65]. NARCIS is a national system that collects information on Dutch scientists, scientific publications and projects, and datasets. Although it has its own native data model, its creators decided to try to disambiguate their domain with the pan-European model. By ensuring a common language, it is possible to ensure high interoperability of

[23] researchportal.be/en (accessed 18 June 2021 and 03 July 2025)
[24] frisr4.researchportal.be/ws (accessed 17 November 2025)
[25] narcis.nl (accessed 18 June 2021 and 04 July 2025)

the data and to provide answers to anyone who can ask a question in that language. As of 3 July 2023, the NARCIS website was taken offline, and the NARCIS service was terminated.

Currently, information on research conducted by Dutch institutions can be found on the Research Results Portal for the Netherlands. The comprehensive collection of Dutch open research information system includes publications, research datasets, research software components from Dutch institutional repositories, and CRISs. It is also linked to research grants (NWO) and research performing organisations (UNL, VH, KNAW, NFU), all linked together through citations and semantics[26].

Science and research based on very large datasets can advance because of the increasing availability of such datasets. Open access to scientific publications and scientific datasets is a key element in terms of both usability and impact on scientific activity [22]. An open model of science communication supports the progress of research. One task of CRIS-type systems is to provide open access[27] to datasets and scientific publications, as well as monitoring policies related to that issue.

The developers of the FRIS system emphasise the opening of access to scientific publications and datasets. The driving force behind this approach was the 2017 pilot of the European Commission's Open Data programme. The initiative is also supported by the Flemish government[28].

Monitoring the open access situation among Norwegian scientists is one of the tasks of CRIStin[29], a Norwegian national system that enables access with institutional, local credentials. It is characterised by its enablement of multiple users to collaborate on a single record while appointing 'super users' from each institution who are responsible for the correctness and completeness of data. The system is being used to monitor the open access status of Norwegian researchers' publications [32] after the country's government established national guidelines for open access to scientific publications in 2017.

The most common task conducted by European information systems that collect information on science is to ensure the availability of information on the state of science for the needs of various stakeholders, including the general public. A model example of a system largely dedicated to this purpose is the Polish system RAD-on[30], an aggregate of information from POL-on. The main task of RAD-on is to transform data collected from various domain systems into unified information in a given business context. The system offers open reports, predefined analyses, and raw data to everybody.

[26] netherlands.openaire.eu (accessed 17 November 2025)
[27] open access, OD, OA
[28] researchportal.be/en/open-science (accessed 18 June 2021 and 04 July 2025)
[29] cristin.no (accessed 17 November 2025)
[30] radon.nauka.gov.pl (accessed 17 November 2025)

The Estonian Research Information System (ETIS)[31] boasts an extensive search engine for access to the portal (and, therefore, access to information for users who do not use open API, for example). The Estonian solution focuses on information on scientific institutions, researchers, and their projects and scientific publications. Simultaneously, it is also used to support the grant application process and to report progress in grant-based research projects. The scientific project search engine on the portal, established by the Estonian Ministry of Education and Research, offers a search of the project database through up to eighteen search criteria that can be combined. It presents the details of projects, including the budgets allocated to them from public funds.

The accountability of public funds and their transparent distribution in the scientific world were the ideas behind the creation of the Swedish SweCRIS[32] system. It enables every person who has access to the internet to access information on how science funding is distributed to Swedish recipients. Twelve agencies (both public and private) that distribute grants feed the system with data. Information is collected on the recipients of and the destinations of the funds. The system has unified the way grant-making agencies report data. Simultaneously, providing data to the national repository saves time; recipients previously had to report to multiple institutions interested in the information (fields of research include: natural sciences, engineering and technology, medical and health sciences, agricultural and veterinary sciences, social sciences, humanities and arts, and unclassified elements).

Meeting the needs of many institutions simultaneously is a desirable operating model for a CRIS-type system. An excellent example of the implementation of such a model is POL-on, which is the subject of analysis in this book. Other European systems also seek to translate the effort from their creation into benefits in the elimination of unnecessary and repetitive bureaucracy. In 2019, Statistics Norway [32], which is the main provider of statistical reports and analyses on Norway[33], joined the CRIStin system.

In the face of large, nationwide tasks, the systems that collect information on science do not forget those without whom science would not exist: its creators, the scientists. ETIS provides researchers with the opportunity to consent to the publication of their CVs, predefined by the system and supplemented on the basis of data collected therein. Researchers' profiles can also be viewed on the NARCIS system. Information on a scientist in an extended form is provided by the *Ludzie Nauki* portal[34] (previously known as the *Nauka Polska* portal)—the oldest database maintained by OPI PIB.

Despite their universal origins, CRIS-type systems in Europe vary in the extent of their functionality and in their links to internal systems. While 80% of the CRIS-type systems whose administrators responded to the 2016 EuroCRIS questionnaire supported handling data on scientific publications, only half of them collected the full content of

[31] etis.ee (accessed 17 November 2025)
[32] vr.se/swecris (accessed 17 November 2025)
[33] ssb.no (accessed 17 November 2025)
[34] ludzie.nauka.gov.pl (accessed 17 November 2025)

those publications. Only 30% of the systems were designed to support researchers' evaluation with their functionalities. Of the CRIS-type systems examined, 68% were integrated with human resources systems and 63% with repositories of individual institutions, and only 8% were linked with library systems [53].

In 2025, based on the information that can be found in the euroCRIS Institutional Research Systems Catalog (DRIS)[35], by scope of activity, there are approximately twenty-six national (this number also includes POL-on in Poland), eight regional, 1,430 institutional, and nineteen other CRIS systems. The CRIS-type systems listed in this subsection are characterised by many individual features; however, a visible commonality of goals exists to which they were established. This is strong evidence of the sense of cooperation between European CRIS-type systems and the intensive exchange of knowledge, experience, and good practices between their creators.

1.3. Information systems for scientific evaluation

Systems for the evaluation of scientific achievements play a key role in the functioning of contemporary science and higher education systems. Their primary goal is to evaluate the quality, effectiveness, and impact of research conducted by scientists, research teams, institutions, and entire academic units. This evaluation is not merely an administrative process: it serves as the foundation for making decisions on the allocation of funding, the granting of academic credentials, the shaping of scientific policy, and the promotion of research excellence. In the context of growing international competition, limited public resources, and the need to enhance the societal impact of science, countries are implementing increasingly complex and diverse evaluation systems. Although their goals are similar—promoting research quality and rationally managing funds—the methods, criteria, and approaches used vary considerably, depending on the country and its academic tradition.

The first evaluation system was the British Research Assessment Exercise (RAE) in 1986, which was initially based solely on expert evaluation. The number of units assessed and the evaluation parameters changed considerably during subsequent evaluations conducted in 1989, 1992, 1996, 2001, and 2008[36]. Scientific institutions, such as universities and research units, were required to prepare detailed applications that presented their research achievements—primarily scientific publications, ongoing projects, and other forms of scientific activity. Those applications were subjected to expert review, which determined the quality of the research conducted. The final assessment was based on a percentage that indicated the percentage of a given unit's research activity that met certain quality standards, such as world-class, higher national, or lower local standards.

[35] dspacecris.eurocris.org/cris/explore/dris (accessed 17 November 2025)
[36] RAE homepage (archived): webarchive.nationalarchives.gov.uk/ukgwa/20091118100524/http://rae.ac.uk (accessed 17 November 2025)

While the RAE played a significant role in streamlining and professionalising research assessment in the United Kingdom, it faced increasing criticism as time progressed [41]. The system was considered to be overly formal and bureaucratic, as well as exceptionally time-consuming for universities preparing applications. Moreover, the RAE focused almost exclusively on publications, largely omitting other aspects of research activity, such as the real social, economic, and cultural impact of research results. As expectations for science rose—particularly in the context of innovation, commercialisation of results, and collaboration with the societal environment—the RAE began to be perceived as a tool ill-suited to new challenges. These limitations prompted the creation of a more modern and sustainable system: the Research Excellence Framework (REF)[37].

Introduced in 2014, as a direct successor to the RAE, the REF aimed not only to streamline the evaluation process but, above all, to adapt it to the changing realities of science and the growing importance of its societal impact. Another edition was conducted in 2021, and the next is planned for 2029.

The REF was designed to assess the research achievements of UK research institutions in a more balanced and multidimensional manner. Beyond the quality of publications, the system focuses on the impact of research on society, the economy, and culture, as well as the institutional conditions in which it is produced. The REF not only classifies research institutions, but also promotes research excellence and encourages universities to develop sustainable strategies for scientific development, and to invest in research infrastructure and human resources.

The REF is based on three key evaluation components. The greatest weight is given to outputs: scientific publications, such as articles, books, and patents. This element represents 60% of a researcher's final grade. The second pillar involves assessing the impact of research beyond academia, encompassing its application in public, economic, and social life. This criterion was introduced in response to earlier criticism of the RAE and currently accounts for 25% of a researcher's overall grade. The third component is research environment, which includes analyses of the research environment, development strategy, researcher support, and infrastructure, with a weight of 15%.

Unlike many other evaluation systems, the REF does not use mechanical scoring; instead, it uses a quality scale ranging from 1* (minimal scientific value) to 4* (world-class quality). Experts play a key role in that scoring; the assessments are based on thorough analysis of submitted articles' content and descriptions of research's impact. The system, as a result, can be considered more qualitative than quantitative.

[37] impact.ref.ac.uk (accessed 17 November 2025)

The REF has enormous practical significance, as it is the basis for the allocation of considerable public funds for research activities (so-called quality-related funding). Simultaneously, the system is an important strategic tool: its results are used by academic institutions to plan development, recruit staff members, and build prestige at home and abroad [63].

Following the UK's introduction of its national research evaluation system, many other countries decided to implement their own mechanisms for evaluating research activities. While they did not imitate the British model in its entirety, many drew inspiration from its general principles, such as regular evaluations, expert participation, a focus on research quality, and linking results to institutional funding.

The Netherlands introduced a structured evaluation system early on. The country's Standard Evaluation Protocol (SEP), in use since the 1990s, is based on a six-yearly assessment of the quality of research teams and institutes. The process focuses on three key aspects: scientific quality, societal relevance, and the long-term strategy and sustainability of research activities[38]. Unlike the British REF, the SEP neither assigns numerical ratings nor directly determines funding, but rather serves as a basis for internal reflection and research development strategies at individual institutions.

A similar approach has been adopted in Finland, where local evaluation systems have been developed since the 2000s. Since 2015, these efforts have been coordinated by the Research Council of Finland. The council's evaluation focuses not only on research quality, but also on openness and international collaboration[39].

Italy has introduced its own research quality evaluation system, VQR (Valuation of Quality of Research), managed by the national agency ANVUR. This process combines bibliometric elements with expert assessment and influences the allocation of funding among institutions[40].

Norway and Denmark, on the other hand, have implemented so-called 'publication models', in which articles are evaluated based on the prestige of the journals in which they are published. These systems are simpler and are used primarily to distribute funding among universities[41].

[38] VSNU, NWO, KNAW / Standard Evaluation Protocol 2021–2027, nwo.nl/en/evaluations-nwo-institutes (accessed 17 November 2025)
[39] Research Council of Finland / Evaluation of Research Quality, aka.fi/en/from-research-to-society/responsible-science (accessed 17 November 2025)
[40] *Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca* (ANVUR), www.anvur.it (accessed 17 November 2025)
[41] Norwegian Centre for Research Data / Norwegian Publication Indicator, npi.hkdir.no (accessed 17 November 2025)

In contrast, Germany has maintained a decentralised scientific evaluation model. Evaluation is conducted by individual institutions, such as the *Deutsche Forschungsgemeinschaft* (DFG), the Max Planck Society, or the Helmholtz Association. Research teams and projects are evaluated, with expert reviews and grant-winning effectiveness playing a key role[42]. The German system, despite its lack of central oversight, is considered to be highly effective and is conducive to high-quality, competitive research.

The evaluation of scientific achievements at the level of research institutions and universities is based on three basic models: peer review, publication count analysis, and citation analysis. Currently, mixed approaches that combine quantitative and qualitative methods are being used more frequently. Each approach has its advantages and limitations, and the selection depends on the purpose of the evaluation, the nature of the science system in a given country, and the type of activity being assessed.

Expert-based models are among the most traditional forms of evaluation. They involve a group of independent experts assessing selected elements of scientific achievements—most often publications, research projects, and also, increasingly, the societal impact and strategic nature of research activities. A key advantage of this approach is the ability to assess the quality of scientific content precisely and contextually, with consideration for the specificity of the field, the language of publication, the form of the work (e.g. monograph), and the presence of innovative solutions (which is not always reflected in bibliometric indicators). This system is used, among others, in the British REF, where experts evaluate not only publications, but also impact case studies, i.e. evidence of research's impact on society and the economy. The disadvantage of expert-based models lies in their subjectivity and high organisational cost; they require much time and resources, as well as transparent procedures to avoid accusations of bias or conflicts of interest[43].

The second approach is a model based on the number of publications. Its main advantage lies in its simplicity: publications can be counted easily and compared between units or researchers. Such a system may be used as part of internal reward systems or in some countries' institutional evaluations. However, the model has serious limitations. First and foremost, it rewards quantity over quality, encouraging the publication of large numbers of articles—even those of marginal scientific value. It also fails to account for differences between disciplines; in the humanities, the publication rate is slower than, for example, in the natural sciences. This model was used in Poland until the 2018 evaluation system reform, before which publications were evaluated almost exclusively based on number and inclusion in a scored list of journals [66].

The third model, based on citations, is considered more advanced than any one that merely counts publications. Its goal is to measure the real impact of scientific works on the academic community. Citations are considered evidence that a given work has been

[42] *Deutsche Forschungsgemeinschaft*, dfg.de/en (accessed 17 November 2025)
[43] REF 2021 Guidance, ref.ac.uk (accessed 17 November 2025)

noticed and used by other researchers. Citation models are particularly popular in the natural sciences, medicine, and engineering, where the community's response time to publication is relatively short and databases such as Web of Science or Scopus cover a significant portion of the body of work. A drawback of this approach is its uneven distribution across disciplines. In the humanities and social sciences, citations appear much less frequently, and in some cases, their significance is marginal. Moreover, citations might result from controversies or negative reviews, which do not necessarily indicate high research quality [78].

In response to the shortcomings of single-sided approaches, many countries have introduced mixed systems that combine quantitative and qualitative evaluation. Such models typically rely on the numerical analysis of selected aspects of scientific output, such as publications or citations, and expert assessment of other elements, such as research impact, the quality of the research environment, or development strategies. This approach aims to balance objective indicators with expert context, increasing the validity and legitimacy of the entire evaluation.

Evaluation commenced in Poland in the early 1990s. Assessments were conducted *ex post*. The evaluation procedure was highly flawed, excessively complex, and required the definition of an increasing number of indicators for each research unit under evaluation. It was based on annually submitted Research Unit Surveys, which were aggregated for each evaluation. Later, The Ministry of Polish Higher Education eliminated the requirement for units to submit annual Research Unit Surveys. Initially, units had been required to submit the required data that covered the entire evaluation period only prior to the evaluation (more recently, evaluation has been based solely on data contained in the POL-on system). Such evaluations were conducted in 1999, 2003, 2006, 2010, 2013, and 2017. Higher education institutions, Polish Academy of Sciences institutes, research institutes, and others were evaluated separately, depending on the number of employees that declared they conducted research. Research units were evaluated in joint evaluation groups defined by the Minister of Science. Since 2003, evaluations have been conducted using computer software, which in recent years has evolved into a system specifically for that purpose: the Scientific Achievement Evaluation System (SEDN). Another evaluation is planned for 2026, under revised principles, in which entire institutions will be evaluated in the specific scientific disciplines in which their research is conducted.

Under the most recent reform of higher education and science in Poland, known as Law 2.0 or the Constitution for Science [72], a new, centralised system for the evaluation of scientific activity was introduced. The goal of the reform was not only to simplify previous procedures, but also to increase the transparency and comparability of the achievements of scientific institutions at the discipline level. This system is based on periodic evaluations conducted by the MNiSW, which consider three main criteria[44].

The first is the quality of scientific publications. This criterion evaluates only publications by selected academics, identified by an institution within the limits assigned to the number of employees. Only works submitted by an institution for evaluation are assessed. Publication scores are determined based on ministerial lists of scientific journals and publications, which classify individual sources according to their assigned point values[45].

The second criterion concerns the financial impact of scientific activity. This pertains primarily to the amount of funding obtained by a research unit through national and international research projects. The efficiency indicator considers not only the total value of grants obtained, but also relates it to the number of research staff at a unit, which ensures greater fairness when comparing institutions of different sizes.

The third criterion, which is relatively new in the Polish system, is the impact of scientific activity on the socioeconomic environment. Research units are required to submit so-called impact descriptions: narrative case studies that illustrate how research findings have contributed to economic development or improved public policy, healthcare, culture, education, or other areas of social life. This element of evaluation aims to demonstrate that science reaches beyond academia, and has a real impact on society and the economy.

Based on the results obtained in the three areas described above, each scientific discipline in which research is conducted at a given institution (e.g. a university or research institute) is assigned one of five scientific categories: A+, A, B+, B, or C. The category is not merely a ranking label: it also has particular legal and organisational implications. Primarily, it determines whether a given institution has the right to award academic degrees (doctorates and habilitations) in a given discipline. In addition, it influences the level of funding received from the state budget, as well as the overall prestige of the institution in the national and international scientific communities.

[44] KEN, gov.pl/web/nauka/komisja-ewaluacji-nauki (accessed 17 November 2025)
[45] Ministerstwo Edukacji i Nauki, Wykaz czasopism i wydawnictw naukowych, gov.pl/web/nauka/ujednolicony-wykaz-czasopism-naukowych (accessed 17 November 2025)

1.3.1. Scientific achievement evaluation system

In Poland, evaluation is conducted using SEDN[46], which forms part of the POL-on[47] ecosystem and operates on the basis of data from, among others, the PBN[48], OSF[49], and other registers and databases, including international ones. This enables the automated analysis of publications, assigning points based on ministerial lists, verifying the authors' participation, and linking data to the scientific disciplines in which they operate [38].

1.3.2. Optimisation algorithm

The core of SEDN is an optimisation algorithm whose task is to select a set of achievements (primarily publications) in a discipline that maximises the final score for the individual. The problem is complex: each publication can have multiple authors who work at different institutions, and regulations impose various limits on the number of achievements that can be included for a single person, institution, or publication type.

Each article or monograph is divided into 'shares'. For example, if a publication is worth 100 points and has a single author, that author's share is 100 points. If it has two authors, each receives fifty points. Therefore, the algorithm does not work on entire publications, but on the shares within them.

The goal of the algorithm is to maximise the total points in a given discipline while adhering to constraints such as:

- the maximum number of shares that can be assigned to an individual
- limits for individuals (e.g. a maximum of four publications per person)
- limits for monographs (e.g. not to exceed 100 points per person per book)
- time limits (e.g. a minimum number of articles from 2019–2021).

Mathematically, this is an NP-hard combinatorial problem; to find the perfect solution, one would have to check all possible combinations of shares (e.g. for ten publications, it is 2^{10} combinations, and for hundreds of thousands of publications, the numbers become incomprehensible).

[46] sedn.opi.org.pl (accessed 17 November 2025)
[47] polon.nauka.gov.pl (accessed 17 November 2025)
[48] A portal of the Polish Ministry of Science and Higher Education that collects information on publications by Polish scientists, the publication output of research institutions, and Polish and foreign journals. It is part of POL-on, pbn.nauka.gov.pl (accessed 17 November 2025)
[49] A system dedicated to Polish scientists and entrepreneurs that is used to register and handle applications for the financing of research and development projects and studies with funds provided by the MNiSW, the Polish NCN, and the Polish NCBR, osf.opi.org.pl (accessed 17 November 2025)

The optimisation algorithm uses a genetic algorithm-based approach. Each potential solution (a set of selected contributions) is represented as a binary vector (0 – contribution not considered, 1 – contribution considered). The algorithm then randomly generates initial solutions and 'evolves' them in subsequent generations using crossover (combining solution fragments) and mutation (random change) operations. In each generation, only the solutions that yield the best scores while satisfying the constraints are selected. The process is repeated multiple times until a stable solution is obtained. Importantly, ready-made integer programming solvers, such as ORTools[50], were used to generate the initial solutions, helping to find suitable starting points more quickly.

For each discipline, the solution generated hundreds of results and the process was repeated nightly for several months before the final evaluation. Institutions could test the algorithm in the SEDN demo version and discover which publications maximised their results. Although the genetic algorithm does not guarantee the absolute best possible result, it does produce solutions that are most likely optimal in an acceptable time frame. This is a compromise between accuracy and the feasibility of computations on a scale of hundreds of thousands of publications and over 100,000 researchers and scientists.

[50] OR-Tools, developers.google.com/optimization (accessed 17 November 2025)

CHAPTER 2

SYSTEM OVERVIEW

Iwona Kucharska
Dr Anna Łobodzińska
Marek Michajłowicz
Dr Piotr Rodzik
Sławomir Rybka

The POL-on system represents one of the most comprehensive and integrated national-level IT infrastructures that supports science and higher education in Europe. Following the contextual and comparative analysis presented in Chapter 1 of this book, this chapter offers a detailed examination of POL-on itself: its genesis, structure, legal foundations, and operational scope. The chapter aims to provide a coherent understanding of how POL-on functions as both a data repository and a strategic tool for governance, evaluation, and public accountability.

The chapter begins by tracing the historical and institutional context that led to the creation of POL-on, highlighting the sociopolitical and technological drivers behind its development. It then explores the system's modular structure, charting its evolution from a monolithic architecture to a distributed ecosystem of microservices. That section pays special attention to the system's interoperability mechanisms, which enable seamless data exchange with other government platforms and international standards. Subsequent sections delve into the key data domains that POL-on manages, the legal and ethical safeguards that surround personal data, and the business processes that POL-on supports—from student registration and academic staff management to scientific evaluation and financial reporting. The chapter also addresses POL-on's broader impact on the social environment, illustrating how the system contributes to transparency, policy-making, and public trust in the Polish higher education sector. By presenting the POL-on system not merely as a technical solution, but as a dynamic instrument of public governance, this chapter sets the stage for the more detailed architectural and methodological discussions that follow in Chapter 3.

2.1. The genesis of POL-on

The Polish science and higher education sector is notably diverse and extensive. In 2010, immediately before the launch of the POL-on system, Poland had over 460 higher education institutions, including more than 330 nonpublic universities and over 120 scientific research institutions. This high institutional density, particularly in the nonpublic sector, is a legacy of both postwar restructuring and the socioeconomic transformation following the fall of communism. During the 1990s and early 2000s, the liberalisation of the education market and the growing demand for higher education degrees, driven by aspirations for improved social mobility and labour market competitiveness, led to a rapid expansion of the sector.

The higher education landscape in Poland at that time was highly fragmented. Institutions operated with considerable autonomy and limited regulatory oversight, particularly in terms of internal governance models and data reporting standards. This lack of standardisation hindered the integration of data across the sector, making it difficult for policymakers to access reliable, up-to-date, and comparable information. As a result, strategic planning and evidence-based decision-making at the national level were constrained by inconsistent and incomplete data flows.

The creation of POL-on addressed this gap by introducing a centralised, legally mandated platform for collecting, validating, and analysing data from all higher education and research institutions in Poland. The system became a cornerstone of the country's science and education policy infrastructure, enabling transparency, accountability, and data-driven governance.

POL-on was developed between 2010 and 2014 as a result of the Creation of an information system on higher education project, cofinanced by the European Union under the Human Capital Operational Programme (Priority IV, Action 4, Sub-action 4.1.3). The project leader was the Polish Ministry of Science and Higher Education (MNiSW). The project was implemented by three contractors: the National Information Processing Institute (OPI PIB), Index Copernicus International (ICI), and the Interdisciplinary Centre for Mathematical and Computational Modelling of the University of Warsaw (ICM UW).

The name POL-on is a reference to the chemical element discovered by the famous Polish scientist Maria Skłodowska-Curie[51]. Simultaneously, it is a combination of the words 'POL' (Poland) and 'on' (to switch on).

POL-on is a natural successor to the Polish Science Knowledge Base (BWNP), which was created at OPI PIB in 1998 and maintained as a continuation of the *Nauka Polska* database, which had existed since 1990 and was also freely available on the internet from 1999. In 2018, the BWNP database ceased to exist and its resources were taken over by POL-on.

[51] The year 2011, in which the system was created, was dedicated to the same scientist

Thus, POL-on contains data that was entered before the system was officially established[52], with data originating from BWNP treated in POL-on as archival.

Ultimately, OPI PIB was the author of the modules[53]: Units; Register of Nonpublic Higher Education Institutions; Academic Staff; Study Programmes; Students, Graduates, Doctoral Candidates; Financial Aid; Financial Reports; Operational and Financial Plans; Higher Education Subsidies; Statutory Subsidies; Real Estate and Infrastructure; Laboratories and Equipment; Libraries; Investments; Patents and Achievements; Scientific Projects; Science Popularisation Activities.

Index Copernicus developed the subsystems the Scientific and Technological Journal Evaluation System and the first version of the System for the Evaluation of Scientific Achievements (SEDN). The Interdisciplinary Centre for Mathematical and Computational Modelling created the subsystems the Polish Scientific Bibliography (PBN), the Polish Citation Index (POL-index), and the Polish National Repository of Theses (ORPPD).

The aim of POL-on was to support the work of the MNiSW and other government ministries, as well as Statistics Poland and the Polish Central Commission for Degrees and Titles (CK), by collecting information (both individual-level and aggregated) that concerns Polish science in a broad sense.

Initially, POL-on chiefly contained data related to higher education[54] and was based on the amendment to the Higher Education Act published on 18 March 2011 ([70]), and on implementing acts based on that law, including regulations on: the professional titles awarded to graduates, the conditions for issuing diplomas and certificates of postgraduate studies, and a diploma supplement template ([56]); as well as data included in the national student register ([58]), on the central register of academic teachers and researchers ([59]), and others.

In 2012 and 2013, POL-on was expanded with modules for reporting by scientific units on their scientific output and research potential. Several modules, such as Real Estate and Infrastructure, Laboratories and Equipment, Libraries, Investments, Patents and Achievements, Scientific Projects, and Science Popularisation Activities were created. Modules related to the financing of scientific units were also developed, including Financial Reports, Operational and Financial Plans, Higher Education Subsidies, and Statutory Subsidies. The rest of the system was rebuilt gradually in line with changes in legal regulations (e.g. [60]) and decisions of staff members of the MNiSW who interpret the provisions of legal acts.

[52] By the Higher Education Act ([70])
[53] The modules that constitute the main part of the POL-on system
[54] In the first development phase, the system included: a register of nonpublic universities, a list of higher education and scientific units, a list of study programmes offered by universities, a list of students, and a central register of academic teachers and researchers

POL-on supported the elections to the Central Commission for Degrees and Titles in 2012—the first fully electronic elections in Poland[55]. The system also supported the evaluation processes of scientific units conducted in 2013 (and on the same principles in subsequent years), serving as a container for transferring data from another program[56].

In 2014, the EU-cofinanced project for POL-on's implementation was completed. In the same year, the first version POL-on was fully launched and maintained.

In 2015 and 2016, POL-on underwent significant reconstruction. The scope of data handled by the system was expanded. This related to the publication in January 2015 of the amended Science Funding Act ([71]) and, in June 2015, the regulation on the Science Information System ([61]).

During those years, the PBN, POL-index, ORPPD[57], and SEDN systems were taken over for maintenance and development by OPI PIB.

OPI PIB was established in 1990. Its key role in creating the BWNP, informatising the complex process of science funding (OSF), building the Reviewer Selection Support System (SWWR) based on advanced reviewer recommendation algorithms (via text mining), and its support in evaluation processes made OPI PIB one of the leading institutions responsible for implementing IT solutions for science in Poland.

In recognition of OPI PIB's contribution to the initial phase of POL-on's development and its growing role in informatising science and higher education, the ministry decided to formally legitimise its role in maintaining and expanding the system. The institute's role as the main technical operator was included in the amended law ([71, Art. 4c para. 5]).

Moreover, OPI PIB had a well-developed IT infrastructure and a workforce that was capable of maintaining, rebuilding, and expanding POL-on at the required high level.

POL-on also supported the elections to the Polish Central Commission for Degrees and Titles held in 2016, when the procedures that support the electoral process were modernised. More on the 2016 elections can be found in a 2019 presentation at the European University Information Systems Organization (EUNIS), which was subsequently published in the European Journal of Higher Education IT ([50]).

In 2017 (and on the same principles in subsequent years), evaluations of the scientific output of research units were conducted, using only data entered into POL-on[58]. The system was also adapted to the needs of public statistics (integration with Statis-

[55] Legal details can be found in [57]
[56] A dedicated program for evaluating scientific units
[57] The National Repository of Theses created by ICM was later transformed into the National Repository of Written Theses (ORPPD)
[58] This decision was made to avoid discrepancies between data reported in a unit's survey and in POL-on

tics Poland) under the Integration of the Higher Education Information System (POL-on) with the Central Statistical Office project, financed by the Human Capital Operational Programme. Since 2017, POL-on has been the sole official source of data for university reporting to Statistics Poland and the official source of data for the higher education subsidy allocation algorithm.

Ultimately, the first version of POL-on included nearly forty modules dedicated to reporting and processing data on science and higher education in Poland. The system has always been based fully on applicable legal regulations. At every stage of system development, users—including staff of universities and other scientific units, as well as employees of the MNiSW—have been supported extensively by: dedicated help systems, helpdesk teams[59], responses to user inquiries and system issues, training systems, and instructional videos for selected system modules that are published on help pages.

In 2018, POL-on was awarded the EUNIS Elite Award for Excellence in the category of IT systems in higher education in Europe. Almost every year since POL-on (or similar systems) has been presented in EUNIS postconference materials: [50], [38], [10] or [47].

The changes introduced in the 2018 Act ([72]) necessitated the POL-on system to be rewritten 'from scratch'—switching from a relational database to a microservices model. A number of decisions were made to 'phase out' (i.e. not to transfer to the new application) unused modules. An example of this is the Academic Staff Development module, which has been transformed into the Document Database for Promotion Procedures module in the updated system. The remaining modules were 'rewritten'. As a result of these changes, POL-on has evolved into its current form.

2.2. An interoperable network of services and information

Cross-sectoral cooperation is crucial not only in a private business, but also in the management of the higher education and science sector. It enables better alignment of education and research with the real needs of the labour market and social development. Such cooperation enables the creation of development strategies based on dialogue between universities and public administration. As a result, decision-making processes become more effective, and grounded in data and expert knowledge from various fields. Such partnerships also promote the optimal use of resources, both human and financial. Higher education and science can play an active role in the innovative and sustainable development of a country.

Effective collaboration requires the rapid and secure exchange of information. This capability is essential in achieving success across all sectors. To enable such processes, organisations must be equipped with modern tools that ensure reliable and safe data

[59] Including the core team deployed at OPI PIB

sharing. Experts at OPI PIB have been developing the POL-on system in accordance with the latest trends in digital interoperability. For several years, POL-on has served not only as a comprehensive data repository, but also as a flexible ecosystem of integrated services [40].

POL-on operates as the central information system of the MNiSW. It is designed to gather, validate, and disseminate structured data across a wide array of public and institutional stakeholders. Its interoperability architecture enables bidirectional communication with other government platforms, including those operated by the Statistics Poland, the Ministry of Family and Social Policy, the Social Insurance Institution (ZUS), and many others [39]. The system plays a vital role in reducing administrative burden by enabling institutions to report data once and reuse it across multiple regulatory and analytical contexts.

The scope of the data contained in POL-on is best illustrated by the numbers. The system comprises nearly forty modules and subsystems, which are continuously developed by experts at OPI PIB [39]. It processes over 11 TB of data, which is updated on an ongoing basis. POL-on contains information on over 4.6 million active students and over 9 million students who have enrolled at Polish universities since 2011. In addition, the system holds data on more than 14,000 students enrolled at doctoral schools, and nearly 8,000 doctoral candidates who study under the so-called 'old system'. It also includes 3.9 million diploma theses, over 3 million scientific publications, and information on more than 98,000 academic staff members[60]. Moreover, POL-on supports the annual allocation of approximately 10 billion PLN in public funding.

2.2.1. What exactly is interoperability?

In today's landscape of digital governance, interoperability—defined as the seamless exchange and integration of data across independently managed systems—has become a pillar of effective, transparent, and accountable public administration. The study of interoperability *per se* is highly domain-specific and is used in many different fields to describe a variety of system characteristics. Digital interoperability describes the ability of different IT systems, applications, and devices to work together efficiently, exchanging and using data consistently without the need for human intervention. This means that systems can communicate with each other and interpret the information exchanged—even if they originate from different providers or operate in distinct technological environments. Systems that are fully interoperable are experienced by the relevant group of stakeholders as a single 'integrated' system [3].

In the case of national IT systems in science and higher education like POL-on, interoperability is not only a technical feature, but also a strategic instrument that enables

[60] polon.nauka.gov.pl/siec-polon (accessed 17 November 2025)

the coordination of activities among various entities in the public sector. For example, data entered into POL-on by a university is disseminated automatically to support national statistics, funding applications, the processing of student benefits, and institutional evaluation. Importantly, POL-on's integration capabilities are embedded in national legislation.

The Law on Higher Education and Science of 20 July 2018 ([72]) mandates the electronic submission of institutional data through POL-on, making the system the legal backbone of the digital information infrastructure for the Polish science and higher education sector. POL-on's role is further reinforced by the Polish government's strategic documents on public digitalisation, such as the Programme for Integrated Informatization of the State[61] and the Digital Competences Development Strategy[62]. The implementation of interoperability in POL-on also aligns with broader European Union objectives under the Digital Europe Programme[63], particularly in areas that concern open science, data standardisation, and cross-border digital services. By providing common interfaces and APIs, POL-on enables real-time data exchange with systems like ORCID, OpenAIRE, and national Current Research Information Systems (CRISs), which supports the international visibility and comparability of Polish research outputs.

As a result, POL-on has transformed from a standalone database into an advanced, multilayered ecosystem of interoperable services and registries. It serves as the foundation for core academic administration processes, public policy, and digital governance, setting a benchmark for other national platforms in terms of its scalability and regulatory compliance.

Due to the interoperability of the POL-on system, it is possible for many important cross-sectoral activities to be performed. When a social need arises—often later formalised in new legislation—experts at OPI PIB update POL-on accordingly to fulfil the given task. The most recent example that stemmed from legislative changes is the introduction of the new Foreigners module. Since 1 July 2025, students and doctoral candidates without Polish citizenship who have been admitted to a study programme or doctoral school must be registered. The Foreigners module is the result of provisions in the Act of 4 April 2025 amending certain laws, aimed at eliminating irregularities in the visa system of the Republic of Poland ([73]) which grants the Minister of Foreign Affairs and consuls access to specified data in POL-on. The change was introduced to facilitate proceedings regarding the issuance, revocation, or annulment of Polish visas.

The obligation to register foreign students in POL-on supports the work of consular services in Poland by providing them with quick access to reliable data on student visa applicants. This enables consuls to verify whether an applicant has indeed been admitted to a legally operating university. The system also enables the verification of a

[61] gov.pl/web/cyfryzacja/program-zintegrowanej-informatyzacji-panstwa (accessed 17 November 2025)
[62] gov.pl/web/cyfryzacja/strategia-cyfryzacji-polski-do-2035-roku (accessed 17 November 2025)
[63] digital-strategy.ec.europa.eu/pl/activities/digital-programme (accessed 17 November 2025)

student's current status, such as whether they are continuing their studies, have been removed from the register, or are on a dean's leave of absence. The information contained in POL-on helps to prevent visa fraud and false recruitment. Based on this data, consuls can make informed decisions to grant or refuse visas. Registration also enables the monitoring of migration trends related to education and the suitable adjustment of visa policies. The system supports cooperation between consulates, universities, and public administrators. It also facilitates control of the legality of foreigners' stays in Poland. The enhanced interoperability of the POL-on system enables improved security and the monitoring of migration management, as well as increasing transparency and the efficiency of services provided to international students.

Improving the visa process is not the end of the development of POL-on's system interoperability. The experts at OPI PIB now face a new challenge: the introduction of the e-Diploma service, on which Intensive development and legislative work is currently underway (as of August 2025). The project [48] aims to build and implement a central, national system for the comprehensive management of electronic diplomas for completed studies, doctoral degrees, and habilitation degrees, which will ensure the integrity, consistency, and reliability of the documents issued by Polish universities. The project requires cooperation among many partners. According to ongoing legislative work, the e-Diploma service is to be available to every graduate from 1 January 2027. The project targets several important groups. Graduates and those pursuing doctoral and habilitation degrees will benefit, as they will have easier access to their documents without the need for in-person collection. The system will also assist universities and other scientific institutions in issuing diplomas more efficiently. Employers will gain the ability to verify candidates' educational qualifications quickly and reliably. Professional licensing organisations that need to verify diplomas will also benefit from the system. Moreover, it will serve as a tool for the governmental ministries that oversee higher education in Poland, enabling them to better control the documents issued. The project's objective is to transition from issuing documents exclusively in paper form to issuing them in an electronic format that is convenient for the public and accessible from anywhere via a web browser and the *mObywatel*[64] application. Therefore, POL-on will need to be connected securely with other systems and applications. Without the data contained in POL-on, it would not be possible to launch the e-Diploma service at all universities and doctoral schools across Poland in 2027.

The POL-on ecosystem is not a single system but a network of interconnected registries and services that collectively support science and higher education in Poland (Figure 2.1.). It consists of:

- **POL-on** (a central system that contains the main registries)[65]: the primary system with which all other applications communicate

[64] info.mobywatel.gov.pl (accessed 17 November 2025)
[65] polon.nauka.gov.pl (accessed 17 November 2025)

- **Polish Scientific Bibliography (PBN)**[66]: a system that contains data on the publication achievements of Polish researchers
- **Scientific Output Evaluation System (SEDN)**[67]: a system that supports the evaluation process of scientific units under the MNiSW, using data from POL-on and the PBN
- **Polish Citation Index (POL-Index)**: a system that collected information on articles published in scientific journals until 2015, aimed at calculating the so-called Polish Impact Factor[68]. The system was discontinued in 2020
- **National Repository of Written Diploma Theses (ORPPD)**[69]: a repository that stores all written diploma theses of Polish graduates since 2009
- **Polish Graduate Tracking System (ELA)**[70]: a portal that presents the annual survey results on the employment circumstances of students, graduates, and doctoral candidates. The survey is based on anonymised data from POL-on and Statistics Poland. The results are presented as interactive reports and infographics
- ***ELA Uczeń***[71]: an interactive information tool dedicated to students, their parents, and career advisors that offers reliable data on the labour market circumstances of graduates of Polish universities quickly and easily
- **Electronic Voting System for the Central Commission for Academic Degrees and Titles**[72]: a system that manages the electronic election process among Polish researchers to the key body responsible for awarding and supervising the quality of academic degrees and titles in the country[73]
- **People of Science (*Ludzie Nauki*)**[74]: a modern portal which, in March 2024, replaced the older database *Nauka Polska*. It is the largest knowledge base on Polish scientific achievements, holding data on over 200,000 Polish scientists and more than one million scientific achievements in one place. The portal features an advanced search engine that enables searches by name, academic degree and title, field, discipline, or specialisation, as well as keywords that appear in publication abstracts, patent descriptions, protective rights, research projects, and artistic output. It enables fast assessment of the significance of results through contextual keyword reading
- **RAD-on**[75]: a publicly accessible portal that serves as a source of reports, analyses, and data on higher education and science in Poland, sourced exclusively from trusted institutions
- ***Studnia***[76]: a system used to collect evidence of the impact of scientific activities on society and the economy.

66 pbn.nauka.gov.pl (accessed 17 November 2025)
67 sedn.opi.org.pl (accessed 17 November 2025)
68 http://pci-www.vls.icm.edu.pl/polindex/info/polski-wspolczynnik-wplywu (accessed 17 November 2025)
69 polon.nauka.gov.pl/orpd (accessed 17 November 2025)
70 ela.nauka.gov.pl (accessed 17 November 2025)
71 ela-uczen.nauka.gov.pl (accessed 17 November 2025)
72 rdn.gov.pl (accessed 17 November 2025)
73 In 2019, the system was updated following the establishment of the Council of Scientific Excellence, for which an electronic election process was also implemented. This service is one of the recurring processes supported by the POL-on system
74 ludzie.nauka.gov.pl (accessed 17 November 2025)
75 radon.nauka.gov.pl (accessed 17 November 2025)
76 studnia.opi.org.pl (accessed 17 November 2025)

Additionally, POL-on cooperates with or integrates the following services:

- ***Studia.gov.pl***[77]: a search engine that presents study programmes and information for prospective students. The tool presents the current educational offer of universities across Poland and outlines admission rules for all study programmes nationwide
- **Inventorum**[78]: a system that recommends information in science and business. The database partly uses information from POL-on and People of Science, regarding data on projects, innovative enterprises, scientific institutions, experts, and conferences
- **Unified Anti-Plagiarism System (JSA)**[79]: a modern tool created in response to the needs of the Polish academic staff members, which has organised the antiplagiarism market in Poland. Previously, many programs existed based on different algorithms and reference databases. Since its launch in 2019, all engineering, bachelor's, master's, and doctoral theses have been analysed using this same tool. Since February 2024, every institution that conducts higher education or doctoral study programmes can use JSA free of charge to check whether a diploma thesis was written with the use of artificial intelligence
- **OSF**[80]: a system that manages funding streams and exchanges information with POL-on on science and higher education institutions, initiated research projects, and publications [5]
- **ORCID**[81]: a system developed as part of a project to standardise data on researchers' achievements
- ***Empatia***[82]: a government portal for registering applications for social benefits, where student status is considered. Information is exchanged with POL-on
- **REGON**[83]: a business entity register maintained by Statistics Poland, which communicates with the POL-on registry of science and higher education institutions
- **TERYT**[84]: a territorial division dictionary maintained by Statistics Poland and used throughout the country
- **Login.gov.pl**[85]: a single-sign-on solution maintained by the Polish government for state domain systems. It is an authentication component of the state's IT architecture and, at the time of publication, is integrated with the RAD-on system for citizen data services, presenting detailed information on individuals processed in source systems, primarily POL-on.

77 studia.gov.pl/sfos/r/wsapp/wyszukiwarka-studiow/home (accessed 17 November 2025)
78 inventorum.opi.org.pl (accessed 03 August 2025)
79 jsa.opi.org.pl (accessed 17 November 2025)
80 osf.opi.org.pl (accessed 17 November 2025)
81 orcid.org (accessed 17 November 2025)
82 empatia.gov.pl (accessed 17 November 2025)
83 wyszukiwarkaregon.stat.gov.pl/appBIR/index.aspx (accessed 17 November 2025)
84 eteryt.stat.gov.pl (accessed 17 November 2025)
85 gov.pl/web/login (accessed 17 November 2025)

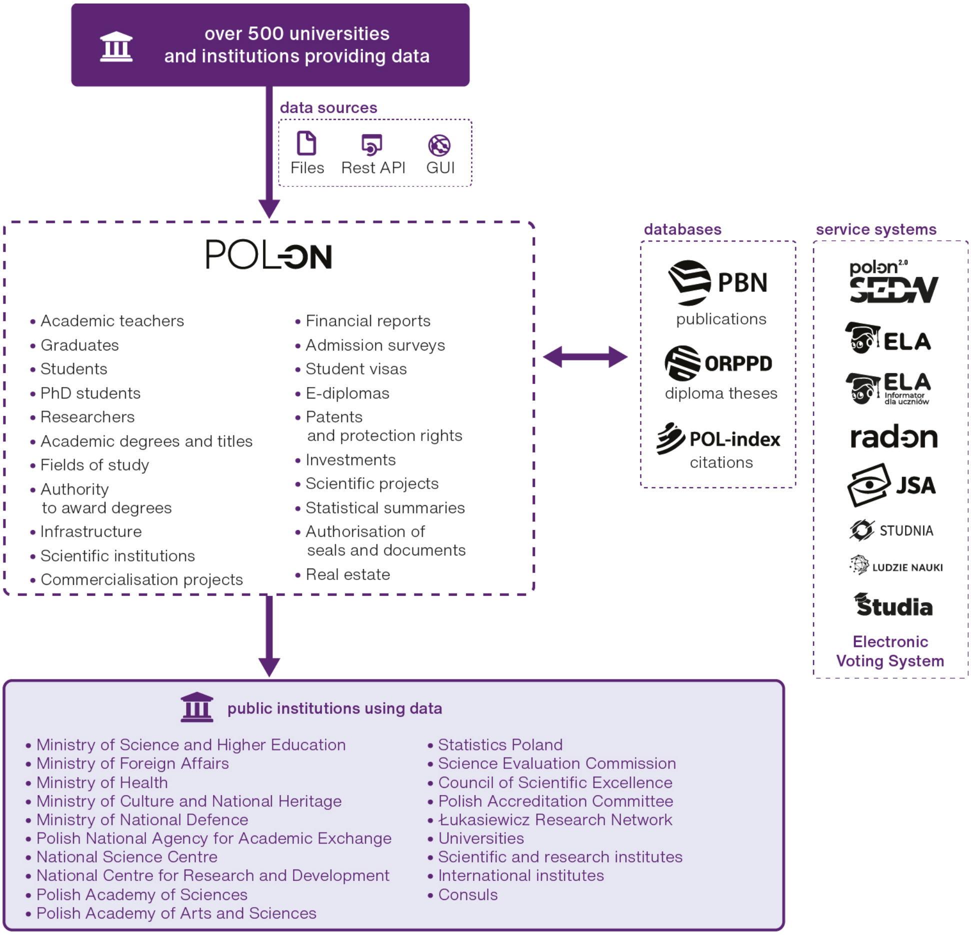

Figure 2.1. The structure of the POL-on ecosystem

2.3. Key data domains

The data is provided by all universities, research institutes, and scientific institutions in Poland, of which there are over 500. This is required by Polish law. All of those institutions are also responsible for keeping the data up to date.

For an institution to fulfil its reporting obligations in POL-on, it must be able to access the system.

Changes introduced in legal acts have resulted in changes to the content of the IT system that processes data on higher education and science in Poland. Currently, this system (in accordance with Article 342(3) of the Law on Higher Education and Science [72]) must include:

- a list of academic teachers, other individuals who conduct classes, individuals who conduct scientific activities, and individuals who participate in their conduct
- a list of foreigners admitted to studies and doctoral schools
- a list of students
- a list of doctoral degree applicants
- a list of institutions in the higher education and science system
- a repository of written theses
- a database of documents in promotion proceedings
- a database of individuals authorised to sign documents
- a database of planning and reporting documents
- a repository of electronic diplomas (from 1 January 2026).

A detailed list of data to be included in any such system is specified in the act and in the Regulation on data processed in POL-on [62] issued on its basis.

As a result, the following data must be collected in POL-on:

- for academic teachers, other individuals who conduct classes, individuals who conduct scientific activities, and individuals who participate in their conduct, in addition to personal, identification, and employment data, the system should contain data on professional titles, degrees and academic titles, classes conducted by a given individual both at a university and at a doctoral school, the scientific disciplines in which such individuals conduct scientific research, and the time spent on scientific activities in particular disciplines, their classification by the employing institution as individuals who conduct scientific research, information on the skills and experience that enables them to properly conduct classes in a study programme (in the case of other individuals who conduct classes), data on the doctoral school where an individual conducts classes, management functions performed at higher education institutions, disciplinary penalties imposed on employees, and their scientific and artistic achievements (authorship or coauthorship of patents, protective rights, and exclusive rights to plant protection, authorship or coauthorship of scientific articles, monographs and chapters in monographs, editing scientific monographs, and authorship or coauthorship of artistic achievements)
- for the list of foreigners admitted to studies and doctoral schools, in addition to personal and identification data, information on admission to studies or doctoral schools and information on whether the foreigner holds a Pole's Card should be included
- for students, in addition to personal and identification data and the level of advancement of their fields of study, such a system must collect information on their place of residence before the commencement of their studies, and the type and period of social benefits and ministerial scholarships granted to them
- for doctoral candidates, in addition to personal and identification data, the system must contain information on the mode of preparation of the doctoral dissertation, information on the doctoral school (in the case of doctoral students, information on the doctoral degree awarded, the scientific achievements of the individual (authorship or coauthorship of patents, protective rights, and exclusive plant protection rights; authorship or coauthorship of scientific articles, monographs and chapters in mono-

graphs; editing of scientific monographs; and authorship or coauthorship of artistic achievements), the first and last name of the supervisor and their PESEL number (in the absence thereof, the number of the identity document and the name of the country that issued it, as well as the place of employment, and in the case of a foreign doctoral student, information on: the name of the country of birth, admission to the doctoral school and education there, and possession of a Pole's Card

- for institutions of higher education and science, in addition to the identification data of an entity and address data (including the history of the institution's transformations) and information on the individuals who manage that entity, the list should include information on the conducting of activities outside the registered office, scientific categories, scientific activities, permits to establish studies, studies conducted, fees charged to students, specialised education provided, cost-effectiveness ratios, doctoral schools run, bodies that award academic degrees and degrees in the arts, scientific and research equipment, and IT infrastructure (with a value exceeding 500,000 PLN), investments, expenditure on scientific research and development work, sources of the origin of funds and financial results, as well as revenues from the commercialisation of scientific results or know-how related to such results, as well as information on whether the entity conducts classified research and information on scientific research conducted by the entity's employees, the impact of the institution's scientific activities on the functioning of society and the economy, and the projects it conducts involving scientific research, development work, or the dissemination of science
- for the repository of written theses, metadata on the thesis itself and its author, information on the thesis supervisor(s) and reviewers, and on the institution awarding the professional title and the student's field of study, as well as information on the diploma examination that concluded the thesis procedure, must be held
- for the database of documents in promotion proceedings, the personal and identification data of an individual applying for a degree or academic title must be provided, as well as the data and opinions of reviewers (and in the case of a postdoctoral degree, the data of the members of the postdoctoral committee), information on the award or revocation of a degree/title; in the case of individuals applying for a doctoral degree, also the content of the doctoral dissertation together with its summary and date of submission, as well as reviews and dates of their preparation; and in the case of a postdoctoral degree or professorship, the application for the degree/title together with the date of its submission; in the case of individuals applying for the title of professor, also the opinions of reviewers regarding the fulfilment of requirements and the dates of their preparation
- for the database of individuals authorised to sign documents at scientific institutions, information on the individual (personal data), information on the function performed by that individual, a digital reproduction of the individual's signature and initials, and the details of the qualified electronic signature certificate (certificate from 01.01.2026), as well as information about the official seal template and information about the templates: graduation diplomas, doctoral diplomas, and postdoctoral diplomas, copies of those documents, including copies in foreign languages, postgraduate diplomas, and certificates of completion of studies and postgraduate studies, or certificates of the awarding of scientific degrees or degree in the arts

- for the database of planning and reporting documents of scientific institutions in Poland, the following are provided: material and financial plans (of public universities and the CMKP), annual financial plans of the Łukasiewicz Center and the Łukasiewicz Network institutes, reports on the implementation of material and financial plans of public universities and the CMKP, reports and statements on the use of funds described in Article 365 of the Act, annual financial reports of public universities and the CMKP, as well as annual financial reports of the Łukasiewicz Centre and the Łukasiewicz Network institutes
- for the electronic diploma repository (from 1 January 2026), in addition to the personal and identification data of an individual who completed their studies, obtained a doctoral degree or a postdoctoral degree, the necessary elements of the diploma of completion of studies and the data contained in the diploma supplement template must also be entered, additional elements of the diploma of completion of studies, specified by the university, necessary elements of the doctoral diploma and postdoctoral diploma, and additional elements of the doctoral diploma and postdoctoral diploma, specified by the higher education and science system entity that issued the diploma.

2.4. Supported business processes

The POL-on system collects data of institutions of the Polish higher education and science system, including higher education institutions, federations of entities of the higher education and science system, scientific institutes of the Polish Academy of Sciences, research institutes, international scientific institutes, the Łukasiewicz Centre, Institutes of the Łukasiewicz Research Network of other entities that conduct chiefly scientific activities in a separate and continuous manner. These institutions register the data of academic teachers, research workers, students, and applicants for degrees and titles, respectively, in POL-on. In addition, they feed documents to the repository of written dissertations and to the database of documents in promotion proceedings, as well as submitting reports and reports in the planning and reporting documents database (the scope of which is detailed in chapter 2).

The set of data collected in POL-on is defined strictly by Polish law. The data is used by the MNiSW to supervise the system of higher education and science. That data is also used by the Polish National Agency for Academic Exchange (NAWA)[86], the Polish National Centre for Research and Development (NCBR)[87] and the Polish National

[86] NAWA was established on 1 October 2017 to coordinate state activities that drive the internationalisation of Polish academic and research institutions. NAWA's mission is to foster the development of Poland in science and higher education.

[87] The NCBR is an executive agency within the meaning of the Act of 27 August 2009 on Public Finance, operating on the basis of the Act of 30 April 2010 on the National Centre for Research and Development and the statute attached to the regulation of the Minister of Science and Higher Education of 9 September 2010 on the statute of the National Centre for Research and Development. The functioning of the NCBR also regulates a number of executive and legal acts that relate to the implementation of programmes financed from European financial instruments.

Science Centre (NCN)[88] to fulfil their statutory tasks. They support scientific research, technology development, and the mobility of scientists. In particular, they serve to:

- establish and implement the state's science policy
- conduct evaluations of the quality of education, the activity of doctoral schools, and the quality of scientific activity
- conduct proceedings in cases of awarding doctoral degrees, habilitation degrees, and professorships
- determine the size of subsidies and grants.

Among the many business processes supported by POL-on, those that relate to the control of potential fraud deserve special attention. The system makes it possible to detect cases of the noncompliance of data entered in the register with the applicable law. Examples of such irregularities include:

- taking financial aid more than twice in one academic year as defined by polish law
- simultaneous use of financial aid in several institutions
- concurrent study at more than one doctoral school
- multiple employment of the same person as a core staff member at the same time
- representation by a staff member of more than two scientific disciplines at the same time
- the recording of a staff member's key scientific data after a deadline.

In the case of an individual receiving financial aid more than twice in one academic year, the system automatically detects anomalies and either blocks the entry of erroneous data (Figure 2.2.) or, if it allows it to be saved, flags it as requiring further verification or clarification.

[88] The NCN is a government agency, supervised by the MNiSW, established in 2011 to support basic research in Poland. Basic research is defined as empirical or theoretical endeavours undertaken to gain new knowledge of the foundations of phenomena and observable facts, without any direct commercial use. The NCN funds projects in arts, humanities, and social sciences; life sciences and physical sciences; and engineering. It launches funding schemes dedicated to researchers at different stages of their careers.

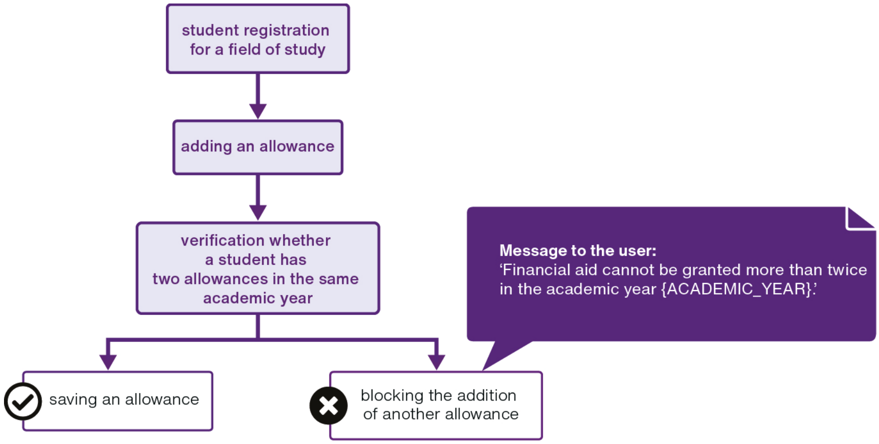

Figure 2.2. Verification of the size of registered financial assistance of the 'allowance' type

Another key element in the functioning of the Polish higher education system is the distribution of public funds, or so-called subsidies. How much funding an institution receives depends on the algorithms set out in the relevant regulations. Those algorithms are based on the data that universities and institutes enter independently into the POL-on system (Figure 2.3.).

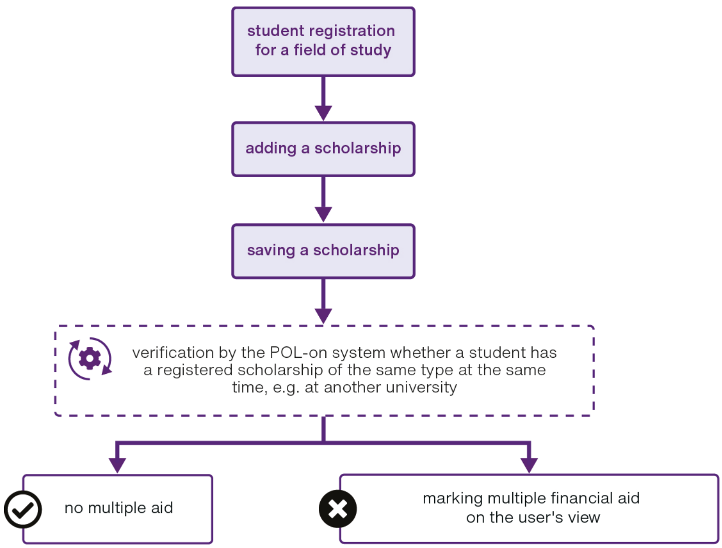

Figure 2.3. Verification of the size of registered financial aid of the 'scholarship' type

On the basis of the same data, but under separate procedures, evaluation of the quality of an institution's activity may also be conducted. This includes the quality of education, the quality of education at a doctoral school, and the quality of scientific activity.

The evaluation of the quality of education is conducted annually by the the Polish Accreditation Committee (PKA)[89] at the request of a university or a government minister.

In the course of conducting a programme evaluation, data is considered from a range of study programmes and learning standards, teaching and research staff, infrastructure used for the delivery of a study programme, cooperation with the socioeconomic environment, internationalisation, and support for students in the learning process.

The first doctoral schools began teaching on 1 October 2019, which means the evaluation of doctoral schools is a relatively new process. In accordance with the provisions of the Act on Higher Education and Science of 20 July 2018, the first evaluations of doctoral schools began in the fourth quarter of 2024. The National Education Commission (KEN)[90] conducts the first evaluation of an institution no earlier than five years after the commencement of teaching at that school, and the next evaluation of a given school may take place no earlier than two years after the KEN resolution on the evaluation becomes final and no less than once every six years.

The assessment of the quality of scientific activity concerns academic universities, institutes of the Polish Academy of Sciences, international institutes, vocational schools, research institutes, and entities that conduct mainly scientific activity in an independent and continuous manner and have their registered offices in the territory of the Republic of Poland. It covers the achievements of all employees who conduct scientific activity. Assessment in this area may be performed upon request, and in other cases every four years. The scientific or artistic level of the activity, the financial results of scientific research and development work, and the impact of scientific activity on the functioning of society and the economy are considered (Figure 2.4.).

[89] The PKA is an independent institution that acts to ensure and improve the quality of education

[90] The KEN operates under the MNiSW. Its tasks include evaluating the quality of scientific activity, and preparing draft lists of publishers of peer-reviewed scientific monographs and lists of scientific journals, as well as peer-reviewed materials from international conferences; submitting proposals to the Polish Minister of Science and Higher Education regarding scientific categories for entities undergoing evaluation; preparing opinions and assessments on matters specified by the minister or on its own initiative; evaluating the quality of education at doctoral schools; preparing analyses of the evaluation of the quality of scientific activity and the quality of education in doctoral schools; and cooperating with national and international institutions involved in the evaluation of the quality of scientific activity.

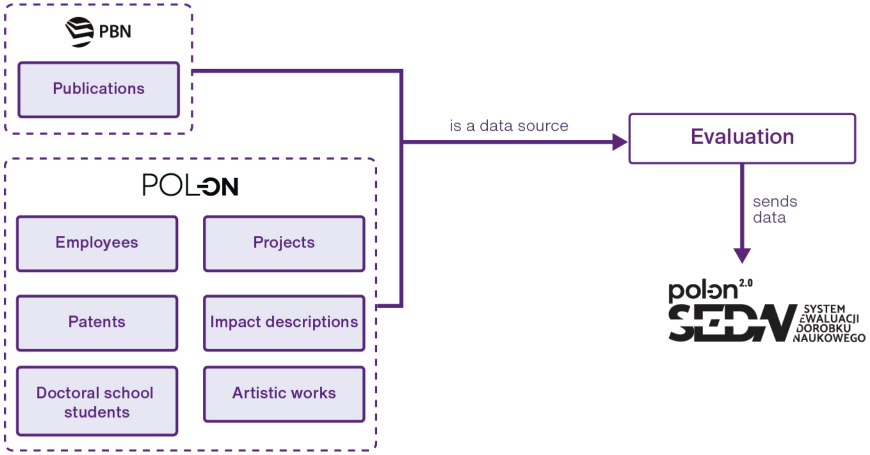

Figure 2.4. Evaluation of entities in terms of the quality of education, doctoral schools, and scientific activity

The financial reporting process is conducted within the Planning and Reporting Documents Database module. It enables the handling of the reporting document submission process and the aggregation of data from selected areas that support decision-making processes in financing. It also enables the storage of archived documents, which are listed below:

- material and financial plans of public universities
- annual financial plans of the Łukasiewicz Centre and the institutes of the Łukasiewicz Research Network
- reports on the implementation of material and financial plans of public universities
- reports and statements on the use of funds, e.g. a report on the use of subsidies for tasks related to ensuring that individuals with disabilities have the conditions to fully participate in the admission process to doctoral schools, education at universities and doctoral schools, or conducting scientific activity; reports on the use of scholarship funds; or annual and final reports on the use of funds for investments related to education and scientific activity
- annual financial statements of public universities
- annual financial statements of the Łukasiewicz Centre and the Łukasiewicz Research Network institutes.

Most of the reports described above are submitted annually, with the exception of final reports, which are submitted for the purpose of accounting for funds that an institution has received from the MNiSW for the implementation of selected activities. The reporting data is entered into POL-on in generated and shared templates defined by the MNiSW. In the case of financial statements, the data is entered in XML format in accordance with the guidelines of the Polish Ministry of Finance, as specified in a published XSD file. After being imported into POL-on, the data is approved by an employee of the entity and a statement of compliance of the data with the true state of affairs is generated. Then, once signed by the university rector or institute director (outside of POL-on), the data is attached when the report is sent to the MNiSW. An employee of the MNiSW who is responsible for reporting verifies the data, then accepts or returns it for correction. After a correction has been made by a university employee, the report form is sent back to the MNiSW via the same path (Figure 2.5.).

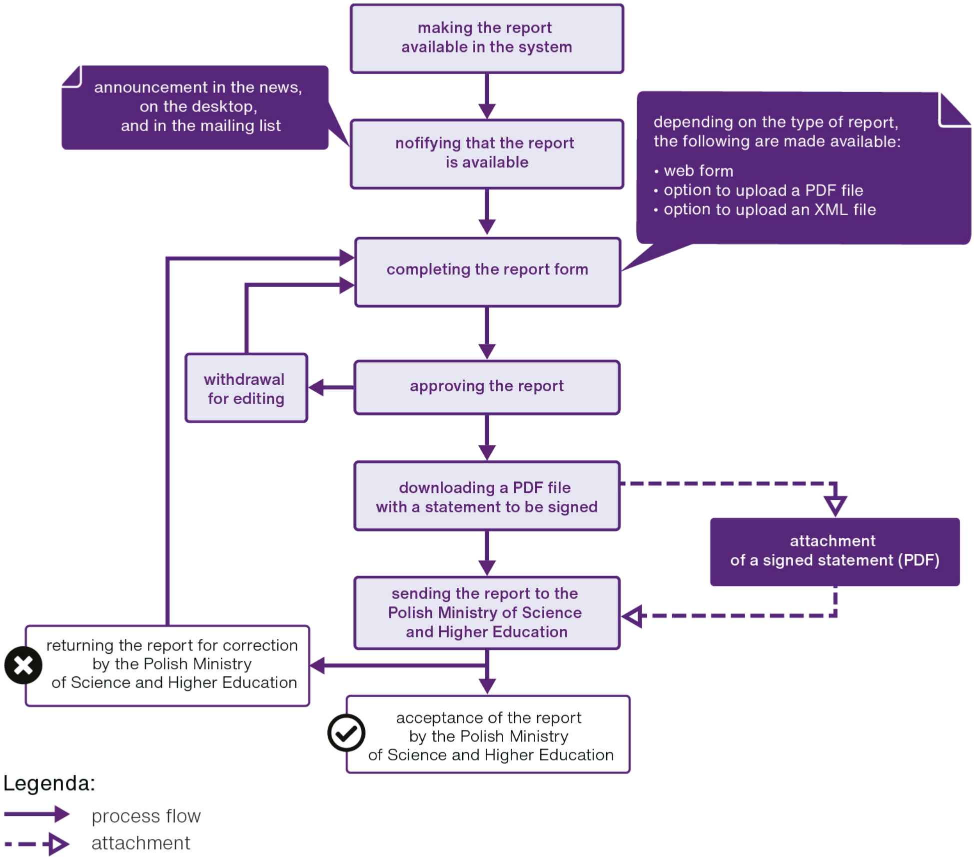

Figure 2.5. The reporting process in POL-on

Reports uploaded to POL-on are archived one year after submission and then stored in the system for ten years.

Another process supported by POL-on is statistical reporting. It simplifies the current process of statistical data submission by universities in the form of paper forms to Statistics Poland; until recently, this was a challenge for higher education institutions. Currently, based on the data entered into POL-on in the appropriate modules, aggregated values are calculated for the S-10[91], S-11[92], S-12[93], and S-M[94] reports. The reports are prefilled, which reduces the burden on users and requires only verification and possible correction of the data in the system by the institutions. In addition, this process eliminates possible errors in data entry by Statistics Poland employees, and thus improves the quality of the data transmitted, which, after approval, is forwarded directly to Statistics Poland (Figure 2.6.).

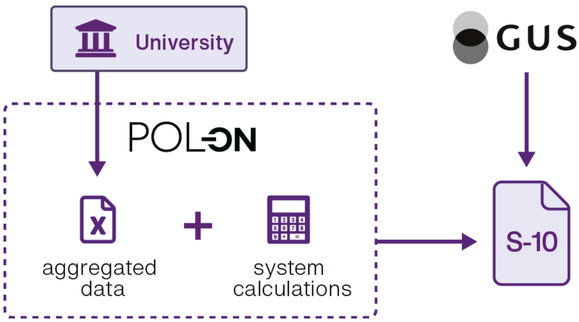

Figure 2.6. The reporting process of Statistics Poland in the POL-on system

The next step in academia after being awarded a master's degree is an application for a doctoral degree. POL-on collects data on both doctoral education and subsequent degree conferral. This is because the promotion process is not only a decision to confer a degree, but also a series of information on the subsequent stages of the academic careers of applicants for degrees or titles, together with the documents that accompany that process. The promotion proceedings database contains a set of information that makes it possible to verify whether and when an individual obtained a degree or title and in which discipline (Figure 2.7.). The content of dissertations included in the promotion proceedings database, like the theses from the ORPPD, is part of the reference collection used by the JSA to review works.

91 Higher education report S-10, S-10-W (for military universities)
92 Report on benefits for students and doctoral students S-11
93 Report on postgraduate studies, specialist training, doctoral students, and employment at universities S-12, S-12-F (for federations), S-12-W (for military universities)
94 Report on the mobility of S-M-POLON graduates

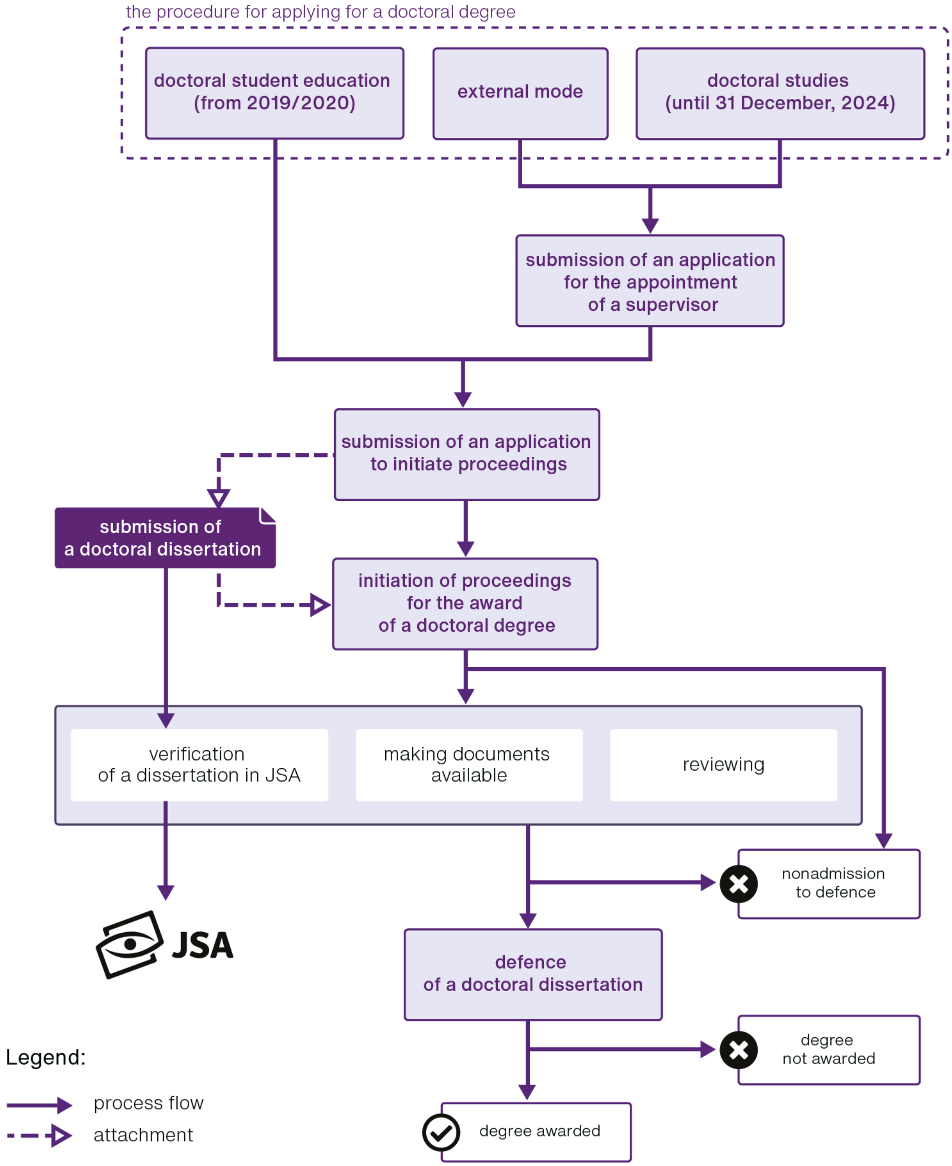

Figure 2.7. The process of applying for a doctoral degree as an example of a promotion process, the data for which is collected in POL-on

In 2024, a system was launched to manage the process of evaluating the quality of education in doctoral schools. The first doctoral schools were established in 2019 in accordance with the Polish Law on Higher Education and Science. Their activities are subject to evaluation after five years. The evaluations are conducted by experts supervised by the KEN and supported by the MNiSW. The submission of self-assessment reports by entities that run doctoral schools and the preparation of evaluation reports with assessments takes place in the SEDok system (Figure 2.8.).

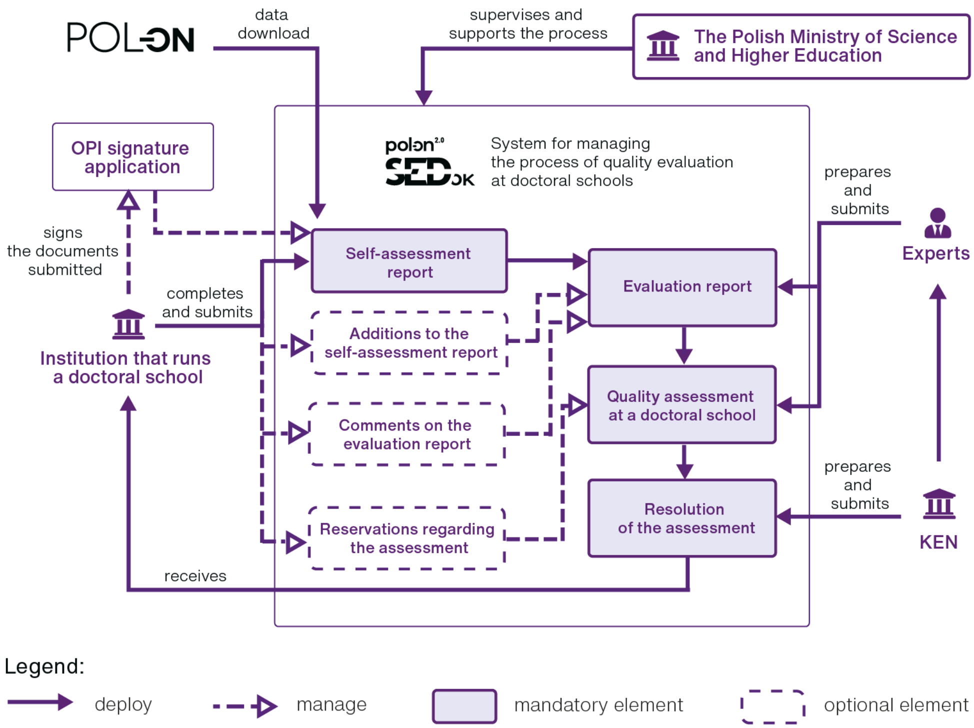

Figure 2.8. The process of evaluating the quality of education at doctoral schools

The database also contains information on any possible revocation of a degree/title, if such a situation has occurred.

POL-on also supports document control. The Polish National Agency for Academic Exchange (NAWA) can issue, through authentication by its director, documents connected with the course or completion of studies, such as:

- diploma of graduation and diploma supplement
- copies of already-issued diplomas, including copies in foreign languages
- duplicates of already-issued diplomas
- certificate of graduation, intended for legal transactions abroad.

The process is based on the comparison between the data on the signatures and functions of individuals who are authorised to sign, the templates of authenticated documents, and the templates of seals stored centrally in the POL-on system with the document actually presented. In addition, the director of NAWA may decide on the education of foreigners, and the POL-on system enables the monitoring of such courses of education.

Work is underway to add a process for verifying identification data stored in the POL-on system with the central PESEL register. When a user registers their data, POL-on will additionally communicate with the PESEL register to confirm whether a person with such identification data exists. If an error is detected, it will prevent that registration or mark the record as incorrect. This measure aims to improve the quality of data collected in POL-on (Figure 2.9.).

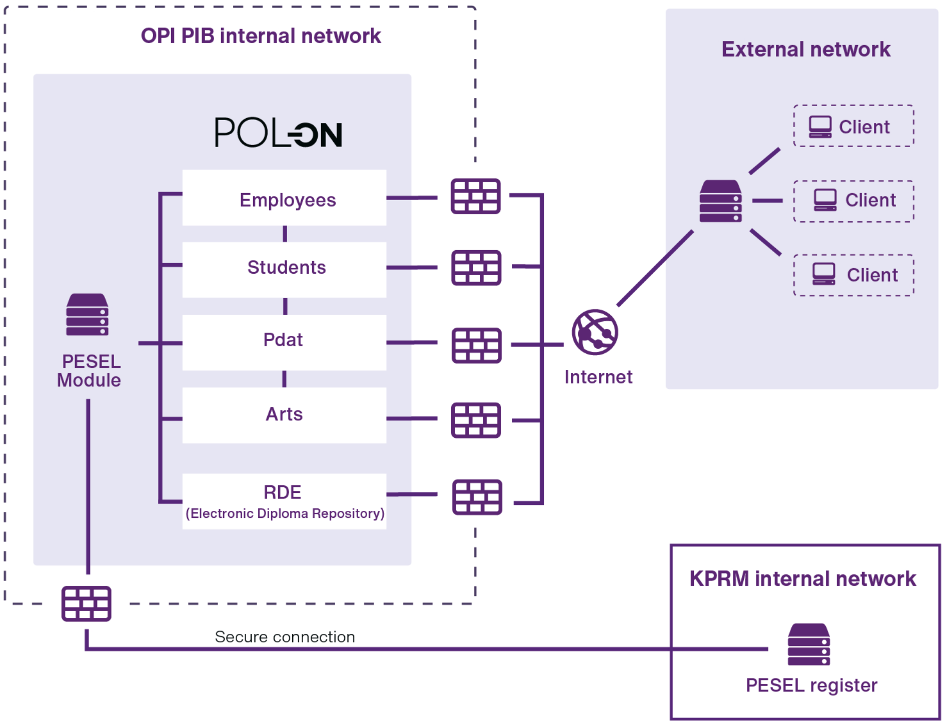

Figure 2.9. PESEL number verification process

This is particularly important in the context of the further development of POL-on in the issuance of electronic documents, such as diplomas or supplements (Figure 2.10.).

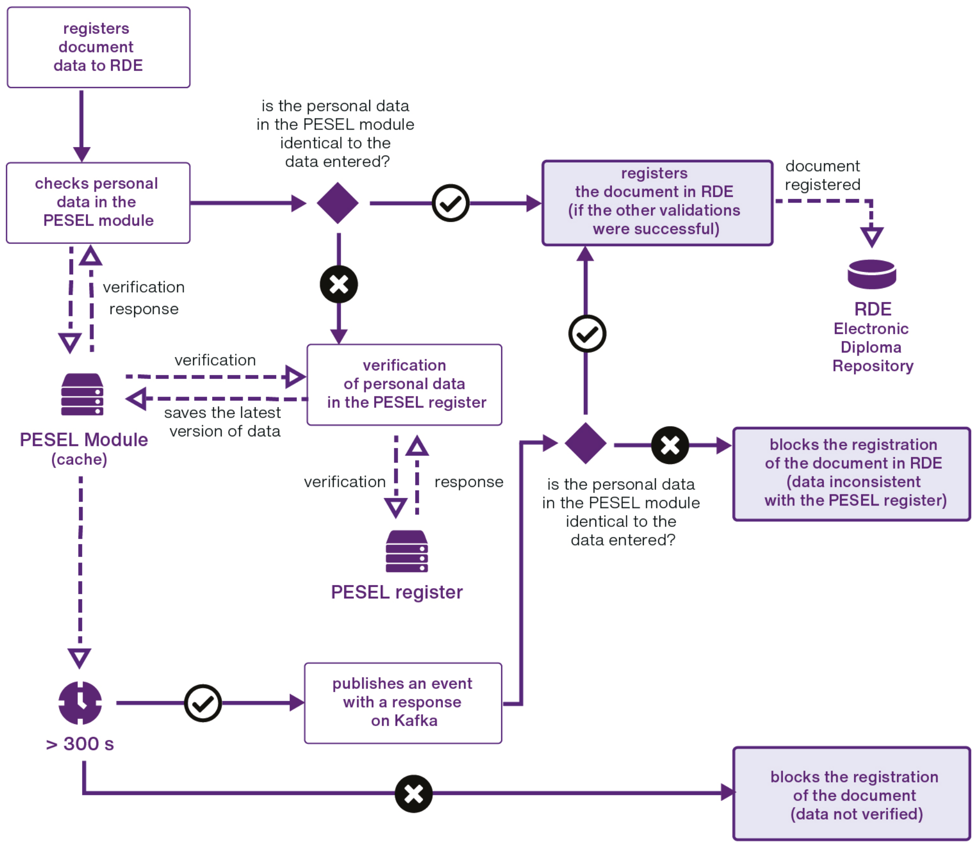

Figure 2.10. Verification of a PESEL number when issuing a diploma

In 2025, the Foreigners module was implemented in POL-on, which enables consuls to verify student visa applicants. Only individuals registered in the POL-on system as foreigners can obtain visas for the purpose of studying at Polish universities (Figure 2.11.).

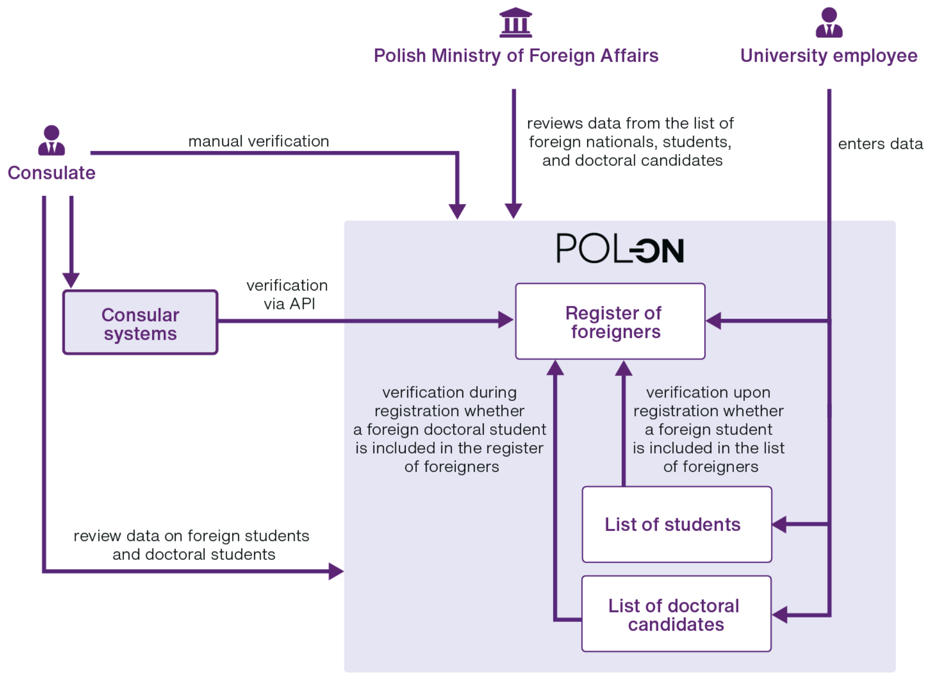

Figure 2.11. The process of verifying a foreigner

Some of the data collected in POL-on is made available publicly to implement the open data policy in public administration.

The aim of the Polish Graduate Tracking System (ELA) is to collect data based on co-operation between the POL-on system and the Polish Social Insurance Institution (ZUS). Based on the collected data, reliable information on the circumstances of Polish university graduates on the labour market is then provided in the form of visualisations or reports (Figure 2.12.).

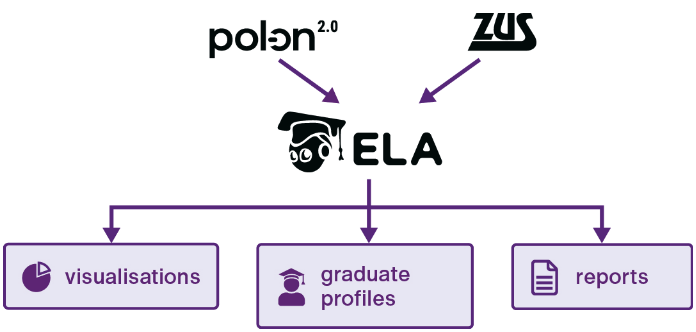

Figure 2.12. The process of collecting data used in the ELA system

The Initiative of Excellence – Research University (IDUB) programme, introduced by the Law on Higher Education and Science, is one of the key tools for reforming science and higher education in Poland. Its main objective is to improve the quality of scientific research and education, as well as strengthening the position of Polish universities internationally. The programme provides for the allocation of additional funds to selected universities that commit to implementing ambitious development strategies. The priority of IDUB is to identify universities capable of competing effectively with the best European and global universities. Another key element of IDUB's operation is the programme management process, which is conducted in the Funding Stream Support System (OSF)[95]. It includes both the submission and consideration of applications, as well as the preparation and evaluation of interim and final reports, which are based on data from POL-on.

In the long term, the IDUB programme aims to create strong research centres in Poland that will drive economic development and innovation, as well as raising the prestige of Polish higher education globally.

The Studia.gov.pl[96] portal has been made available to support the decision-making processes of university applicants. It includes a tool for recommending study programmes to candidates, where users can find verified and reliable information on fields of study. Initially, the tool utilised four sources of information:

- constantly updated data from POL-on
- assessments of study programmes conducted by the PKA
- evaluation of scientific activity conducted by the KEN
- information submitted by universities.

Because the chief purpose of the portal is to support young people in choosing the best field of study for themselves, a number of changes were introduced to the portal in 2023–2025. These include (Figure 2.13.):

- the simplification of the search interface
- the limiting of data sources to:
    - data from POL-on regarding institutions and fields of study
    - data from the recruitment survey completed by universities regarding individual fields of study
    - data from ELA on graduates.

[95] A cutting-edge system dedicated to Polish scientists and entrepreneurs. It is used to register and handle applications for financing R&D projects and studies with funds provided by the MNiSW, the NCN, and the NCBR
[96] studia.gov.pl/sfos/r/wsapp/wyszukiwarka-studiow/home (accessed 17 November 2025)

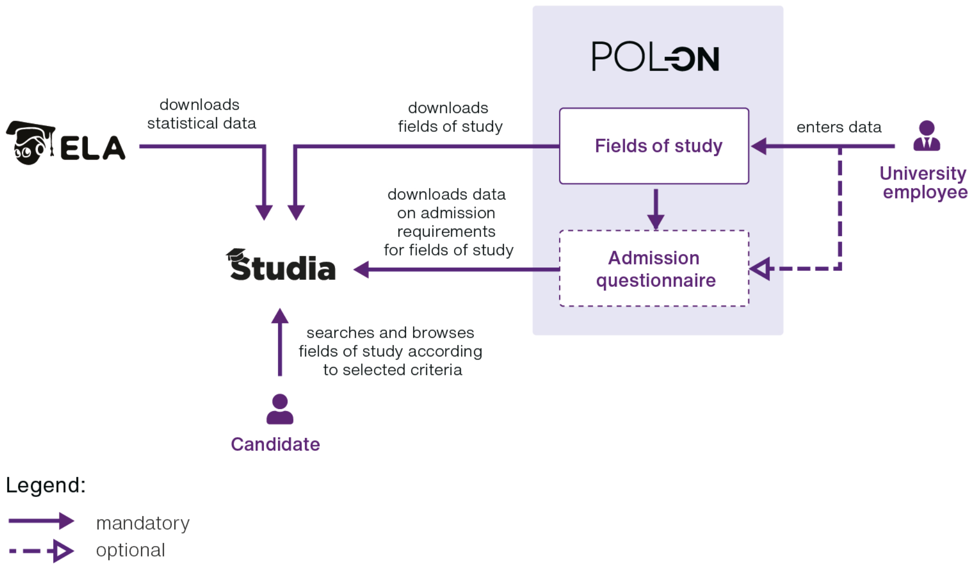

Figure 2.13. Supporting the decision-making processes of university applicants

The collected data can be searched in many ways. Users must only select one or more categories available on the home page, i.e. university names, locations, fields of study, high school subjects (the official list of required subjects can be found on universities' recruitment websites), full-time (stationary) and part-time (non-stationary) programmes, and types of university.

The processes supported by POL-on are crucial for the efficient functioning of the Polish higher education and science system, because:

- They ensure state supervision and control: data from POL-on is used as the basis for shaping science policy, determining the size of subsidies and grants, as well as for evaluating the quality of education, scientific activity and doctoral schools
- They increase the reliability and quality of data: verifications (e.g. with the PESEL register), the automatic detection of fraud, and the archiving of reports guarantee the reliability of information
- They facilitate reporting and accounting: POL-on automates the transfer of financial and statistical data (e.g. to Statistics Poland), which reduces the burden on universities and limits the risk of errors
- They support promotion and academic processes: the registration of theses, doctoral dissertations, and data on the awarding of degrees and titles enables full control of academic career paths and prevents plagiarism
- They enable strategic development activities: data from POL-on feeds into programmes such as IDUB, ELA, and the Studia.gov.pl portal, which supports the development of universities and informed selections by prospective students

- They strengthen the openness and mobility of science: due to cooperation with NAWA and the Foreigners module, it is possible for documents to be verified and for services to be provided to foreign students.

Together, these processes create an integrated tool for managing information on higher education; one that improves the transparency, security, and efficiency of public administration, as well as supporting the development and internationalisation of Polish science.

2.5. The socioeconomic impact of POL-on

Access to reliable, comprehensive, and verified data is a crucial asset in the modern world. Data helps its users to make informed decisions, and drives the advancement of social and economic systems. Therefore, any analysis and description of the impact of POL-on, as a centralised source of data on institutions in the Polish science and higher education system, must embrace the following dimensions:

- **organisation**: the ways in which the introduction of POL-on impacts the operation of scientific and higher education institutions
- **monitoring and governance**: the influence of POL-on on decision-making processes and problem-solving in the sector
- **research support**: the contribution of POL-on to enhancing knowledge of the sector.

Additionally, POL-on influences the socioeconomic environment by providing reliable, up-to-date data to services and platforms that disseminate that information and provide open access to it. Therefore, two further areas of influence emerge at the intersection of the dimensions described above:

- **open data**
- **informing citizens**

Organisation

The introduction of POL-on required adjustments at the level of higher education institutions. These changes have both technological and personnel dimensions. The technological aspects of POL-on's impact on organisations include the encouragement of institutions to develop their own technological solutions and the integration of those solutions with the central system.

At a personnel level, the need to support the system creates the requirement for roles—or even whole organisational units—that are responsible for cooperating with POL-on's developers and central team. Such changes typically necessitate either an increase in the size of a workforce (and the additional costs that it entails) or changes to the duties of current employees, which might, in turn, result in increased workloads rather than true restructuring of work. One of the aims of the remote and centralised reporting for universities and research institutes that POL-on provides is to reduce bureaucracy. The system enables authorised stakeholders and supervisory bodies to reuse its data for their needs after it is uploaded, without the necessity of completing a variety of specific reports (e.g. for the purposes of Statistics Poland). POL-on's effectiveness has been demonstrated by the complete elimination of a form used in statistical reporting: *The Report of the University Activity*—which, for large institutions, typically covered approximately 1,000 pages of detailed data.

Another dimension of POL-on's influence on scientific and higher education institutions lies in its economic outcome—particularly in cost reduction. The system is fully financed by the MNiSW. Consequently, universities, scientific institutions, and supervisory bodies are unburdened by any additional financial costs related to the operation of POL-on. Moreover, POL-on is the only IT tool that supports educational and performance reporting processes for some institutions, particularly for small, nonpublic universities.

Despite the many challenges that the institutions had to face during the implementation of POL-on, the system has since became one of the most influential factors in shaping the organisation and management of institutional information at Polish universities, as well as highlighting the importance of information in their activities [6].

Monitoring and governance

Perhaps the most significant dimension of the socioeconomic impact of POL-on is that observed in the monitoring and governance of the Polish science and higher education system. By gathering up-to-date and reliable data, POL-on supports a host of processes in the governance of science and higher education system: 'funding, evaluating, controlling, monitoring, and informing' [39]. By storing and sharing complex data, POL-on also enhances the transparency of the Polish science and higher education system.

POL-on impacts almost all institutions that relate to science or higher education in Poland. The primary beneficiary of the data stored in POL-on is the MNiSW; however, other institutional stakeholders exist that are frequent users of information gathered through POL-on, such as Statistics Poland, the Science Evaluation Committee, the NCBR, the NCN, NAWA, or The Council of Scientific Excellence. Other ministries—including the Ministry of Health, the Ministry of National Defence, and the Ministry of Culture and National Heritage—are also recipients of POL-on's data[97]. Moreover, POL-on provides data to institutions that operate outside of the public sector, such as nonprofit organisations and private firms, including research and consulting agencies. One

[97] polon.nauka.gov.pl/siec-polon (accessed 17 November 2025)

example is the annual report on the business services sector in Poland published by the Association of Business Service Leaders (ABSL), which utilises data on students and graduates [2].

The introduction of POL-on also impacted the quality of data that concerns science and higher education in Poland positively. First, it made the data more accurate, by, for example, eliminating the practice of students being counted more than once for studying more than one subject. POL-on supports three ways to upload data: via REST API, XML files, or forms accessible after logging in to the system via a browser. Reporting institutions are responsible for keeping their data up to date. Moreover, all of the records are validated in line with the rules derived from law and best practice. These rules are a key element of the system architecture. They ensure that the independent policies of universities and research institutes comply with the applicable laws, particularly with regard to the organisation of studies and the employment of academics. Second, channelling the reporting to POL-on helps to minimise delays in analysing current trends in science and higher education in Poland. Previously, only the situation at the end of each calendar year was reported and the key data for processes connected to public statistics and allocation algorithms were calculated on that basis. Third, the introduction of an extensive data history in POL-on has enabled the tracking and analysis of longitudinal data , thus allowing its users to observe how trends in science and higher education change over time.

POL-on is the practical implementation of legislative acts designed to describe and evaluate science and higher education in Poland. In this legal context, the system's implementation has helped to systematise the concepts and assumptions that underlie the operation of institutions that were once often interpreted colloquially. POL-on contributes to data-driven decision-making processes by delivering information that is essential for the cyclical evaluation of Polish science and higher education institutions, which since 2017 has been based fully on data stored in the system [38]. Moreover, POL-on contains algorithms that help to allocate funds to specific institutions—for educational purposes and scientific activity [39]—thereby supporting the effective distribution of financial resources.

POL-on also enables the elimination of undesirable phenomena in the Polish higher education system. The centralisation and monitoring of information ensured by POL-on helps to detect and limit the abuse of scholarships. The system enables the identification of students who have received scholarships from several sources. In addition, POL-on supports antiplagiarism measures. The introduction of the ORPPD[98] together with the JSA[99] has provided universities and supervisory bodies with a robust, consolidated tool for controlling the issue.

98 polon.nauka.gov.pl/orpd (accessed 17 November 2025)
99 jsa.opi.org.pl (accessed 17 November 2025)

One of POL-on's original objectives was to support the unification of the methodologies for collecting data and calculating indices for various rankings. Presently, POL-on serves as a source of data for the most influential rankings of Polish higher education institutions, such as the *Perspektywy* Foundation Ranking[100] and the *Rzeczpospolita* Ranking[101]. Therefore, POL-on contributes to greater visibility of Polish universities, as well as supporting prospective students in making choices with regard to their education.

Last but not least, POL-on provides valuable input into assessments of the influence of scientific activities on society, innovation, and the economy. This is because the system contains impact descriptions of research projects, as well as information on revenue generated from the commercialisation of scientific research results and from research services commissioned by entities outside of the higher education and science system.

Research support

As a centralised database of institutions in the Polish science and higher education system, POL-on not only contributes to the transparency of the sector and serves as a source of information for the financial and evaluation decision-making processes, but also supports the realisation of research and the in-depth analysis of current trends in the sector. The data stored in POL-on is used widely by OPI PIB's experts in research projects commissioned by the MNiSW.

Access to the data stored in POL-on enables the engagement of the scientific community in debates on the state of science and higher education in Poland and, through that, the collection of additional empirical data. Data gathered in POL-on is used in the creation of sampling frames and the design of sampling procedures for surveys conducted as part of research projects commissioned by the MNiSW. Examples include *The analysis of open science in Poland* [21] and *Evaluation of the National Scientific Policy document adopted in 2022* [81]. The official contact details of the managers of universities, research institutes, and other scientific entities derived from POL-on were used to invite them to participate in the surveys as part of these projects. Importantly, using data from POL-on made it possible for all institutions that form the science and higher education system in Poland to participate and express their opinions on vital issues such as the challenges of open science or the direction of national scientific policy.

[100] 2025.ranking.perspektywy.pl (accessed 17 November 2025)
[101] edukacja.rp.pl/badania-rozwoj-biznes/art42526751-ranking-ogolny-liderzy-polskiego-szkolnictwa-wyzszego-2025 (accessed 17 November 2025)

Moreover, data derived from POL-on has been used for complex, in-depth analyses of crucial issues in Polish science and higher education, such as dropout. Analysing the data gathered in POL-on enabled experts from OPI PIB to quantify the scale of university dropout and readmission, identify the characteristics of the students who drop out, and investigate student transfers between fields of study [79].

As a comprehensive and unique source of data that enables the design and conducting of engaging and inclusive research on the key challenges facing the sector, POL-on impacts the practices of evidence-based policymaking in Poland.

Open data

Contemporary research has shifted from free enquiry towards problem solving. Simultaneously, rising research costs necessitate the incorporation and (re)use of existing data to a greater extent [23, p. 54]. POL-on contributes to the opening of data on Polish higher education and science to a wide range of potential users, thus putting the open data policy into practice [39] and supporting problem-solving and decision-making processes in research and innovation [49]. These aims are achieved through Reports, Analyses and Data on Higher Education and Science (RAD-on)[102], a platform that provides open access to selected data with visualisations supplemented with expert commentary, as well as to research reports and analyses of the Polish science and higher education system. POL-on is one of the sources of the data that RAD-on shares. RAD-on aligns with the Open Government Data (OGD) concept. OGD initiatives adhere to the principles of the Open Data movement and acknowledge the assumption that integrating, providing free access to, and reusing public sector data can create added value and potentially impact society and the economy, e.g. by increasing citizens' engagement or promoting innovation [31], [67, p. 10].

The target groups of RAD-on services are: 'universities and institutional authorities that are responsible for research and development, public administration staff, representatives of NGOs, scientists, journalists and other groups of individuals and entities interested in science and higher education, such as innovative enterprises', [68]. The scope and quality of published data, as well as the various access options (suitable for both novice and advanced users, including browsing, downloading, and API access) mean that almost everyone can benefit from it. Moreover, RAD-on enhances the impact of POL-on on decision-making processes by providing IT tools that help in the analysis and interpretation of the data [68].

[102] radon.nauka.gov.pl (accessed 17 November 2025)

Informing citizens

Data reported to POL-on also feeds two platforms that aim to provide the people of Poland with information on higher education in the country: Studia.gov.pl[103] and ELA[104]. Studia.gov.pl is a portal that enables high-school graduates and prospective students to search through the study programmes offered by Polish universities. The search results provide basic information on all fields of study (for which POL-on is a source of data) and the admission rules (excluding master's programmes). ELA, on the other hand, is a periodic scientific study that compiles data derived from POL-on with that from the Polish Social Insurance Institution (ZUS) to provide credible information on graduates' circumstances in the labour market. ELA enables users to view both country-level data and data disaggregated by university and field of study.

Given the aims of the Studia.gov.pl and ELA portals and the scope of the information they provide, POL-on plays a vital role in supporting rational study choices. The scope of influence of the findings provided by ELA is even wider. The MNiSW can use these findings as support in the planning of educational policies as well as evaluating actions undertaken. For universities, such data may serve as a tool in the effective adaptation of study programmes to labour market needs, and for employers as a reference for setting pay scales.

In summary, the impact of POL-on on the socioeconomic environment in Poland is evident in its provision of reliable data on the science and higher education system. This ensures transparency and supports research, as well as delivering information to citizens and evidence for decision-making purposes. The data stored in POL-on supports several services aimed at different users—ranging from academia and governmental bodies to private businesses, nonprofit organisations, and members of the public. POL-on gathers and provides the data that forms the foundation of evidence-based policy in science and higher education. By enhancing the transparency of the sector and helping to combat the negative phenomena, POL-on plays a vital role in increasing trust in science in Poland.

[103] studia.gov.pl/sfos/r/wsapp/wyszukiwarka-studiow/home (accessed 17 November 2025)
[104] ela.nauka.gov.pl (accessed 17 November 2025)

CHAPTER 3

TECHNOLOGY AND SYSTEM ARCHITECTURE

Szymon Roziewski
Marek Michajłowicz
Dr Jakub Kierzkowski

The technological foundation of POL-on reflects the evolution of large-scale, data-driven public information systems in science and higher education. This chapter explores the architectural principles, design strategies, and implementation choices that enable POL-on to handle vast, heterogeneous datasets while maintaining reliability, scalability, and interoperability. It discusses how the transition from a monolithic architecture to a modular, domain-oriented and microservice-based environment has improved performance, resilience, and development efficiency. The chapter also discusses automation, testing, and quality control methods that underpin the system's stability. Together, these aspects illustrate how POL-on has matured into a robust national platform that is capable of supporting both regulatory and analytical functions across Poland's research and higher education landscape.

This chapter focuses exclusively on presenting conclusions from our experiences of the development of the POL-on system on technical grounds. The guiding theme of the experiments is the transition from the original version of POL-on, which had been built based on a monolithic architecture model, to its newest version, which constitutes a conglomerate of distributed services in line with the microservice architecture pattern. This chapter focuses less on presenting the theoretical aspects that describe similar experiences in the rich literature on the subject, such as [13, 15, 24] or [43].

3.1. Applied architecture

The ongoing architectural transformation of the POL-on ecosystem is best understood as a managed transition. This hybrid operational state is a direct consequence of a strategic imperative to migrate from a legacy monolithic architecture that had become incapable of supporting the complex new demands mandated by Polish law. Specifically, a major legislative overhaul known as the Constitution for Science[105] imposed significant new requirements for data integration, architectural flexibility, and rapid feature development that the original monolithic system could not meet, making a complete redesign both necessary and inevitable [46]. This inflexibility rendered the legacy platform unsustainable, necessitating a phased migration that has resulted in the current coexistence of two distinct versions of POL-on:

1. **POL-on 1**: a legacy system built between 2011 and 2018 on a monolithic architecture Characterised by a large, tightly-coupled codebase and a centralised database, this system had become a considerable technical liability. Its rigidity made implementing the legally mandated changes prohibitively slow and risky, leading to development bottlenecks and frequent regression issues. It now serves as the system of record for core functionalities that are progressively being decommissioned as their replacements are deployed.
2. **POL-on 2**: the strategic successor, in development since 2018, which has been built following a distributed, microservice-based architecture. It is not a rewrite, but a strategic decomposition, in which functionality is extracted carefully from the monolith and reimplemented as independent, containerised services.
3. This modern approach, guided by principles like domain-driven design, is designed to provide the scalability, resilience, and development agility required for the future of the POL-on ecosystem [46].

The two systems represent fundamentally different technological eras and design philosophies: the rigid, synchronous world of the monolith versus the flexible, event-driven paradigm of microservices. The prolonged migration period has concluded, and despite their previous differences, POL-on 1 and POL-on 2 now interoperate seamlessly. The integration layer, which once bore the immense burden of translating data and processes between incompatible worlds, has successfully completed its sophisticated bridging role. That layer now forms a critical component of the wider modernisation strategy, ensuring operational continuity while the legacy system is dismantled piece by piece. To quantify the scale of that undertaking, it is important to note that the new development effort constitutes a complex distributed ecosystem of over fifty distinct microservices and integrated REST-based systems. Many of those services are comprehensive, comprising both back-end functionalities and their corresponding graphical user interfaces. The statistical data on the number of lines of code, presented in Table 3.1., serves as a direct indicator of the magnitude of both the legacy system being replaced and the new development landscape. The line counts refer to logical lines of code and exclude comments and blank lines. While lines of code is, admittedly, an imperfect measure of functional complexity, it serves as a powerful and direct indicator

[105] The 2018 act ([72]) necessitated a complete overhaul of POL-on; one that fundamentally reshaped the legal and structural framework of Polish science and higher education. This significant reform marked a pivotal moment, leading to the development of POL-on 2.0

of the project's magnitude and helps to contextualise its status as one of the largest and most critical systems that supports administrative processes at the state level in Poland[106] [46].

Table 3.1. System-wide lines of code in the POL-on 2 ecosystem, categorised by technological function

Category	Technology / Language	Lines of Code (k)
Back-end Logic	Java	903.1
	Python	10.2
	Subtotal	**913.3**
Front-end / UI	JavaScript	892.3
	TypeScript	231.9
	HTML	92.9
	CSS	13.1
	Subtotal	**1,230.2**
Database & Persistence	SQL	241.4
	Subtotal	**241.4**
Configuration & Markup	XML	616.9
	YAML	24.6
	Subtotal	**641.5**
Data & Data Exchange	JSON	878.8
	Text files	676.4
	CSV	58.8
	Subtotal	**1,614.0**
	Grand Total	**4,640.4**

3.1.1. Layers and system components

Despite the differences between the architectures of both versions of the POL-on system, both use analogous components to achieve their goals. In broad terms, they include:

- a database
- business modules responsible for fulfilling individual requirements
- a user interface.

Additionally, POL-on 2 includes a message broker and a data warehouse.

[106] polon.nauka.gov.pl/siec-polon (accessed 17 November 2025)

However, the way these elements are organised differs fundamentally. Figure 3.1. schematically presents the individual parts of POL-on 1 and how they communicate with each other. The schematic illustrates the classic monolithic architecture of POL-on 1. Its defining characteristic is the intense centralisation of resources. All business modules (*1...N*) are tightly coupled through two shared, critical components: a single, large database and a unified user interface. This shared database becomes the central point of integration and contention. Any change to the database schema required by one module carries the considerable risk of creating unintended side effects or breaking functionality in other modules. Over time, this shared-state model leads to a system that is difficult to change, scale, and maintain—a phenomenon often described by experts as a primary driver for migration away from monolithic designs [20]. Moreover, all development teams must coordinate their deployments around this single codebase, which creates significant bottlenecks and hinders the delivery of new features.

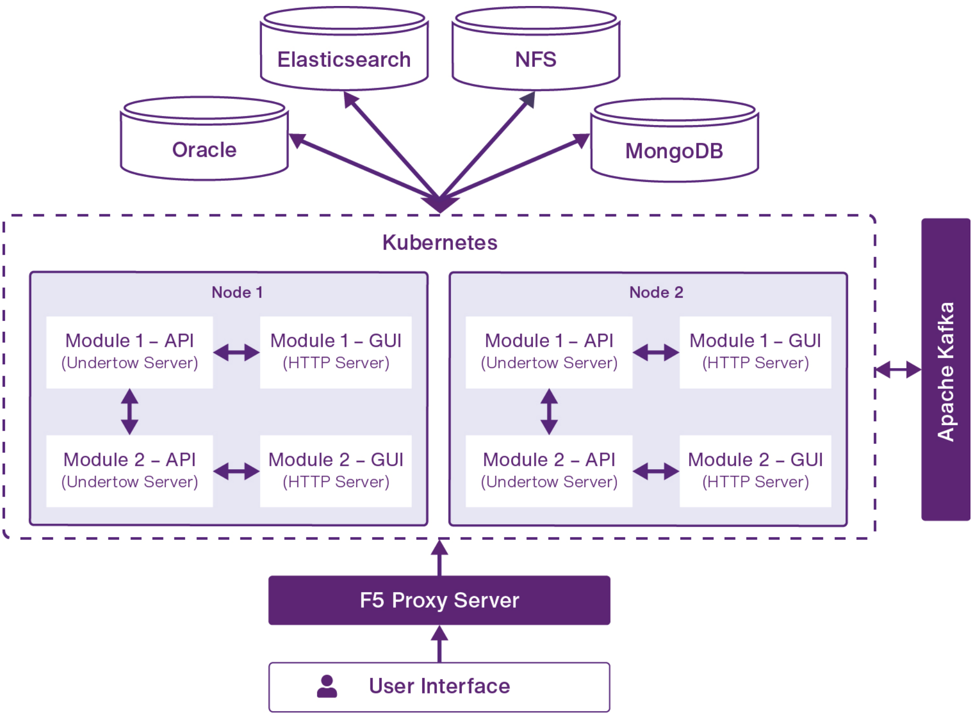

Figure 3.1. Components of POL-on 1

For comparison, Figure 3.2. presents the components of POL-on 2, which embodies a modern microservice-based architecture, in a similar way. This model fundamentally reorganises the system's components to prioritise autonomy and resilience. The most critical difference is the implementation of the 'database per service' pattern [42]. Each business module is a self-contained vertical slice of functionality that possesses its own dedicated database and user interface. This enforces strong encapsulation; the internal data and logic of a module are a private implementation detail, inaccessible directly by any other module. This aligns with the principles of domain-driven design, under which each microservice represents a 'bounded context' with clear, explicit boundaries [16].

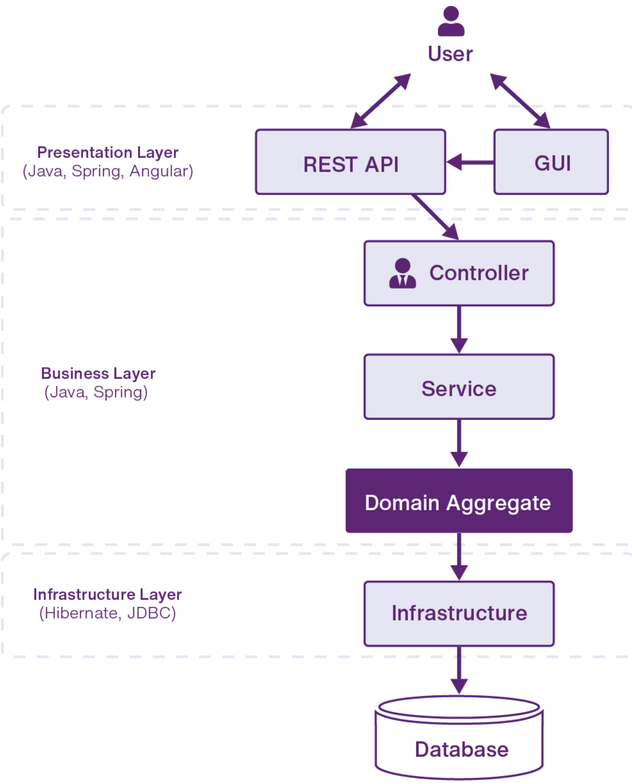

Figure 3.2. Components of POL-on 2

This decentralisation introduces a new challenge: communication. In POL-on 2, this is solved by the introduction of an asynchronous message broker. Instead of making direct, synchronous calls to each other, modules communicate by publishing events to the message broker. Other interested modules can then subscribe to these events and react accordingly. This event-driven approach creates a loosely-coupled system that is far more resilient.

The need for a robust data warehouse in the RAD-on and POL-on IT ecosystems arose from significant challenges in data integration and reporting. Before its implementation, analytical tasks, cross-sectional reviews, and periodic reports relied directly on production systems. This led to fragmented data exchange processes, unstandardised data integration, and large operational overhead on transactional systems. The absence of a centralised and harmonised data source meant that complex cross-domain analyses required arduous manual integration, cleaning, and organisation, significantly increasing the cost and effort associated with generating reports and responding to information requests. The subsequent decision to implement a central data warehouse and a business intelligence system addressed these deficiencies by providing a unified and credible source of data, thus optimising reporting processes and facilitating comprehensive cross-sectional analyses that had been prohibitively costly [55, 67]. For instance, during the evaluation of institutions' scientific activity, the data warehouse proved indispensable by collecting, integrating, cleaning, and converting data from the POL-on, PBN, and Web of Science systems to feed SEDN.

The architectural design of the OPI PIB data warehouse is distinguished by its capacity to integrate a heterogeneous landscape of data sources, ranging from various relational databases (Oracle, PostgreSQL, MySQL, MSSQL) to object databases (MongoDB), message brokers (Apache Kafka), XML documents, REST APIs, and static files. Using a sophisticated extract, transform, and load procedure, the data warehouse leverages Oracle Golden Gate for real-time replication from Oracle databases, achieving near-instantaneous data availability with minimal impact on production systems. For other sources, Oracle Data Integrator or custom Python scripts orchestrated by it are used, with replication delays tailored to business requirements [39]. This multifaceted approach, combined with the strategic implementation of data marts using both classic star schemas and data vault methodologies, ensures the provision of clean, organised, and thematically grouped data for various analytical purposes, significantly reducing the development burden on software engineers and serving as a 'single source of truth' for both internal and external system integrations.

The data warehouse plays a crucial and symbiotic role in this distributed architecture. Although the 'database per service' pattern grants autonomy, it makes performing cross-module queries and system-wide analytics extremely difficult. The data warehouse solves this problem by aggregating data from the individual module databases into a centralised, read-optimised repository. This design effectively implements the command query responsibility segregation pattern at an architectural level. Individual microservices handle the 'command' side (writing and updating data), while the data warehouse handles the 'query' side (complex reporting and business intelligence) [16, 74].

In summary, the transition from POL-on 1 to POL-on 2 is not merely a technical upgrade, but a paradigm shift from a highly coupled, shared-state architecture to a decentralised, message-driven ecosystem in which module autonomy is paramount. The message broker and the data warehouse are not add-ons, but essential enablers of this modern architecture that provide the necessary mechanisms for communication and data aggregation that make the distributed model viable at scale.

3.1.2. Logical architecture

Since POL-on 2 is implemented in a microservice architecture, individual microservices may differ slightly from each other, depending on the technical solution used. This strategic choice enables a 'polyglot' approach to both programming and persistence, which enables each microservice to be built with the technology stack best suited to its particular business requirements. While the presentation layer is standardised on Angular for a consistent user experience, the back-end technologies vary. Most business layers are implemented in Java using the Spring framework; however, the architectural freedom of microservices permitted the use of Python for modules in which it was deemed more effective. Similarly, the data storage strategy is highly pragmatic: relational databases (Oracle), document databases (MongoDB), search indices (Elasticsearch), and the file system are all utilised. The selection of persistence mechanism—accessed via libraries like Hibernate, Spring JDBC, or Spring Data—is dictated by the needs of the functionality, including its data structure, consistency requirements, and performance characteristics. Some institutions preferred to send complete data records rather than atomic operations. POL-on 2 introduced a document model that enabled partial and invalid data to be saved. This improved usability and simplified validation. Each change creates a new version, enabling rollback and history tracking.

According to Figure 3.3., which presents the canonical logical architecture of a single microservice, the design follows established principles of clean architecture, which ensures a clear separation of concerns and protection of the core business logic from infrastructure details [37]. The flow of a user request through these layers is deliberate and enforces a strict dependency rule in which outer layers depend on inner layers, but not *vice versa*.

1. **Presentation layer:** This is the entry point for all external interactions. A user or an external system can interact via either the graphical user interface (GUI) or via a programmatic REST API. This layer's responsibilities are to handle the transport protocol (e.g. HTTP), deserialise incoming requests into data transfer objects, and serialise application responses. It acts as the 'front door' to the application's use cases.
2. **Business layer:** This is the heart of the microservice that contains all of the business logic, carefully partitioned into distinct components:
    - **Controller:** The first stop in the business layer. Its primary role is to perform initial, stateless validation on the incoming data structures and formats. It ensures that a request is syntactically correct and well-formed before the request is processed further. It then orchestrates the execution of the relevant business use case by invoking the appropriate service.
    - **Service:** Often described as an 'application service', this component coordinates the execution of a business operation. It is responsible for handling cross-cutting concerns like transaction management and for performing simple, stateless business validations that do not require deep knowledge of the domain's state. Crucially, the service itself contains minimal business logic; instead, it loads the relevant domain model objects and delegates the core work to them.
    - **Domain aggregate:** This component is the centrepiece of the business layer and a core concept from domain-driven design An aggregate is a transactional consistency boundary: a cluster of domain objects that can be treated as a single unit [16]. The domain aggregate component is responsible for enforcing the most complex business rules and invariants by considering its own state and that of re-

lated objects. All modifications to the objects in the aggregate must pass through the 'aggregate root', which ensures that the model remains in a valid state at all times. This is where the true, unadulterated business logic resides, completely isolated from any technical or infrastructure concerns.

3. **Infrastructure layer:** This layer implements the interfaces defined by the business layer. It contains the 'how' of the application: the concrete code that interacts with the database, message brokers, or other external systems. By adhering to the dependency inversion principle, the business layer defines what it needs (e.g. an interface for a *UserRepository*), and the infrastructure layer provides the implementation (e.g. a *HibernateUserRepository* class). This separation is vital, as it enables the persistence technology to be changed with minimal impact on the core business logic, which remains stable and protected at the centre of the architecture.

In conclusion, this structured, layered approach ensures that each component has a single, well-defined set of responsibilities. It protects the invaluable domain logic from being entangled with technical details, making the system easier to test, maintain, and evolve.

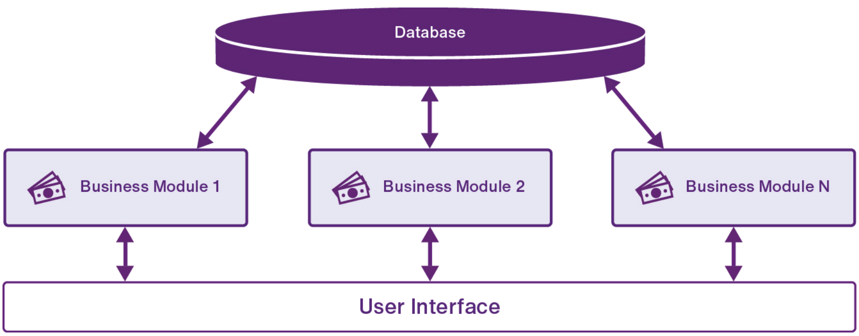

Figure 3.3. Logical architecture

3.1.3. Architectural design principles

The transition to microservices required a strategic redefinition of the POL-on system's boundaries. Although POL-on was conceptually modular, the technical dependencies between its components were strong. To avoid excessive fragmentation and maintain coherence, we adopted the following principles:

- Limit the number of services to avoid the antipattern of nanoservices
- Define module boundaries based on business logic and functional cohesion
- Separate the API and GUI layers to enable independent updates
- Minimise intermodule dependencies and promote loose coupling
- Use direct API calls and asynchronous messaging (e.g. Kafka) for interservice communication.

3.1.4. Communication between POL-on 1 and POL-on 2

With the need for communication between systems that remain part of the POL-on 1 ecosystem, experiments related to the implementation of an enterprise service bus commenced. Although it was intended to be responsible for communication between applications, this solution was ultimately rejected for the following reasons:

- The necessity of establishing a canonical data model, as well as managing, maintaining, and extending it, as proposed in [75]. In [33], such a model requires a combination of data from different sources, while [36] presents such a model that is appropriate for data warehouses. The key problem was the great difficulty in any modification of such a model, as such a change could entail changes in all systems already integrated with the bus
- Moving the logic of the data consumer to the enterprise service bus layer. It was also noted that using a data bus causes the consumer logic to penetrate into the bus
- The labour intensity associated with maintaining another layer of the system, resulting from the additional elements that duplicate the business logic and data model of integrated applications.

As a complementary solution to the data bus, we also tested a solution based on REST services. Such services are provided by modules with which such services should be integrated. Such a module is ‘aware’ of the sharing of its data, and therefore stores the data in a model optimised for reading. To reduce network traffic, we also introduced a message broker. Its purpose is to inform the users of changes made in the source system. If the recipient of such a message is interested in such a change, they may download it and store it in their cache. During the work, we tested variants of messages sent by the broker:

- Only changes to objects were sent along with the identifiers of such objects. This approach was intended to minimise network traffic. Nevertheless, the chief problem that arose was the inability to determine whether the recipient of such a message had data on the previous value of the object to correctly apply the received change
- Entire changed objects were sent along with information about the type of change. Although this approach was intended to eliminate the weakness of the previous solution, it soon became apparent that not all message recipients needed to receive all messages
- Only information on the changes was sent along with information on the type of change with links that enabled the new data to be downloaded.

We are currently conducting tests and evaluating which of the solutions presented above will work best in our environment.

The exchange of information between systems is a complex operation. When analysing ways of exchanging messages, emergency situations that might arise during a system’s operation is an element that is frequently overlooked. Applications communicate with each other via REST services and keep the information thus obtained in a cache, which, in turn, is updated by a message from the message broker. Depending on the implementation, the cache may persist continuously or for a defined period. The key task is to ensure the consistency of the cache with the source of truth, which, depending on the implementation, might be based on messages from the message broker, periodic refreshing of the cache, or a combination of the two. The POL-on 2 system communicates with POL-on 1 in the manner described above.

For the sake of completeness, we will now also describe the method of integrating POL-on 2 and POL-on 1 systems through the database. Some data can only be changed in POL-on 2. Such data is transferred to POL-on 1 through a database change induced by a trigger in the database. This is a process that is difficult to test. If an error is detected, it is difficult to determine which data, specifically, is incorrect. It is also difficult to assess whether a process related to data migration has already been fixed and whether the error continues to occur. This is a temporary solution until all modules are transferred to POL-on 2.

3.1.5. Physical architecture

Figure 3.4. schematically presents the physical architecture of the POL-on 2 system, which serves as the concrete implementation of the logical architecture described in the sections above. The design is rooted in modern cloud-native principles, prioritising scalability, resilience, and high availability.

A user's request first reaches an F5 proxy server, which acts as a load balancer and a single entry point into the system. Its primary role is to distribute incoming traffic across the available computing resources, preventing any single node from becoming a bottleneck and ensuring service continuity even if a downstream server fails. This component is critical in the achievement of high availability in a production environment.

From the proxy, a request is routed to the core of the application's runtime environment: a Kubernetes cluster. As a container orchestration platform, Kubernetes is responsible for automating the deployment, scaling, and management of the application's containerised components [9]. The illustration depicts a multinode cluster (showing 'Node 1' and 'Node 2'), which is fundamental to the system's design. This multinode setup enables:

- **Horizontal scalability:** The system's capacity can be increased by the simple addition of more nodes to the cluster
- **High availability:** Kubernetes automatically distributes replicas of each microservice (e.g. 'Module 1 API', 'Module 1 GUI') across multiple physical nodes. If one node fails, Kubernetes automatically reroutes traffic to the healthy replicas on other nodes, which ensures uninterrupted operation.

A key architectural decision illustrated in the diagram is the separation of the stateful persistence layer from the largely stateless compute layer running on Kubernetes. Although the primary database for POL-on remains Oracle, individual modules also leverage MongoDB, Elasticsearch, or a network file system. Critically, these persistence services are managed externally by the Kubernetes cluster. The decision not to containerise the Oracle database is a pragmatic one, driven by the high licensing costs and the operational complexity of managing traditional, monolithic relational databases in a containerised environment.

Similarly, the Apache Kafka message broker, which provides the asynchronous communication backbone for the microservices, is deployed as an external service available to all Kubernetes nodes. This is a common and robust pattern, as stateful systems like Kafka typically require dedicated management and infrastructure to guarantee the high throughput and durability required for an event-driven architecture [35].

While the services for Elasticsearch and MongoDB also currently lie outside of the cluster, the text notes that this is due to historical legacy rather than technical limitations. The plan to migrate these services into the Kubernetes cluster reflects a growing industry trend towards managing stateful applications on Kubernetes using advanced features like StatefulSets and Operators. This future-state architecture aims to unify infrastructure management, enabling more efficient and consistent operations across the entire ecosystem. The physical architecture of POL-on 2 represents a mature, hybrid approach that leverages the power of container orchestration for its stateless services while pragmatically integrating with external, specialised systems for its stateful data management.

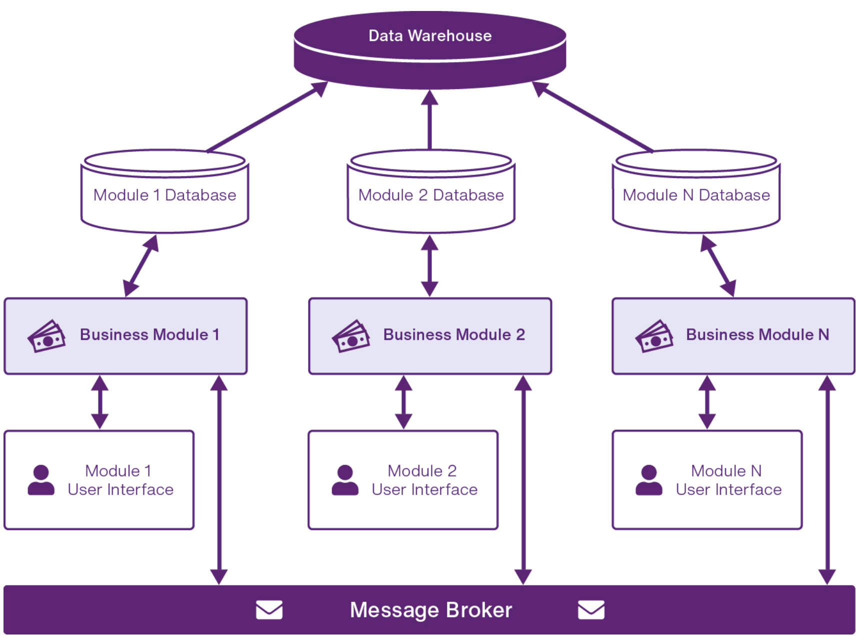

Figure 3.4. Physical architecture

3.1.6. A Platform-centric, domain-oriented approach

An architectural analysis of the POL-on 2 ecosystem reveals a mature and deliberate implementation of a distributed system. The design illustrates a strategic migration from a monolithic paradigm towards a cohesive ecosystem of domain-oriented microservices [42]. This modern architecture is underpinned by a set of well-defined, cross-cutting platform services that provide a unified development and delivery platform with a focus on consistency, resilience, and operational efficiency across the enterprise. The key architectural patterns include a platform-based approach for foundational services, clear domain decomposition following domain-driven design principles [16], a hybrid communication model that leverages both synchronous and asynchronous patterns, pragmatic polyglot persistence, and a managed transition from a legacy system using the Strangler Fig pattern [17]. The analysis below will elaborate on these findings, demonstrating how the POL-on 2 architecture represents a robust and scalable solution for a complex, enterprise-level domain.

The most salient architectural pattern that is emerging is the establishment of a common platform that provides foundational, cross-cutting services. Rather than obligating each microservice team to build its own infrastructure for common needs, the architecture centralises these concerns into a shared platform. This approach is critical for maintaining governance and efficiency in a large-scale microservice landscape, enabling domain-focused teams to achieve high velocity with reduced cognitive load [64].

The architectural model, which provides a standardised and supported technological foundation, comprises several key pillars. This is illustrated conceptually in Figure 3.5. A centralised identity and access management provider—such as Keycloak, MCL, or Polon2 Security—serves as the single source of truth for authentication and authorisation, decoupling security from business services and ensuring consistent access control. Moreover, the ubiquitous use of a centralised configuration server indicates a strategic decision to externalise configuration from application code, which enables dynamic changes without redeployment. The bus for the ecosystem is an asynchronous messaging platform, Apache Kafka, whose presence for publishing domain events signifies a deep commitment to an event-driven [7] architecture. This promotes loose coupling, enhances system resilience, and enables asynchronous processing and future integrations by treating the log as the central nervous system of the enterprise data architecture [34]. This is complemented by an externalised business rules engine, the Ruleski service, which decouples complex validation logic from the service's core implementation, thereby increasing agility. Finally, a data and analytics platform, the data warehouse, acts as a central aggregation point, solving the problem of distributed data by providing a unified view for reporting and business intelligence without imposing performance-draining queries on the live transactional systems.

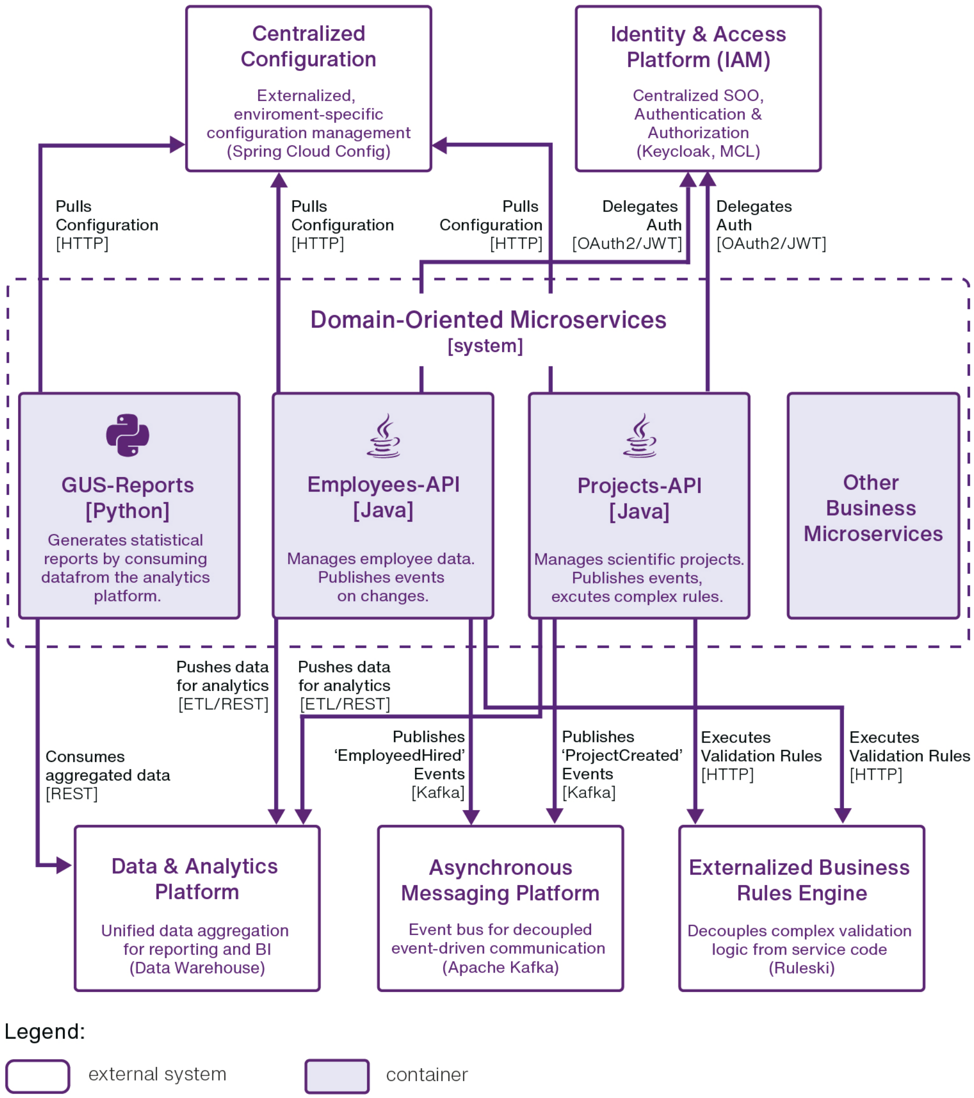

Figure 3.5. The conceptual platform architecture of the POL-on 2 ecosystem. This illustrates the relationship between domain-oriented microservices and the five core platform services

This platform-centric approach, visualised by a C-4 diagram in Figure 3.5., provides the foundation for the second key architectural principle: a clear domain-oriented decomposition. Each service aligns with a distinct business capability (e.g. Employees-API, Projects-API, Investments-API), clearly delineating a bounded context as prescribed by domain-driven design [16]. This decomposition is crucial for granting teams the autonomy to evolve their services independently, enabling faster delivery and flow of value [64]. This autonomy naturally extends to data storage, which represents a pragmatic approach to polyglot persistence [42]. While many services leverage separate schemas in a shared Oracle instance—a common choice that balances cost and operational complexity—the architecture does not enforce a single database technology. For example, Projects-API, a Java-based microservice, embodies a distinct business capability by managing scientific project data autonomously. Its design demonstrates robust integration with the broader platform by publishing 'ProjectCreated' events via Apache Kafka for asynchronous communication, interacting with the externalised business rules engine for complex validation logic, and pushing relevant data to the data and analytics platform for centralised reporting, all while delegating authentication to the identity and access platform. Similarly, the Employees-API microservice, also implemented in Java, focuses exclusively on managing employee data, which represents another critical bounded context in the POL-on ecosystem. It follows an identical pattern of platform integration, publishing 'EmployeeHired' events to Kafka, leveraging the Ruleski engine for business rule execution, and contributing its data to the analytics platform. The two services collectively illustrate the architectural goals of clear functional delineation, enabling independent evolution and development while consistently using the shared platform infrastructure for security, configuration, messaging, and analytical insights. This principle of autonomy extends beyond functional boundaries to encompass technological choices. This demonstrates a pragmatic approach to polyglot persistence. For instance, Foreigners-API utilises PostgreSQL, and GUS-Reports is a Python service that interacts with the data warehouse and file storage. This exemplifies the principle of using the most appropriate data storage technology for a service's particular needs, a hallmark of a mature microservice architecture. In addition to its internal decomposition and data management, the architecture employs a sophisticated hybrid communication model. Although asynchronous, event-driven communication via Kafka is favoured for decoupling and resilience, the design also incorporates extensive use of synchronous, direct REST API calls between services. For example, Foreigners-API and Patents-API both rely on Users-API and Institution-API for referential data. This synchronous pattern is appropriate for queries that require immediate responses. This demonstrates a nuanced understanding of communication trade-offs rather than dogmatic adherence to a single pattern [42]. This strategic approach to internal communication is complemented by a clear pattern of external evolution: consistent integration with the POL-on 1 (legacy) system in multiple services is a clear implementation of the Strangler fig pattern [17]. This enables a gradual, controlled migration of functionality from the old monolith to the new microservice ecosystem, minimising risk and ensuring business continuity during a prolonged transitional period.

In conclusion, collective evidence paints a picture of a well-designed and modern enterprise system that moves beyond a simple collection of services to establish a cohesive platform ecosystem. By providing centralised solutions for identity, configuration, messaging, and data analytics, the architecture empowers domain-focused teams to deliver business value efficiently and consistently. The pragmatic and deliberate combination of domain-driven design, hybrid communication models, polyglot persistence, and a managed legacy migration strategy demonstrates a high degree of architectural maturity. The system is designed not only to meet current requirements, but also to be adaptable, scalable, and resilient in the face of future evolution.

3.2. Managing business domain complexity

The development of POL-on 2 revealed that different modules required different implementation strategies, in accordance with the varying degrees of complexity and functional specificity throughout the system. A central challenge in the system's design was the accurate representation and consistent enforcement of business rules. To address that, the development team adopted domain-driven design as a strategic approach to managing complexity in the business domain.

3.2.1. Domain-driven design

Domain-driven design was applied to modules characterised by intricate business logic and frequent changes. Although it introduces a steep learning curve, domain-driven design offers clear modular boundaries, a shared language between stakeholders, and a maintainable codebase, making it a robust foundation for long-term system evolution.

Lessons learned from the previous version of POL-on underscored the need for a domain model that could evolve alongside the business domain. Consequently, the team embraced domain-driven design as a comprehensive software development methodology [16]. This approach stresses close collaboration between cross-functional development teams and domain experts, fostering a shared understanding of the domain through techniques such as event storming [7]. Workshops on event storming facilitated effective domain exploration, enabling the identification of key business concepts and their relationships. As a result, the system was decomposed into modules that reflect the structure and semantics of the higher education domain. Domain experts played a pivotal role in defining bounded contexts and identifying real-world entities to be represented in the source code (Figure 3.6.).

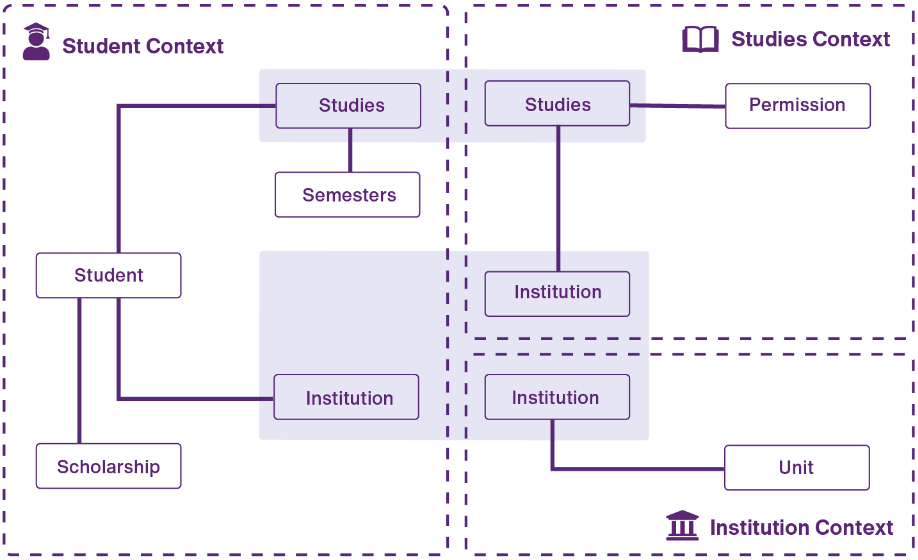

Figure 3.6. Bounded contexts with unrelated concepts (student, permission, unit, scholarship) and shared concepts (studies, institution)

The domain model serves as the backbone of a ubiquitous language, used consistently by developers and domain experts to communicate and reason about a system. Using strategic design principles [74], the development team partitioned the domain of higher education into subdomains: core, supporting, and generic. The core domain, being the most critical to stakeholders, received the greatest investment in terms of design and implementation effort. Domain experts were involved in the process of defining the boundaries of modules (bounded contexts), and identifying the real-world objects (Figure 3.6.) that should be reflected in the source code. In this way, the model of the domain is forged by all key participants.

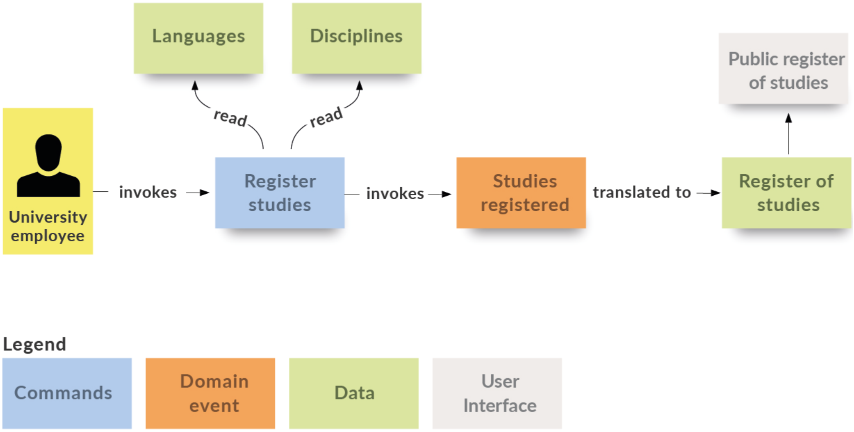

Figure 3.7. A simplified result of event storming regarding one domain event

A key outcome of the event storming was the identification of domain events and business entities, which enabled the modelling of state transitions and lifecycle behaviours (Figure 3.7.). This laid the groundwork for implementing event sourcing, in which changes in entity state are captured as a sequence of domain events [18]. Event sourcing enables the reconstruction of current or historical states by replaying event logs, and facilitates intermodule communication through published domain events. To address performance and scalability concerns—particularly those that arise from a centralised relational database and high-read operation volume—the system architecture was designed to support dual data models: one optimised for updates, and another for reads. The implementation of the domain model was realised using microservices, with each bounded context mapped to an independently deployable service (Figure 3.8.). This architectural style promotes loose coupling, enabling teams to develop, test, and deploy modules autonomously. It aligns with the principles of microservice architecture, offering agility and resilience in system evolution [46].

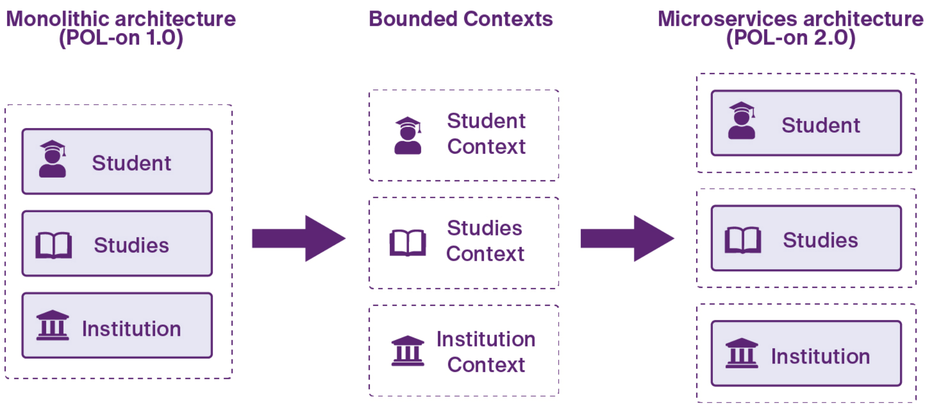

Figure 3.8. A simplified result of event storming regarding one domain event

3.2.2. Benefits and trade-offs of microservices in POL-on

The transition to microservices in POL-on has yielded substantial benefits:

- Increased system availability
- More frequent and targeted deployments
- Greater flexibility in technology selections
- Improved scalability and fault tolerance
- Enhanced developer autonomy and productivity
- Optimised infrastructure costs.

However, these advantages come with trade-offs:

- More complex design and planning
- Increased architectural overhead

- Loss of global transactional integrity[107].

Despite these challenges, the microservices model has proved effective for POL-on and reflects broader trends in the design of IT systems for higher education.

3.3. Quality control method automation

Throughout its existence, POL-on has been developed using the Scrum methodology, through the implementation of user stories during two-week sprints, followed by the deployment of the system to the production environment. As a young system that comprised approximately a dozen modules at that time, with a relatively low level of complexity, POL-on was tested by a team of four who had experience in manual testing. A new version of the system was deployed daily in the development environment. As part of the deployment procedure, unit tests created by members of the development teams were run automatically on that environment. The testers conducted manual story tests and retests[108] of any system defects discovered there.

After completing a sprint, implementing all stories, repairing defects found during earlier tests in the development environment, and conducting confirmation tests, a pre-implementation version of the system was prepared and deployed to a separate preproduction test environment. There, the tests team examined two aspects.

The first was the correctness and completeness of the new version of POL-on. The testers verified that all newly-created system features were available with all patches (i.e. that they corresponded to what had been deployed on the development environment at the end of the sprint).

The second aspect was regression testing. All members of the test team performed these tests in parallel, simultaneously for system functions available to users logged in in the context of a supervisory unit, to users logged in in the context of a science and higher education institution, and to users who were not logged in (publicly available reports). All main views displayed by the system, buttons that trigger basic action paths, and wizards in the system were tested. The tests pertained to the main use cases of POL-on (adding and displaying data, without attempting to enter incorrect data, and without editing or deleting data), possibly supplemented by other cases based on the testers' experience, performed manually. Regression tests in the preproduction environment lasted up to approximately one working day.

[107] Inconsistencies in the data shared by many independent modules (e.g. if the name of a person registered in the Employee module differs from a name of the same person registered in the Doctoral Student module, this situation is treated as an inconsistency). The solutions for this type of risk are described in [46]
[108] The definitions of the terms used in this subsection are taken from the ISTQB® glossary [27]

In the case of significant defects, corrections were made and the amended version was implemented in the preproduction testing environment, followed by another quality control check. Minor defects were fixed in the following sprint and retested in the development environment. After the tests were completed satisfactorily, the system was deployed to the production environment.

Works on the preparation of test automations were performed concurrently. The developers prepared a test framework and functional tests for the first POL-on modules using Cucumber and Selenium. Further development of the framework and the creation of tests for the remaining modules were taken over by the test team. Another test environment was then created, intended exclusively for daily automated regression tests. The use of test automata also replaced the preproduction environment tests; some of the automated tests (that correspond to the manual tests previously performed in that environment) were selected, marked, and configured so that they could be performed after deployment on the preproduction environment.

3.3.1. Test automation tools and usage patterns

The test framework is a module in the POL-on source code repository that contains code written using the Cucumber and Selenium libraries. It contains text files with the feature extension written in Gherkin[109], Java classes that translate Gherkin commands to Java code[110], and the actual Java code that uses Selenium WebDrivers[111] to launch POL-on and perform test actions.

Automatic tests are performed concurrently using the Selenium Grid technology. A Selenium application server is run as the network hub on one machine (with which the test automata communicate) and as nodes on multiple machines, to which the hub directs instructions to perform each test. Every Selenium node is configured to run a specific number of instances of individual web browser drivers. The grid used for POL-on testing comprises five virtual machines running Linux Mint (two machines, one originally installed with a 32-bit system and later upgraded to 64-bit, the other with 64-bit from the outset), Windows XP (later replaced by Windows 10), Windows 7, and Windows 8.1, each with Selenium WebDriver drivers for Google Chrome and Mozilla Firefox browsers (the machines with Windows systems also have Microsoft Internet Explorer, and on Windows 10, the Microsoft Edge browser). One of the machines that runs Linux Mint performs the function of both a hub and a node. During the development of the test suite, the team

[109] Gherkin is a component of Cucumber; a set of keywords used to compile test cases from steps defined in a natural language; Listings 3.1.–3.3. contain examples of the Gherkin code, i.e. the translations of the original steps' names, as the definitions were written in Polish

[110] The method in which the test step is implemented has a Cucumber annotation that contains the name of the step used in Gherkin (or a regular expression to which the names are matched)

[111] Part of Selenium; the driver launches and controls the selected web browser

considered running the test in twenty threads to be optimal under the existing infrastructure conditions (for the full test pool; in certain situations other numbers of threads were also used: ten, twenty-five, fifty, or 100).

The limitations of the tools used[112] are also noteworthy: the pool of tests to be performed is divided *a priori* into twenty parts, each run in a separate thread. When the automated testing process is launched, a request to perform twenty tests is sent to the Selenium hub (the first test from each part is used).

The hub distributes them to subsequent nodes, four instances of WebDriver (a web browser) per node. The aim of the distribution is to use as many nodes as possible, so that each one is used to a similar extent. If all of the tests planned for a given thread are completed, the remaining tests run in nineteen threads; the Selenium hub receives fewer requests. In practice, in the case of POL-on, it is repeatedly encountered that some of the tests assigned to one thread are incomparably more time-consuming than others, which means that for a large part of the automated testing process, they are not actually performed in parallel. This phenomenon is described in the 3.3.2. section of this chapter.

Automated regression tests were performed according to the following scheme in a separately designated environment. First, a new version of the system was deployed (the code came from the same branch of the repository as for the development environment). Next, a snapshot of the POL-on database was created. Then, the test automata were launched. After the tests were completed, the database was restored from the created snapshot to its previous state. The entire process, unless disrupted by external factors, required approximately eight to twelve hours. Tests on the preproduction environment were performed without creating and restoring snapshots, with fewer threads, and lasted approximately eight hours.

3.3.2. Challenges in the utilisation of testing automata

The monolithic architecture of POL-on 1 forced POL-on's automated tests to adopt an equally monolithic architecture . Automatic functional tests were implemented as end-to-end tests: system tests at the graphical user interface (GUI) level. The disadvantages of this approach to automation, combined with other factors, caused a number of problems that the test team had to overcome.

The first was the time-consuming nature of the tests. While running nightly tests that lasted over eight hours was not burdensome as such, the many hours of regression testing in the preproduction environment did not enable decisions to be made on deployment to and testing in that environment, and deploying POL-on to production on

[112] At least, the versions of these tools used at the time, which were later no longer used in the same manner

the same day. Reports from tests that lasted longer than a tester's work day inevitably had to be checked during the next work day. When test automatons (or manual tests conducted simultaneously) detected a system defect, the information reached the programmers with a delay. Even the immediate repair of defects led to another deployment and further hours of testing (unless a decision was made to abandon the use of test automata when the risk of regression errors resulting from the introduction of a fix was considered low and any new errors were considered possible to be found by a tester manually). This rendered the set of automated tests an inconvenient tool.

The second difficulty was test instability. Many automated tests returned false positives or failed to run entirely. This is a direct result of the nature of end-to-end tests: they verify the operation of systems in many aspects simultaneously. For a test performed by Selenium on the user interface to be successful, a number of conditions must be met. Failure to meet even one of them results in a test failure, even if the tested component has no defects. This is due to the tests being overly focused on the GUI. Factors that might cause a false positive test result include:

- a mismatch between the test and a change in the GUI (even if the change does not affect the system's operation from the user's perspective, such as a change in the interface component identifier)
- exceeding the time during which the test waits for a change in the system, resulting in the test determining prematurely that no change has occurred
- attempting to perform the next action too early (and the test determining that it cannot be performed at all)
- one test interfering with the data of another test.

A failure to run one test or all tests might be caused by factors such as:

- incompatibility between the version numbers of the web browser and its driver
- excessive load on machines in the Selenium grid (e.g. too many tests running simultaneously)
- instability of the Selenium server (e.g. during prolonged use)
- changes in the system database (inability to log in to test accounts, absence or change of objects used in tests, e.g. specific institutions)
- errors in the system that block access to the part being tested.

The failure of any particular test occurred sporadically and randomly; however, with a large number of tests, it was certain that not every one with a positive result actually indicated an error in the system. Thus, verifying the automatic test report was a separate task, and it was necessary to reverify selected test cases (manually or using test automation run individually), which prolonged the testing process further.

The daily tasks of testers included the creation of new tests. The appearance of newly-written tests depended largely on how the test framework was prepared. This highlighted the difference between the approaches of the people who initially defined the test steps in Gherkin and the methods of translation from Gherkin to Java. The differences in possible approaches are illustrated by code snippets in Listings 3.1. and 3.2., and are based on real POL-on test code.

Listing 3.1. Code of a test of transition from search results to student data (first version, 2014)

```
Feature: STD.UC.004.View student studies data
  Background:
    Given I log into Polon with the role 'INST_PR_WS'
    And expand the browser window
    And navigate to the "STUDENTS -> Students list" menu
    And wait for the page to load for '60' seconds

  Scenario Outline: View the student's studies data from the list: 1 row - 1 person
    When select the value '---All---' for the 'Academic year:' filter
    And wait for the page to load for '3' seconds
    And expand the row labelled 'ADDITIONAL CRITERIA'
    And fill in the 'Surname:' filter with the value '<surname>'
    And press the button labelled 'Search' and wait for '60' seconds
    And click on the link named '<surname>'
    And click on the tab named 'Studies'
    Then check if the page contains '<fieldOfStudy>'
    And check if the page contains 'Study details'
    And I log the user out

    Examples:
      | surname  | fieldOfStudy |
      | Kowalski | Ethnology    |

  Scenario Outline: View the student's studies data from the list: 1 row = 1 field
         of study
    When select the value '---All---' for the 'Academic year:' filter
    And wait for the page to load for '3' seconds
    And expand the row labelled 'ADDITIONAL CRITERIA'
    And fill in the 'Surname:' filter with the value '<surname>'
    And press the button labelled 'Search' and wait for '60' seconds
    And select the option '1 row = 1 field of study'
    And wait for the page to load for '5' seconds.
    And click on the link named '<surname>'
    And click on the tab named 'Studies'
    Then check if the page contains '<fieldOfStudy>'
    And check if the page contains 'Study details'
    And I log the user out

    Examples:
      | surname  | fieldOfStudy |
      | Kowalski | Ethnology    |
```

Listing 3.2. Code of a test of the scientific conferences search engine (first version, 2014)

```
Feature: DUN
  Scenario Outline: DUN_UC_001_Conference search
    When I log into Polon with the role 'INST_KONFERENCJE_ADM'
    And select the context of the institution 'Uniwersytet Warszawski'
    And navigate to the "Activities promoting science -> Scientific conferences"
          menu
    And enter the name of the conference '<name>'
    And enter the year of organisation '<year>'
    And select the type of conference '<type>'
```

```
    And click the Search button
    Then the searched conference '<name>','<year>' is displayed
    And I log the user out

    Examples:
      | nazwa  | year | type          |
      | Name 1 | 2012 | international |
      | Name 2 | 2013 | national      |
```

In the case of Listing 3.1., an incorrect approach can be observed: single business operations are described in multiple steps, separate steps are used to wait for events in the system, and waiting times are fixed. Although Listing 3.2. appears to be correct, it also appears to have flaws of its own. For example, the `Search` button in the conference search engine does not have to be implemented in the same way as the button with the same name in another search engine, which means that the method of finding that button in the GUI might fail. If so, each time the test is performed using the `click the Search button` step in another search engine, the result will be positive (the test will fail).

There are several ways to resolve such a conflict, depending on the situation. It is possible to change the name of the step so that it indicates the context in which it can be used more clearly, and to implement another step with a nonconflicting name separately; however, this might lead to an increase in the number of steps with similar names. On the other hand, it might happen that the same action in the system is implemented independently by different people in different modules. Alternatively, a change can be made to the implementation of the step (in the example: finding any 'Search' button in the interface), which might result in unexpected, incorrect behaviour (e.g. failure to detect the absence of the desired button due to the presence of another button with the same label or pressing the wrong button with the same label, breaking the first test). It is also possible to change the division of the test into steps as early as the Gherkin layer—and in this example, implement pressing the correct button in the step of using the selected search engine. The latter approach can be observed in Listing 3.3. It contains the final version of the test (the first version of which is 3.1.) and has all the advantages seen in 3.2.

Listing 3.3. Code of a test of transition from search results to student data (final version, 2017)

```
Feature: STD.UC 004.View student studies data
  Scenario Outline: Display student studies data
    Given As 'INST_PR_WS2' using menu "STUDENTS -> Students list"
    And select '<option>'
    And search for students:
      | Surname:       | <surname> |
      | Academic year: | ---All--- |
    When navigate to the found student's data
    Then check if the student's data is displayed

    Examples:
      | surname  | option                  |
      | Kowalski | 1 row = 1 person        |
      | Kowalski | 1 row = 1 field of study |
```

The difference in approaches transpired to be another difficulty in the operation of test automata. This made the implementation of new tests longer and more difficult, forced the improvement of old tests (increasing the number of tasks to be performed by the testers), reduced the readability of the tests, and made it difficult to distribute responsibilities evenly among the testers. All of this leads to a lack of time to improve test quality.

For each of the issues described above, better solutions are possible—although this takes time. The worse the tests were written, the more time it took to perform and maintain them, which, in turn, meant that the test team members had less time to devote to improving the automated tests—all while having to perform manual tests of ongoing changes to the system.

The last point to be raised is the lack of experience in the test team: not enough of its members were experienced in using test automation tools that could deal with the difficulties described above (this relates closely to the issue of time).

3.3.3. The solutions to the difficulties

All the difficulties described in Section 3.3.2. have been resolved, partly due to a change to microservice architecture in POL-on 2, and partly due to the development of the test team.

The development of the test team's skills was achieved in several steps. The team was expanded to include a person with experience in programming test automations, who was not assigned any tasks related to manual testing or the ongoing maintenance of existing tests. That person addressed the weaknesses of the entire testing framework, developed best practices for writing new automated tests, implemented new tests, and improved the implementation of existing test steps. The test team was then further expanded, and its members were trained in test automation using Selenium and gained the relevant experience. As a result of the skill development of more experienced employees and external recruitment, each development team now has one test team member assigned to it, who is responsible for the manual testing of new features (delivered by that team) and one who is responsible for maintaining test automation for the old, monolithic POL-on, as well as maintaining and developing automated tests for the new POL-on microservices. As a result, individual employees do not have to divide their work time between manual and automated testing, which leads to greater overall productivity.

The architecture of POL-on 2 is reflected in the organisation of automated tests. End-to-end tests have been limited to the role of GUI tests. The business logic tests of POL-on 2 microservices are performed by REST API tests, which also serve as integration tests between these microservices. The framework for these tests is common to all modules. It was written in Java using Cucumber and the REST Assured library, which ensures a high level of consistency in the approach to test creation. The test code is stored in the code repositories of individual microservices or completely separately, away from the test code of other modules. The tests are run independently of each other. Despite their large number, their execution time is relatively short (details are presented in Table 3.3.).

Under this new approach, automated tests serve not only as regression tests, but also as tools that support the automated testing of new system features. To test the business logic layer using REST, a tool for sending HTTP queries is required, and many such tools are available. They differ, among other things, in the amount of time needed to start testing a new system. However, using a separate tool can lead to situations in which the same work is performed twice: a tester using the separate tool a test developer using the test framework. Therefore, whenever possible, test developers implement the necessary steps in Gherkin and Java, and the manual testers use them for their current work. The test developers can also temporarily take control of all functional testing of new module features, which is an additional benefit in unusual situations, such as when several members of the test team are absent from work at the same time.

The GUI of POL-on 2 is built using the Angular framework. In this situation, to create automated tests for it in 2019, the test team selected the Protractor framework, which uses Selenium for JavaScript and adds support for Angular components. Cucumber and the Selenium grid described in Section 3.3.1. were also used. The tests were written in TypeScript. The role of these tests is to verify the correctness of the GUI itself (the upper layer of the system) and its communication with the business logic layer (the lower layer). The lower layer is not tested through the upper layer, so these are not end-to-end tests as in POL-on 1. The tests ran in five threads and lasted approximately ten minutes.

In 2022, it was announced that Protractor would not be developed in 2023. The test team was subsequently faced with the need to select a new framework. Initially, all frameworks that require programming in JavaScript and TypeScript were rejected. The experience gained in writing tests for the new POL-on in TypeScript demonstrated that maintaining GUI tests written in a language other than API tests is too costly. None of the team members were skilled in JavaScript or TypeScript, and it would have been necessary to repeat the team's development path for a new language (about which the team had a low opinion), which was considered unprofitable. Using the Playwright framework and writing the tests in Python were considered. Playwright was a young tool at the time, rapidly gaining in popularity. The team members decided that due to the advantages of Python, it would be possible to retrain the team and, in the long term, even write REST API tests using Python. Ultimately, however, this idea was also abandoned. The solution adopted was the Selenide framework with programming in Java. That decision was made in 2022 and all necessary tests were subsequently rewritten. Concurrently, the approach to the Selenium Grid architecture underwent modification. The old grid was retained for the purposes of old tests, and new tests use a grid set up on a Docker container with ten Google Chrome browsers. This change had no noticeable impact on the performance of the testing process.

The POL-on 1 automated tests have not been abandoned. Individual test automations for disabled modules are disabled, but not removed. Attempts to improve their stability and speed have been conducted in parallel with the development of POL-on 2 tests.

The use of a separate test environment for daily testing of the entire test pool has been discontinued. Later, they were performed once a week in the development environment. By disabling unnecessary automated tests, the system testing time in the preproduction environment has been reduced to approximately 100 minutes. Improvements to the test code components associated with the longest wait times and the most-frequently-failing cases have reduced the duration of the testing process to approximately forty minutes. Further deactivation of unused tests has resulted in a reduction in testing time to approximately eighteen minutes. Due to these optimisations, full tests in the development environment took approximately 160 minutes.

3.3.4. Statistics of tests and the effects of greater automation

Table 3.2. presents a summary of the numbers of individual types of artefact produced by the POL-on test team. The values for POL-on 1 should be treated as an approximation of the number of tests that were ever performed in parallel. Some tests were created after others were disabled, some tests were removed, and individual cases were removed from some scenario templates[113]. Finally, some tests were intended to be performed in only one of the test environments. The number of modules is an approximate value, as some tests performed operations in multiple modules, and some larger modules comprised several parts that were separate from the perspective of test development and could also be counted separately. Tests for POL-on 2 are being added gradually. The table does not contain data on tests of other systems related to POL-on, such as the ORPPD[114].

Table 3.2. Number of end-to-end tests

	POL-on 1	POL-on 2
Number of modules for which tests were created	34	14
Number of test files	906	35
Number of scenarios	550	26
Number of scenario outlines	1329	136
Number of examples	5309	463
Total number of test cases	5859	489

[113] The Listings 3.1.–3.3. presented above contained scenario templates: tests with the same structure for a host of different test data arranged in a table marked with the keyword ‘Examples’. There are also simple scenarios for which no examples are given

[114] The ORPPD GUI tests did not use the same framework and did not use any of the Cucumber components, while the RPPD, which is replacing the ORPPD, has a common interface with POL-on 2, and its GUI tests are created as part of the POL-on 2 interface tests

In POL-on 2 tests, the number of automated test cases per module (approximately thirty-five) is much lower than in POL-on 1 (over 170). This demonstrates that significantly fewer tests are now being created at this level than in the previous version of POL-on.

Table 3.3. presents a summary of POL-on 2 modules (including the RPPD) with information on the number of automated test cases for REST API layer tests and their approximate execution time. The Employees module tests are performed across four separate processes due to their long execution time and the fact that it is unnecessary to perform all tests on every run.

Table 3.3. Number of REST API tests

Module	Number of test cases	Approximate test duration [min]
Academic degree granting permissions	53	0.5
Arts	636	54
Doctoral schools	426	5
Documents	517	3
Employees 1	302	16
Employees 2	842	15
Employees 3	931	21
Employees 4	15	2
Fields of study	715	64
Finance	80	0.5
Foreigners	1481	9
Institutions	1699	31
Investments	79	0.5
Managers	252	8
NAWA	70	0.5
Patents	124	3
PhD students	983	39
Projects	56	1
RPPD	1178	18
Rector’s statements	124	2
Specialised education	31	0.5
Students	736	38
Users	291	3

The total number of test cases at this level of testing (excluding the RPPD, as the old version of the tests were separate from the old POL-on tests) is 10,443 with nineteen modules (i.e. almost 550 test cases per module), which is much more than the average in end-to-end tests in POL-on 1. The microservice test coverage of POL-on 2 is therefore greater than that of the old POL-on 1 modules. Simultaneously, the time needed to perform the full tests is shorter—partly due to the lower time consumption of REST tests compared to end-to-end tests, and partly due to the division of the system into microservices and the implementation of modules independently of each other. However, as tests are performed independently, the time comparison could be misleading.

The features of the POL-on 1 and POL-on 2 testing processes described in this subsection demonstrate the benefits of placing greater emphasis on the correct and extensive automation of the test team's work and reducing the number of tests on the GUI at the cost of introducing functional tests at a lower level. These observations align with the antipattern of the ice-cream cone and the pattern of the test pyramid ([19, 77], Figure 3.9.), which are well known in the literature of the testing industry. The widths of the layers in Figure 3.9. illustrate the number of automated tests in a given layer. The unit testing layer was not considered in this subsection. The integration testing layer did not exist in the case of the monolithic POL-on 1. The majority (too many) were manual tests, supported by end-to-end tests; this model was our own version of the test cone. The deliberate actions of the test team and of the wider POL-on development team aimed to implement a pyramid model, in which the interface tests act as a mere supplement to the tests in lower layers. This work was successful, benefiting the entire POL-on project.

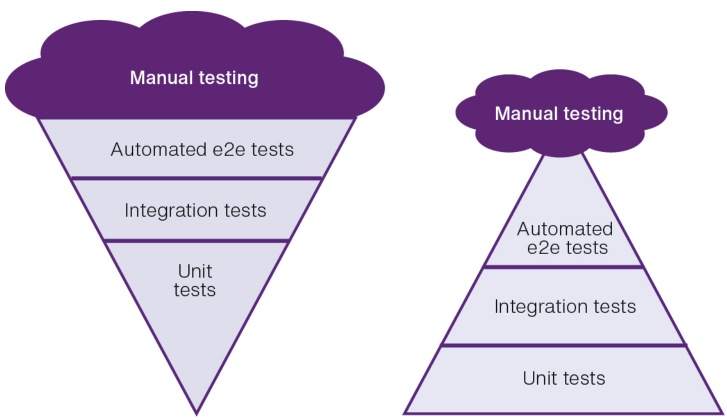

Figure 3.9. The ice-cream cone antipattern and the testing pyramid pattern
Source: own work based on alisterscott.github.io/TestingPyramids.html

CHAPTER 4

CONCLUSION

The POL-on system has become a central component of Poland's digital infrastructure for science and higher education; one that enables comprehensive data collection, regulatory oversight, and strategic planning. Its evolution from a monolithic architecture to a distributed, microservice-based ecosystem reflects a broader shift toward scalable, interoperable, and modular public sector IT systems. Looking ahead, several key challenges and development directions are emerging. One of the most significant is the implementation of Poland's national e-Diploma service. This initiative aims to digitalise the issuance and verification of academic diplomas and degrees, ensuring their authenticity, accessibility, and integration with other public services such as the *mObywatel* platform. The successful deployment of this service will require not only technical interoperability, but also robust legal and organisational frameworks to support secure data exchange and long-term document validity. Another area of growing importance is the enhancement of data on the commercialisation of research results. As the role of universities and research institutions in innovation ecosystems expands, there is a need for more granular and timely information on technology transfer, intellectual property, and collaboration with industry. Although POL-on is well-positioned to serve as a foundational source for such data, this will necessitate the development of new data models, reporting standards, and interfaces with external stakeholders. Finally, strengthening POL-on's capacity to support collaboration between science and business remains a strategic priority. In addressing these challenges, the continued development of POL-on must balance technological innovation with regulatory compliance, user engagement, and data quality assurance. The system's success will depend on its ability to adapt to evolving policy goals, support new forms of academic activity, and contribute to the broader objectives of the digital transformation in public administration and knowledge-based economic development.

ILLUSTRATIONS

TABLES

LISTINGS

REFERENCES

[1] A. Asserson, J. Keith G., and Lopatenko, Andrei. CERIF: Past, Present and Future: An Overview. In Wolfgang Adamczak and Annemarie Nase, editors, *Gaining Insight from Research Information*, pages 33–40, Kassel, 2002. euroCRIS, Kassel University Press. hdl.handle.net/11366/131 (accessed 17 November 2025).

[2] Association of Business Service Leaders. Business Services Sector in Poland 2025. Report, ABSL, 2025.

[3] C. Berg. Interoperability. *Internet Policy Review*, 13(2):1–10, 2024. DOI:10.14763/2024.2.1749.

[4] I. Blümel, S. Dietze, L. Heller, R. Jäschke, and M. Mehlberg. The Quest for Research Information. *Procedia Computer Science*, 33:253–260, 2014. DOI:10.1016/j.procs.2014.06.040.

[5] A. Bochenek, editor. *Systemy informatyczne wspierające naukę i szkolnictwo wyższe. OSF: Obsługa Strumieni Finansowania*. Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy, 2021. opi.org.pl/wp-content/uploads/2021/12/OSF-ebook.pdf (accessed 17 November 2025).

[6] H. Bojkowska and M. Deja. Wpływ zintegrowanego systemu informacji o szkolnictwie wyższym pol-on na zarządzanie informacją w systemie nauki i szkolnictwa wyższego. *Praktyka i Teoria Informacji Naukowej i Technicznej*, 24(4):79–91, 2016.

[7] A. Brandolini. Introducing EventStorming: An act of Deliberate Collective Learning. *Leanpub*, 2017.

[8] R. Bryant, A. Clements, C. Feltes, D. Groenewegen, S. Huggard, H. Mercer, R. Missingham, Maliaca Oxnam, A. Rauh, and J. Wright. *Research Information Management: Defining RIM and the Library's Role*. OCLC Research, Dublin, 2017. DOI:10.25333/C3NK88.

[9] B. Burns, J. Beda, and K. Hightower. *Kubernetes: Up and Running: Dive into the Future of Infrastructure*. O'Reilly Media, 2017.

[10] M. Bylina, E. Podwysocki, and M. Michajłowicz. Data governance at the National Information Processing Institute in Poland. In J.-F. Desnos and M. L. Nores, editors, *Proceedings of European University Information Systems Congress 2023*, volume 95 of *EPiC Series in Computing*, pages 79–88. EasyChair, 2023. DOI:10.29007/1rp9.

[11] D. Celjak, Z. Bekić, M. Cundeković, L. Jertec, M. Milinović, and A. Zubić. Dabar - the national infrastructure for digital repositories. In *EUNIS 23rd annual congress: Shaping the digital future of universities: Book of proceedings*, pages 16–24, 2017.

[12] CERIF in Brief. eurocris.org/eurocris_archive/cerifsupport.org/cerif-in-brief (accessed 17 November 2025).

[13] R. Chen, S. Li, and Z. Li. From Monolith to Microservices: A Dataflow-Driven Approach. In *2017 24th Asia-Pacific Software Engineering Conference (APSEC)*, pages 466–475, 2017. DOI: 10.1109/APSEC.2017.53.

[14] P. de Castro, K. Shearer, and F. Summann. The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. *Procedia Computer Science*, 33:39–46, 2014. DOI: 10.1016/j.procs.2014.06.007.

[15] S. De Santis, L. Florez, D. V. Nguyen, and E. Rosa. *Evolve the Monolith to Microservices with Java and Node*. IBM Redbooks, 2016.

[16] E. Evans. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, 2004.

[17] M. Fowler. Strangler fig application, 2004. martinfowler.com/bliki/StranglerFigApplication.html (accessed 17 November 2025).

[18] M. Fowler. Event sourcing, 2005. martinfowler.com/eaaDev/EventSourcing.html (accessed 17 November 2025).

[19] M. Fowler. Test Pyramid, 2012. martinfowler.com/bliki/TestPyramid.html (accessed 17 November 2025).

[20] M. Fowler. Microservices, 2014. martinfowler.com/articles/microservices.html (accessed 17 November 2025).

[21] A. Frankowska, A. Łobodzińska, M. Czarnocka-Cieciura, P. Drężek-Kłoczko, and J. Bieliński. Analiza stanu otwartej nauki w Polsce. Report, Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy, 2023. radon.nauka.gov.pl/analizy/analiza-stanu-otwartej-nauki-w-Polsce (accessed 17 November 2025).

[22] D. Gagliardi, D. Cox, and Y. Li. What are the factors driving and hindering the adoption of open science? An exploratory study. *Manchester Institute of Innovation Research working paper series*, 10 2014. DOI: 10.13140/2.1.3016.9602.

[23] M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. SAGE Publications Ltd, London, 2010. DOI: 10.4135/9781446221853.

[24] J.-P. Gouigoux and D. Tamzalit. From monolith to microservices: Lessons learned on an industrial migration to a web oriented architecture. In *2017 IEEE International Conference on Software Architecture Workshops (ICSAW)*, pages 62–65, 2017. DOI: 10.1109/ICSAW.2017.35.

[25] N. Gruenpeter. Jak korzystać z zasobów w repozytoriach danych, Uniwersytet Warszawski,2019. drodb.icm.edu.pl/wp-content/uploads/2019/10/Jak-korzystać-z-zasobów-w-repozytoriach-danych.pdf (accessed 17 November 2025).

[26] HIS eG, 2025. his.de/hisinone (accessed 17 November 2025).

[27] ISTQB® Glossary Working Group. *ISTQB® Glossary, V4.6.2*, 2023. glossary.istqb.org (accessed 17 November 2025).

[28] K. G. Jeffery. Cris in 2020. In K. G. Jeffery and J. Dvořák, editors, *e-Infrastructures for Research and Innovation: linking information systems to improve scientific knowledge production: proceedings of the 11th International Conference on Current Research Information Systems : CRIS 2012, Prague, Czech Republic, June 6th–9th, 2012*, pages 333–342, 2012. hdl.handle.net/11366/119 (accessed 17 November 2025).

[29] K. G. Jeffery and A. Asserson. Supporting the Research Process with a CRIS. In A. Asserson and E. J. Simons, editors, *Enabling Interaction beyond the Hanseatic League*, pages 109–119, Bergen, Norway, Leuven, 2006. Leuven University Press.

[30] B. Jörg. CERIF: The Common European Research Information Format Model. *Data Science Journal*, 9:CRIS24–CRIS31, 2010. DOI: 10.2481/dsj.cris4.

[31] E. Kalampokis, E. Tambouris, and K. Tarabanis. Open Government Data: A stage model. In M. Janssen, H. J. Scholl, M. A. Wimmer, and Y.-h. Tan, editors, *Electronic Government. EGOV 2011. Lecture Notes in Computer Science*, volume 6846, pages 235–246. Springer Berlin Heidelberg, 2011. DOI: 10.1007/978-3-642-22878-0_20.

[32] E. S. Karlsen and G. C. Lingjærde. CRIStin in the Norwegian research landscape, 2019. hdl.handle.net/11366/975 (accessed 17 November 2025).

[33] V. Klyuchikov. Language integrated query as a canonical data model for virtual data integration. In *XXI International Conference DAMDID/RCDL'2019 (October 15–18, 2019, Kazan, Russia): Conference Proceedings*, pages 381–394, 2019. ceur-ws.org/Vol-2523/paper37.pdf (accessed 17 November 2025).

[34] J. Kreps. The log: what every software engineer should know about real-time data's unifying abstraction, 2013. engineering.linkedin.com/distributed-systems/log-what-every-software-engineer-should-know-about-real-time-datas-unifying (accessed 17 November 2025).

[35] J. Kreps, N. Narkhede, and J. Rao. Kafka: A distributed messaging system for log processing. In *Proceedings of the NetDB'11 Workshop*, volume 11, pages 1–7, 2011.

[36] M. Manukyan and G. Gevorgyan. Canonical data model for data warehouse. In *East European Conference on Advances in Databases and Information Systems*, pages 72–79. Springer, 2016. DOI: 10.1007/978-3-319-44066-8_8.

[37] R. C. Martin. *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Prentice Hall, 1st edition, 2017.

[38] M. Michajłowicz, M. Kozłowski, M. Kowalski, S. Fijałkowski, J. Drogosz, M. Bylina, and A. Furmankowska-Podniesińska. Scientific activity evaluation in Poland: the IT ecosystem and the optimal selection of achievements. In J.-F. Desnos, R. Yahyapour, and R. Vogl, editors, *Proceedings of EUNIS 2022 – The 28th International Congress of European University Information Systems*, volume 86 of *EPiC Series in Computing*, pages 56–65. EasyChair, 2022. DOI: 10.29007/213t.

[39] M. Michajłowicz, M. Niemczyk, J. Protasiewicz, and K. Mroczkowska. POL-on: The information system of science and higher education in Poland. *European Journal of Higher Education IT 2018-1*, 2018. eunis.org/download/2018/EUNIS_2018_paper_70.pdf (accessed 17 November 2025).

[40] J. Mincer-Daszkiewicz. Success story – 25 years of digitalization of higher education institutions in Poland. In R. Vogl, L. Desnos, J.-F. Desnos, S. Bolis, L. Merakos, G. Ferrell, E. Tsili, and M. Roumeliotis, editors, *Proceedings of EUNIS 2024 annual congress in Athens*, volume 105 of *EPiC Series in Computing*, pages 1–10. EasyChair, 2025. DOI: 10.29007/xq85.

[41] J. Mingers and L. White. Throwing out the baby with the bathwater: The undesirable effects of national research assessment exercises on research. *arXiv preprint arXiv:1502.00658*, 02 2015. arxiv.org/abs/1502.00658 (accessed 17 November 2025).

[42] S. Newman. *Building Microservices: Designing Fine-Grained Systems*. O'Reilly Media, 2015.

[43] S. Newman. *Monolith to Microservices: Evolutionary Patterns to Transform Your Monolith*. O'Reilly Media, 2019.

[44] K. Palavesam and K. Siva Shankar. IRINS: Indian Research Information Network System, 2019. DOI: 10.25610/x923-gd97.

[45] C. S. Pinto, C. Simões, and L. Amaral. CERIF – is the standard helping to improve CRIS? *Procedia Computer Science*, 33:80–85, 2014. DOI: 10.1016/j.procs.2014.06.013.

[46] E. Podwysocki, Ł. Błaszczyk, M. Niemczyk, J. Protasiewicz, M. Michajłowicz, S. Rosiak, and I. Kucharska. Distributed services and a warehouse as an ecosystem on science and higher education. *European Journal of Higher Education IT 2019-1*, 2019. eunis.org/download/2019/EUNIS_2019_paper_22.pdf (accessed 17 November 2025).

[47] E. Podwysocki, M. Bylina, J. Raczko, M. Ulaniuk, M. Michajłowicz, and R. Jendrzejewski. Ludzie Nauki: Data aspects of the Polish national CRIS. In R. Vogl, L. Desnos, J.-F. Desnos, S. Bolis, L. Merakos, G. Ferrell, E. Tsili, and M. Roumeliotis, editors, *Proceedings of EUNIS 2024 annual congress in Athens*, volume 105 of *EPiC Series in Computing*, pages 94–107. EasyChair, 2025. DOI: 10.29007/23qb.

[48] Projekt ustawy zmieniającej ustawę o zmianie ustawy - Prawo oświatowe oraz niektórych innych ustaw, ustawę - Prawo oświatowe oraz ustawę - Prawo o szkolnictwie wyższym i nauce, 2025. gov.pl/web/premier/projekt-ustawy-zmieniajacej-ustawe-o-zmianie-ustawy---prawo-oswiatowe-oraz-niektorych-innych-ustaw-ustawe---prawo-oswiatowe-oraz-ustawe---prawo-o-szkolnictwie-wyzszym-i-nauce (accessed 17 November 2025).

[49] J. Protasiewicz, E. Podwysocki, S. Ostrowska, and A. Tomczyńska. Open access to data on higher education and science: A case study of the RAD-on platform in Poland. In S. Bolis, J.-F. Desnos, L. Merakos, and R. Vogl, editors, *Proceedings of the European University Information Systems Conference 2021*, volume 78 of *EPiC Series in Computing*, pages 9–21. EasyChair, 2021. DOI: 10.29007/gz8q.

[50] J. Protasiewicz, S. Rosiak, I. Kucharska, E. Podwysocki, M. Niemczyk, Ł. Błaszczyk, and M. Michajłowicz. RAD-on: an integrated system of services for science – online elections for the Council of Scientific Excellence in Poland. *European Journal of Higher Education IT 2019-1*, 2019. eunis.org/download/2019/EUNIS_2019_paper_20.pdf (accessed 17 November 2025).

[51] M. L. Ramirez, J. T. Dalton, G. McMillan, M. Read, and N. H. Seamans. Do open access electronic theses and dissertations diminish publishing opportunities in the social sciences and humanities? findings from a 2011 survey of academic publishers. *College & Research Libraries*, 74(4):368–380, 2013. DOI: 10.5860/crl-356.

[52] Ranking Uczelni Akademickich 2025. 2025.ranking.perspektywy.pl/ranking/ranking-uczelni-akademickich/ (accessed 17 November 2025).

[53] L. Ribeiro, P. de Castro, and M. Mennielli. Final report: EUNIS–EUROCRIS joint survey on CRIS and IR, 2016. eunis.org/wp-content/uploads/2016/03/cris-report-ED.pdf (accessed 17 November 2025).

[54] L. M. Ribeiro, R. H. Pereira, O. Pacheco, M. Bernardes, and R. T. Martins. Interoperability between information systems of Portuguese higher education institutions. *European Journal of Higher Education IT 2016/1*, 2016. eunis.org/download/2016/EUNIS2016_paper_31.pdf (accessed 17 November 2025).

[55] M. Richards. Microservices vs. service-oriented architecture. Report, O'Reilly Media, 2015. oreilly.com/radar/microservices-vs-service-oriented-architecture/ (accessed 17 November 2025).

[56] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 1 września 2011 r. w sprawie tytułów zawodowych nadawanych absolwentom studiów, warunków wydawania oraz niezbędnych elementów dyplomów ukończenia studiów i świadectw ukończenia studiów podyplomowych oraz wzoru suplementu do dyplomu. Dz.U. 2011 nr 196 poz. 1167.

[57] Rozporządzenie Prezesa Rady Ministrów z dnia 14 września 2011 r. w sprawie liczby członków Centralnej Komisji do Spraw Stopni i Tytułów oraz warunków i trybu ich wyboru. Dz.U. 2011, poz. 1185.

[58] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 22 września 2011 r. w sprawie danych zamieszczanych w ogólnopolskim wykazie studentów. Dz.U. 2011 nr 204 poz. 1201.

[59] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 29 września 2011 r. w sprawie centralnego wykazu nauczycieli akademickich i pracowników naukowych. Dz.U. 2011 nr 207 poz. 1236.

[60] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 23 sierpnia 2012 r. zmieniające rozporządzenie w sprawie warunków prowadzenia studiów na określonym kierunku i poziomie kształcenia. Dz.U. 2012 poz. 983.

[61] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 29 czerwca 2015 r. w sprawie Systemu Informacji o Nauce. Dz.U. 2015 poz. 944.

[62] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 12 września 2024 r. w sprawie danych przetwarzanych w Zintegrowanym Systemie Informacji o Szkolnictwie Wyższym i Nauce POL-on. Dz.U. 2024 poz. 1403.

[63] Science and Technology Committee. Balance and effectiveness of research and innovation spending: Government and UK Research and Innovation Responses to the Committee's Twenty-First Report of Session 2017–19. First special report of session 2019–21, House of Commons, 2020. committees.parliament.uk/publications/1286/documents/11482/default (accessed 17 November 2025).

[64] M. Skelton and M. Pais. *Team Topologies: Organizing Business and Technology Teams for Fast Flow*. IT Revolution Press, 2019.

[65] W. Steinhoff. Mapping the NARCIS aggregated domain to the OpenAIRE CERIF model, 2019. hdl.handle.net/11366/989 (accessed 17 November 2025).

[66] M. Thelwall. Quantitative methods in research evaluation citation indicators, altmetrics, and artificial intelligence, 06 2024. arxiv.org/abs/2407.00135 (accessed 17 November 2025).

[67] A. Tomczyńska, A. Knapińska, and S. Ostrowska, editors. *Information technology systems that support science and higher education. RAD-on: Reports, analyses, data*. National Information Processing Institute, 2023. archiwum2.opi.org.pl/wp-content/uploads/2023/06/RAD-onEnWWW.pdf (accessed 17 November 2025).

[68] A. Tomczyńska, S. Ostrowska, J. Protasiewicz, and E. Podwysocki. Beyond CRIS: A research and higher education information system in Poland. In *Events EUNIS 2023 Annual Conference*. EUNIS, EuroCRIS, 2023. hdl.handle.net/11366/2477 (accessed 17 November 2025).

[69] UNESCO. Study Report on the Feasibility of a World Science Information System by the United Nations Educational. Technical report, UNESCO, UNISIST, Paris, 1971.

[70] Ustawa z dnia 18 marca 2011 r. o zmianie ustawy – Prawo o szkolnictwie wyższym, ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki oraz o zmianie niektórych innych ustaw, 2011. Dz.U. 2011 nr 84 poz. 455.

[71] Ustawa z dnia 15 stycznia 2015 r. o zmianie ustawy o zasadach finansowania nauki oraz niektórych innych ustaw, 2015. Dz.U. 2015 poz. 249.

[72] Ustawa z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce, 2018. Dz.U. 2018 poz. 1668.

[73] Ustawa z dnia 4 kwietnia 2025 r. o zmianie niektórych ustaw w celu wyeliminowania nieprawidłowości w systemie wizowym Rzeczypospolitej Polskiej, 2025. Dz.U. 2025 poz. 622.

[74] V. Vernon. *Implementing domain-driven design*. Addison-Wesley, 2013.

[75] C. Viegas, A. Vasconcelos, J. Borbinha, and Z. Chora. A canonical data model for records management in the Portuguese public administration. In J. Filipe, M. Śmiałek, A. Brodsky, and S. Hammoudi, editors, *Enterprise Information Systems*, pages 210–249. Springer, 2020. DOI: 10.1007/978-3-030-40783-4_11.

[76] R. Vogl, H. Angenent, D. Rudolph, A. Thoring, C. Schild, S. Stieglitz, and C. Meske. "sciebo — theCampuscloud" for NRW. In *EUNIS Journal of Higher Education*. EUNIS, 2015. hdl.handle.net/11366/461 (accessed 17 November 2025).

[77] M. Wacker. Just say no to more end-to-end tests, 2015. testing.googleblog.com/2015/04/just-say-no-to-more-end-to-end-tests.html (accessed 17 November 2025).

[78] L. Waltman. A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2):365–391, May 2016. DOI: 10.1016/j.joi.2016.02.007.

[79] P. Wiktorko, M. Czarnocka-Cieciura, M. Feldy, A. Łobodzińska, B. Pawlik, and E. Witkowska. Zjawisko drop-outu na polskich uczelniach. Analysis, Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy, 2020. radon.nauka.gov.pl/analizy/dropout (accessed 17 November 2025).

[80] wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1):D520–D528, 08 January 2019. DOI: 10.1093/nar/gky949.

[81] A. Łobodzińska, A. Frankowska, O. Iwańska, J. Bieliński, G. Hołdanowicz, M. Walkowiak, and M. Feldy. Ocena dokumentu Polityka Naukowa Państwa przyjętego w 2022 roku. Analysis, Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy, 2024. radon.nauka.gov.pl/analizy/ocena-dokumentu-polityka-naukowa-panstwa-przyjetego-w-2022-roku (accessed 17 November 2025).

NATIONAL INFORMATION PROCESSING INSTITUTE

Minister of Science and Higher Education
Republic of Poland

The National Information Processing Institute

is an interdisciplinary scientific institute and a leader in software development for Polish science and higher education. We knowledge on almost every Polish scientist, their projects, and their research apparatus. Gathering, analysing, and compiling information on the research and development sector allows us to influence the direction of Polish scientific policy. We develop intelligent information systems both for the public sector and for commercial use.

The key areas of research at the institute include: machine learning algorithms, natural language processing algorithms, sentiment analysis, neural networks, discovering knowledge from text data, human-computer trust, computer assisted decision making systems, and artificial intelligence. Our research is driven by interdisciplinarity, and is conducted in laboratories, which employ specialists in various fields Our team of information technology experts is supported by economists, sociologists, lawyers, statisticians, and psychologists. This fusing of different scientific approaches is conducive to in-depth analysis of research issues, and is a driving force for innovation.

ISBN 978-83-63060-29-9

The task is financed by the Minister of Science and Higher Education.