

# Integrated access to data about science and higher education in the context of General Data Protection Regulation

Jarosław Protasiewicz\*, Emil Podwysocki, Sylwia Ostrowska,  
Aldona Tomczyńska†

The National Information Processing Institute, al. Niepodległości 188b, 00-608, Warsaw, Poland,  
[jaroslaw.protasiewicz@opi.org.pl](mailto:jaroslaw.protasiewicz@opi.org.pl), [emil.podwysocki@opi.org.pl](mailto:emil.podwysocki@opi.org.pl),  
[sylwia.rosiak@opi.org.pl](mailto:sylwia.rosiak@opi.org.pl), [aldona.tomczynska@opi.org.pl](mailto:aldona.tomczynska@opi.org.pl)

## Abstract

By building Reports, Analysis, and Data about science and higher education system – RAD-on, we have provided access to public and non-public statistics about science and higher education, which come from multiple separate databases administrated by the National Information Processing Institute under the leadership of the Polish Ministry of Education and Science. These databases gather high quality information about scientists, students, and graduates, who constitute around 11% of citizens of Poland. This kind of information system requires special handling in order to maintain its integrity and continued availability. In this paper, we present the functionalities and the architecture of the citizen's data service allowing users to request correction or deletion of data, to restrict or object to their processing or to transfer data to another administrator. We explain how this service enhances citizen's trust in the system and enables us to provide the most up-to-date, reliable and credible information about science and higher education.

## 1 INTEGRATED ACCESS TO DATA ABOUT SCIENCE AND HIGHER EDUCATION THROUGH RAD-on

Governments around the world embrace digitalization by creating more complex and interlinked information systems. By doing so new useful services are provided to the citizens, helping them save time and resources. By building Reports, Analysis, and Data about science and higher education system – RAD-on‡ (Michajłowicz et al., 2018; Protasiewicz et al. 2019), we have provided access to public and non-public statistics about science and higher education, which come from multiple separate databases administrated by the National Information Processing Institute (NIPI)§ under the leadership of the Polish Ministry of Education and Science\*\*. NIPI develops computer systems and software for

---

\* <https://orcid.org/0000-0002-9204-921X>

† <https://orcid.org/0000-0002-0832-8081>

‡ Website: <https://radon.nauka.gov.pl/> (only in Polish).

§ Website: <https://www.opi.org.pl> (English version available).

\*\* Website: <https://www.gov.pl/web/edukacja-i-nauka> (only in Polish).

the science and higher education sector. Our computer systems gather data about, among other things, scientists, students, graduates and their achievements. It can be estimated that our institute processes information about around 11% of Polish citizens.

All of the databases built by NIPi have been integrated through the RAD-on platform, which can be seen as an example of an open government data system (OGD, see Azeroual et al.). This kind of system requires special attention when it comes to protection of personal data, especially in the context of General Data Protection Regulation. The regulation came into force on 25 May 2018, which was exactly at the time when RAD-on's functionalities were already being developed.

## 2 RAD-on's SERVICE OF CITIZEN'S DATA ACCESS

The scale of the integrated systems generates enormous potential in terms of data analysis. Open access to data is the key to the development of a data-driven policy in the field of R&D, whereas non-public statistics may be used by administration offices, e.g. in order to validate information about citizens' educational paths. However, the process of data opening or sharing underlines the importance of personal data protection. To stay in line with the General Data Protection Regulation, the IT infrastructure should include functionalities that enable users to access, correct or delete data, to restrict or object to their processing, or even to transfer data to another administrator.

In the RAD-on system these functionalities are available through the service of citizens' data access, which consists of:

- Data warehouse which combines, deduplicates and aggregates data about science and higher education from all domain systems. Personal data from various transactional systems are stored and sent to the data warehouse and integrated into a person's profile. This ensures an integrated access to citizens' data.

- Central authentication module which provides access to the citizen's data service. It is linked to the trusted profile provided through the Polish National Electronic Identification Node<sup>††</sup>. This node provides the identity of a person and our service matches it with data collected by the system. A person is able to use electronic identification means of his or her choice to confirm identity, e.g. identification procedures used in the banking system. Due to the fact that only identification systems that meet specific requirements are connected to the node, a citizen's trust in the system is enhanced.

Our service - as a single point of access - allows users to check whether their personal data is being processed in any of the RAD-on's integrated systems. If so, user can download the report with his or her data to check its correctness and validity. The systems include for example:

- POL-on - system of information about science and higher education,
- Inventorium - system that facilitates science-business cooperation,
- Polish Science - data about Polish scientists and their research,
- SSSR - support system for selection of reviewers,
- PBN - database of Polish scientific publications.

The service of citizens' data access is intended mainly for scientists, experts or students whose data are stored in the abovementioned systems.

---

<sup>††</sup> <https://login.gov.pl/login/main> (only in Polish).

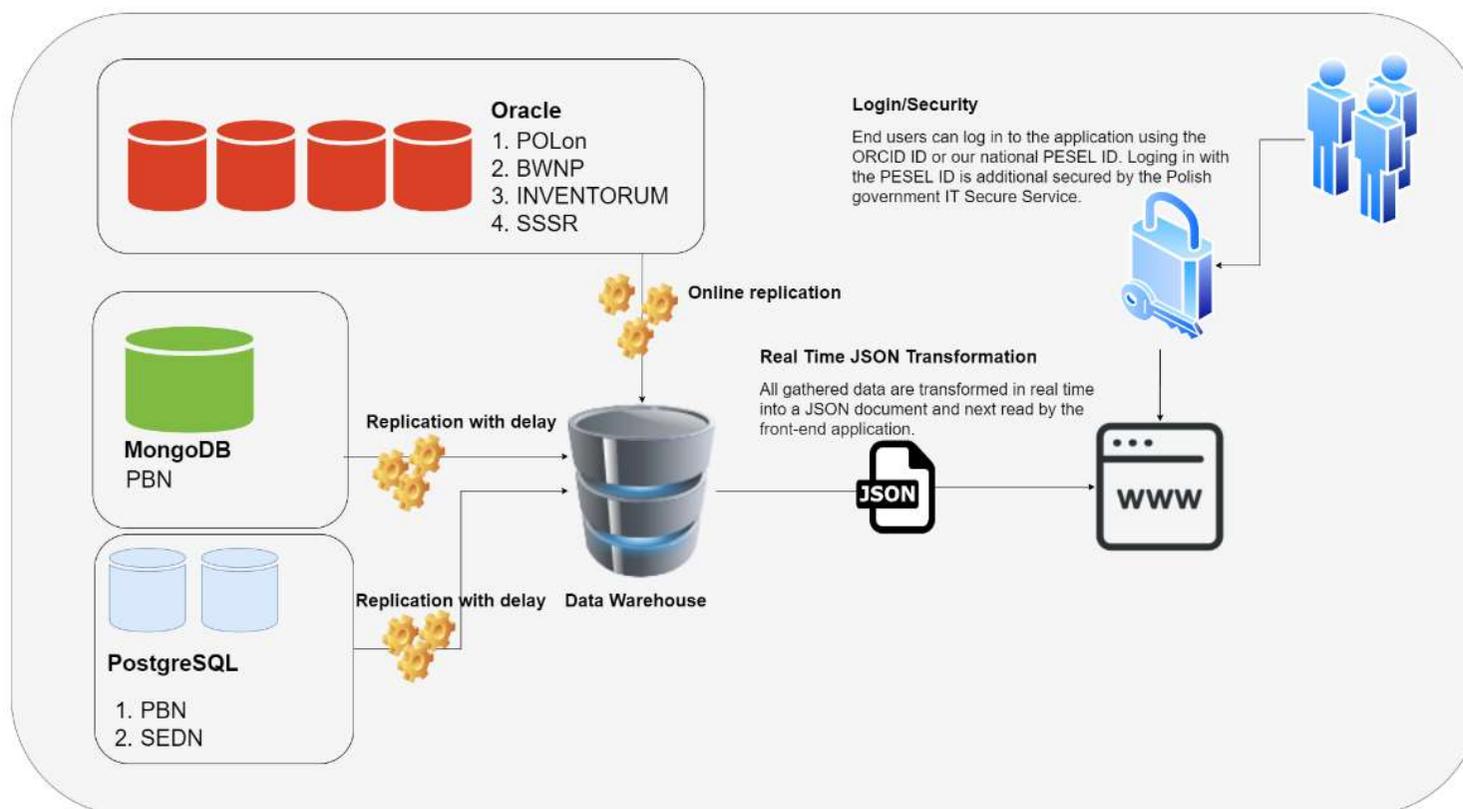


Figure 1. A simplified architecture of 'citizen's data access' service

Abbreviations and detailed information on data source layer: POL-on– system of information about science and higher education, Inventorum – system that facilitates science-business cooperation, Polish Science – data about Polish scientists and their research, SSSR – support system for selection of reviewers, ORPPD – national repository of written diploma theses, PBN – database of Polish scientific publications, SEDN – System for Evaluation of Scientific Achievements in Poland.

The service is available through the user account. As we want citizen's personal data and non-public information to be safe, we need to confirm person's identity before gaining access to the data. We use the Polish National Electronic Identification Node to be sure that data are made available to the person to which they pertain. After logging in, users see a list of systems in which their personal data is being processed (the list may differ depending on whether it is a student, reviewer or author of publications). In some cases, no system will be available, which means we have not found this person's data in the RAD-on source systems. Users can order and download individual reports containing their data from every system. The reports include, among other things, data on the course of studies and awarded grants, data on projects and scientific publications, employment history, and awarded titles and degrees. Files can be downloaded within 24 hours in two formats: PDF (recommended for presenting data in an organized manner), JSON (machine-readable format recommended for further processing e.g. for analyzing or transferring data).

After downloading the data, the citizen can verify it and, if necessary, react to errors in their data so that the period from reporting to updating them is as short as possible. After selecting the "Submit data request" option user is redirected to the the National Information Processing Institute Helpdesk system to specify the scope of requested changes. The processing of the request depends on a system (e.g. not all systems allow removal of data). Each request concerning personal data is handled individually. Updated data feed the RAD-on's integrated systems, therefore citizens have a direct impact on the correctness of the science and higher education data in Poland.

By implementing our service of citizen's data access, which ensures instant correction of data, we are able to provide the most up-to-date, reliable and credible information about science and higher education. At the same time, we are able to fulfill obligations of the General Data Protection Regulation, and lay the groundwork for further digitalization of public administration.

## References

Azeroual, O., Saake, G. & Wastl, J. Data measurement in research information systems: metrics for the evaluation of data quality. *Scientometrics* 115, 1271–1290 (2018).  
<https://doi.org/10.1007/s11192-018-2735-5>

Michajłowicz M., Niemczyk M., Protasiewicz J., Mroczkowska K. (2018). POL-on: The Information System of Science and Higher Education in Poland, In EUNIS 2018 Congress Book of Proceedings.

Protasiewicz J., Rosiak S., Kucharska I., Podwysocki E., Niemczyk M., Błaszczyk Ł., Michajłowicz M. (2019). RAD-on: An integrated System of Services for Science - Online Elections for the Council of Scientific Excellence in Poland, In EUNIS 20189 Congress Book of Proceedings.

### 3 AUTHORS' BIOGRAPHIES



**Jaroslaw Protasiewicz (PhD)** is a director of the National Information Processing Institute and an assistant professor at the same Institute. He received a Ph.D. in computer science at the Systems Research Institute of the Polish Academy of Sciences. His areas of interest include agile project management, software design and development, big data, machine learning, and bio-inspired algorithms.

Email: [jaroslaw.protasiewicz@opi.org.pl](mailto:jaroslaw.protasiewicz@opi.org.pl)



**Emil Podwysocki (MSc)** received a master's degree in telecommunications systems at the Technical University of Lodz. He has 10 years of professional experience related to ETL/ELT, data warehouses and business intelligence. His areas of interest include Oracle technology, big data, business intelligence and data visualizations. Currently, he works as a Database and Business Intelligence Team Leader in the National Information Processing Institute.

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

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



**Sylwia Ostrowska (MSc)**: business analyst and project manager at the National Information Processing Institute. She received a master's degree in Management at the University of Warsaw and completed postgraduate studies in designing information systems at the Warsaw University of Technology. She has professional experience in business analysis and project management in an Agile environment. Currently, she is the project manager of „RAD-on: An integrated System of Services for Science”. Her knowledge in this area is confirmed by following certificates: PRINCE2 Foundation, Professional Scrum Product Owner.

Email: [sylwia.rosiak@opi.org.pl](mailto:sylwia.rosiak@opi.org.pl)

LinkedIn: <http://linkedin.com/in/sylwia-ostrowska-06749b23>



**Aldona Tomczyńska (PhD)**: doctor of philosophy in the field of international political economy. She was awarded her doctorate degree with distinction from the University of Warsaw in 2017. She has worked as a chief researcher at the National Information Processing Institute for over eight years. She conducts research in the area of political economy, innovation, and data science. She also serves as an expert in the field of science and higher education policy. She coordinated multiple scientific projects, participated in evaluation studies and research and support actions from the Horizon 2020 programmes.

Email: [aldona.tomczynska@opi.org.pl](mailto:aldona.tomczynska@opi.org.pl)

LinkedIn: <http://linkedin.com/in/aldona-tomczyńska-phd-8490b718a>