EPiC Series in Computing

Volume 78, 2021, Pages 9–21

Proceedings of the European University Information Systems Conference 2021

Open access to data on higher education and science: A case study of the RAD-on platform in Poland

Jarosław Protasiewicz[1*], Emil Podwysocki[1], Sylwia Ostrowska[1] and Aldona Tomczyńska[1†]

[1] The National Information Processing Institute, Poland

jaroslaw.protasiewicz@opi.org.pl, emil.podwysocki@opi.org.pl, sylwia.rosiak@opi.org.pl, aldona.tomczynska@opi.org.pl

Abstract

In the era of data-driven policies, there is a need to ensure that decision-makers and citizens alike have easy access to the highest quality data. In the creation of RAD-on: an information system comprising reports, analyses, and data on science and higher education in Poland, we have attempted to overcome barriers to decision-making in the field of research and innovation. The system presents open government data (OGD) on subjects that are relevant to science and innovation policy, including: the publication patterns of scientists and academics in Poland; graduates from Polish universities and their incomes; and data on research projects conducted in Poland and financed using national and international funds. RAD-on aims to ensure transparency and participation through interaction with its users and a methodologically coherent approach to data. In this article, we describe the complex architecture of the RAD-on system, and present its reporting capabilities. While doing so, we focus on the development process, which accounted for recent studies on the usability of open data. We address the problem of users' understanding of the possibilities of the data platform, and outline our efforts to mitigate the risk of data being wrongly interpreted and used in decision-making processes.

1 RAD-on as an open and integrated data platform

In the era of data-driven policies, it has become necessary to ensure that decision-makers, citizens, and other stakeholders have access to non-privacy-restricted and non-confidential data, produced

* https://orcid.org/0000-0002-9204-921X
† https://orcid.org/0000-0002-0832-8081

S. Bolis, J.-F. Desnos, L. Merakos and R. Vogl (eds.), EUNIS 2021 (EPiC Series in Computing, vol. 78), pp. 9–21

using public funds (Ruijer et al. 2020). Governments around the world have embraced digitalization by creating increasingly complex and interlinked information systems. Although many sources of data are involved in this process, open government data (OGD) is of particular importance, owing to its scale and status as a source of credible information on a wide range of subjects.

How individuals access and use open data is influenced by the manner in which it is published (Dombrowski et al. 2012). In raw form, open data often fails to be useful for end users (Weerakkody et al. 2017). As Huston, Edge and Bernier (2019) assert, the true benefit of open data is not simply that individual databases can be used more widely, but that their data can be combined with that from other sources to present more contextual, in-depth analysis.

Over the last two decades, the science, innovation, and higher education sectors of many countries have begun to experience a transformation as new forms of digital data are generated, analysed, and utilised to inform decision-making processes (Williamson 2018). Modern IT systems focus not exclusively on data gathering, but also on introducing interoperability standards, dashboards, and visualisations. One example of such an IT system in Europe is that of the Higher Education Statistics Agency in the United Kingdom (HESA) ‡, which offers information on all aspects of the country's higher education landscape, including data on students, staff, and graduates. HESA's strategy for 2016–2021 involved upgrading the data infrastructure, improving sectoral data capability, and enhancing insight with the use of business intelligence (BI), next-generation data analytics, and visualisation technologies (HESA, 2016). Although analytics systems for higher education have been constructed in other countries for similar purposes (e.g. Williamson 2018, Khatibi et al. 2020, see also Samuelsen et al. 2019), they remain less advanced.

In Poland, the Information System for Science and Higher Education (POL-on) has been in development since 2011. It aims to establish a national database of scientific institutions, universities, and students. It supports governmental procedures, including the evaluation of scientific achievements and the distribution of funds for higher education. Some data stored in POL-on is publicly available through its user interface; the system, however, does not provide the facilities to process data automatically, visualise aggregated data, nor comment on its meaning for the science sector. Moreover, it does not integrate fully with other scientific open access databases. As a result, the potential for more comprehensive analysis using POL-on is limited.

In the creation of RAD-on – an information system comprising reports, analyses, and data on higher education and science § (Michajłowicz et al. 2018; Protasiewicz et al. 2019) – we have attempted to bridge the technology gap and offer enhanced support for decision-making in research and innovation. The uniqueness of the system stems from its provision of access to public and non-public statistics on science and higher education from multiple databases administrated by the Polish National Information Processing Institute** (NIPI), including from the POL-on system. The institute has developed computer systems and software for the science and higher education sector since 1991. Its modern computer systems gather data on scientists, students, graduates, research projects, and scientific publications. It can be estimated that the institute processes information on around 11% of Polish citizens. Integrating all of the non-privacy-restricted data sourced from the institute's databases, has enabled us to access the largest sources of information on research and innovation, science, and higher education in Poland, in addition to ensuring wide, free, and comfortable use of that data. The novelty of RAD-on can be found in the following aspects:

- RAD-on integrates data on higher education and science in Poland that is sourced from various autonomous databases – the system currently processes 11.22 TB of data;
- RAD-on delivers mostly open, and occasionally restricted, access to the most up-to-date and reliable data on research and development (R&D), science, and higher education

‡ Website: https://www.hesa.ac.uk.
§ Website: https://radon.nauka.gov.pl/ (English version available).
** Website: https://www.opi.org.pl (English version available).

through an application program interface (API) and an independent user-friendly business intelligence system;
- RAD-on ensures the reliability of its data by facilitating its correction or deletion with the use of a single point-of-entry by any individual whose data is stored in any of the integrated databases;
- RAD-on supports decision-making processes by delivering data analysis tools and descriptions that enable easier interpretation of data.

In this article, we examine the architecture, business process, and functionalities of the RAD-on system, which has only recently been introduced in Poland in its fully developed form. We present the platform's functionalities and potential data delivery, creation of statistics, and more complex analyses that align with best practices of Open Government Data (OGD). Furthermore, we focus on best practices in the creation of OGD platforms that aim to increase users' satisfaction from data exploration (see Weerakkody et al. 2017). More specifically, the goals of this article are:

- to present the rationale behind the RAD-on system and discuss the environment of the project and its constraints;
- to disseminate the project's results and its impact on scientific, educational, and business processes;
- to discuss further development plans and the possible impact of RAD-on in a European context.

The remainder of this article is structured as follows: Section 2 explains the goals of the project; Section 3 describes the environment in which the project occurred, including the technical, administrative, and other project constraints; Section 4 presents the key functionalities of RAD-on and the process of their development; and Section 5 presents the most relevant results and their impact on the R&D and higher education sector in Poland and Europe. The article concludes by presenting the most significant achievements of the system and the institute's plans for further development, as well as its efforts to increase the profile of the RAD-on platform in Poland and internationally.

2 Goals of the project

NIPI has over 30 years' experience implementing IT projects in science and higher education. Plantin et al. (2016: 4) correctly observed that 'fully developed infrastructures are complex ecologies whose components must continually adapt to each other's ongoing change'. The institute constantly adapts its IT systems to new laws and regulations pertaining to the R&D sector in Poland, and develops new IT solutions to problems as they emerge. The growing number of databases, despite their obvious benefits, necessitated the adoption of systemic solutions to address:

- increased data dispersion and the absence of a central source of access to data for citizens, journalists, scientists, and public administrators;
- poor data consistency and integrity. Data from individual IT systems occasionally conflicts with that of other systems, and identifying the most up-to-date data often proved challenging;
- limited possibilities for analysis and machine data processing. There was no central point of access to raw data and machine data retrieval services, and no uniform data model, in addition to low data openness.

Solutions to these problems are viable only via individual integration processes of a limited range of data from selected systems, or via the ad hoc provision of API services. This demanded the launch

of a new IT platform that would integrate all of the systems and state registers, and make data on science and higher education available in a single place and in an open manner.

To this end, the detailed goals of RAD-on were:

- to facilitate the retrieval of information and metadata on science and higher education by implementing solutions, including a full-text search engine, a wide range of filters, division of data into categories, various forms of data presentation (tables, charts, visualisations, and maps) and analyses (detailed studies for download), and compliance with web content accessibility guidelines (WCAG)††;
- to reduce time-consuming and costly procedures by introducing reporting automation solutions through API services;
- to increase the credibility and consistency of data by ensuring direct access to users' data and the ability to react to observed errors, as well as enabling access to machine editing services for the external systems of authorised institutions;
- to increase the openness and transparency of data and ensure its easy interpretation by providing complex analysis and reports;
- to enable access to raw data with aggregation, which allows users to conduct their own analyses and studies, or to reuse data in innovative applications and services.

3 Project environment and constraints

IT infrastructure grows constantly – not only in simple measures (such as the number of lines of code), but also according to other factors (such as complexity or the degree of interoperability). This can serve as the starting point for a variety of technical problems in the life-cycle of software (Duarte et al. 2021). Additionally, most commercial IT projects are managed with the use of agile methods that follow iterative, test-driven approaches. Research funding organisations that support the development of OGD projects prefer, however, to use traditional waterfall approaches. Both approaches have their merits, but also entail their own models of work and conditions. Combining the two can lead to difficulties. In this section, we elaborate on the technical and administrative challenges in the development of the RAD-on system.

The main technical risk associated with a project such as RAD-on is incorrect selection of the technology behind the data exchange model. This is a key technical output of the project, which ensures stable integration between the source systems and microservices. All of the services offered by RAD-on – such as remote reporting and machine sharing of higher education and science resources – depend on the accuracy of the model selected. We considered three technological solutions: Mule Enterprise Service Bus (ESB), Spring Integration, and Apache Kafka. Each of the technologies were verified according to their distinctive features pertaining to stability, scalability, programming limitations, ease of development, and integration with databases. Based on analysis, we opted to use Apache Kafka – a publish-subscribe, high-throughput system that mediates the exchange of messages between applications (systems) and enables their propagation in real-time. The system was originally built at LinkedIn to serve as its centralised event pipelining platform for online data integration tasks (Vohra 2016). Kafka is available as open software. It is also fault-tolerant, as it can run on any number of servers, regardless of their locations. Apache Kafka-based integration provides a medium for efficient and secure communication between modules and systems; it has built-in mechanisms to ensure replication and protection of messages against loss in the production environment (see Le Noac'h et al. 2017; Hiraman et al. 2018; Wu et al. 2019). In the context of the

†† https://www.w3.org/WAI/standards-guidelines/wcag/

RAD-on project, an undoubted advantage of Apache Kafka lies in its scalability and its impressive record of practical applications.

Even after the technical issues had been settled, however, administrative constraints to the project remained that could undermine its implementation. Although RAD-on was developed using agile methods, it was nevertheless co-financed by European Union funds, which allowed relatively little flexibility in terms of financial and timeline adjustments. The budget and scope of the project, approved at the stage of application for funding, could not be altered for its duration. This led to a host of formal obstacles, from which the purchase of the data warehouse was the most impactful. Soon after the project was launched in late 2017, the institute was informed of a planned increase in hardware and software prices of approximately 10-15%. A lower budget for the data warehouse had been approved by the financing institution, and the expenditure itself was planned for the fourth quarter of 2018. As a result, there was a solid risk of exceeding the budget by approximately PLN 400-600 thousand, which would entail:

- cancellation of the tender process due to insufficient funds, or a reduction in planned purchases;
- failure to implement the project according to the schedule and scope specified in the contract.

To compensate for this risk, the institute opted to conduct a tender process and select a contractor as soon as was possible, before the price increase took effect. Ultimately, NIPI obtained approval for the change in the purchase schedule. It should be emphasised, however, that the mitigation strategy for risk comprised the following two measures:

- constant market monitoring and networking with technology suppliers to obtain hardware and software prices estimations in advance;
- close cooperation between the project partners: the Polish Ministry of Education and Science and NIPI, which allowed swift decisions regarding the purchase, as well as the submission of a budgetary change to the project settlement institution;

The remaining risks associated with the development of RAD-on related to users' satisfaction with the final product. Throughout the design process, we paid close attention to user experience, focusing our efforts on eliminating the potential for data to be misinterpreted. In the next section, we describe the functionalities of the system that ensure the delivery of data that is easily interpretable and reusable in custom-made reports.

4 Functionalities and their development

In order to create the most useful tool for analysing data on science and higher education, we conducted 24 in-depth interviews with stakeholders of RAD-on: representatives of the Polish government and non-governmental organisations, research financing organisations, scientific journalists, and employees of universities in Poland (academics and support staff). This qualitative study offered a comprehensive insight into the functionalities required by the system's users. The interviewees alluded to three key difficulties related to statistics on science and higher education: the dispersion of the data into nine separate databases, the low reliability of the data, and the absence of methodological unity. The current version of RAD-on accounts for the conclusions of studies on open data and its usability (Ruijer et al., 2020).

As mentioned above, the primary goal of the RAD-on project was to integrate all available data on science and higher education in Poland that is held on all of NIPI's databases:

- POL-on[‡‡], which gathers information on all scientific institutions, universities, scientists, and students in Poland. The system acquires its data directly from universities and research institutions;
- The Polish Scientific Bibliography[§§] (PBN), which stores data on the research output of Polish academics;
- he Polish Graduate Tracking System[***] (ELA), which monitors the careers of graduates of higher education institutions (HEIs). The system displays the salaries of graduates of faculties and universities in Poland. It allows prospective students to make more informed choices regarding their studies;
- Inventorum [†††] , which facilitates science-business cooperation using intelligent algorithms. As an active recommender system, it provides tailored information on innovations, projects, innovative enterprises, scientists, and research organisations;
- Polish Science[‡‡‡], which stores data on Polish scientists and their research, doctoral, and habilitation[§§§] dissertations;
- SSSR[****] (Support System for Selecting Reviewers) – an adaptive knowledge database of potential reviewers of scientific papers. It generates rankings of reviewers based on specified criteria;
- ORPPD[††††] (The Polish National Repository of Theses), which collects the full content of diploma theses defended at Polish universities;
- OSF[‡‡‡‡] (The Integrated System of Services for Science/Funding Streams), which was developed for registering and servicing scientific grant applications submitted to research funding organisations in Poland;
- SEDN [§§§§] (The System for Evaluation of Scientific Achievements), which provides comprehensive support for the evaluation process of scientific achievements in Poland. Full integration with POL-on and PBN facilitates the collection and processing of reliable information on scientific and artistic achievements.

[‡‡] Website: https://polon.nauka.gov.pl (English version available).
[§§] Website: https://pbn.nauka.gov.pl/ (only in Polish).
[***] Website: https://ela.nauka.gov.pl (English version available).
[†††] Website: https://inventorum.opi.org.pl (English version available).
[‡‡‡] Website: https://nauka-polska.pl/ (English version available).
[§§§] Habilitation is a qualification required in order to conduct self-contained university teaching.
[****] Website: https://recenzenci.opi.org.pl/sssr-web/site/home (English version available).
[††††] Website: https://polon.nauka.gov.pl/orpd/login (only in Polish).
[‡‡‡‡] Website: https://osf.opi.org.pl/app/ (English version available).
[§§§§] Website: https://polon.nauka.gov.pl/sedn2017/ (only in Polish).

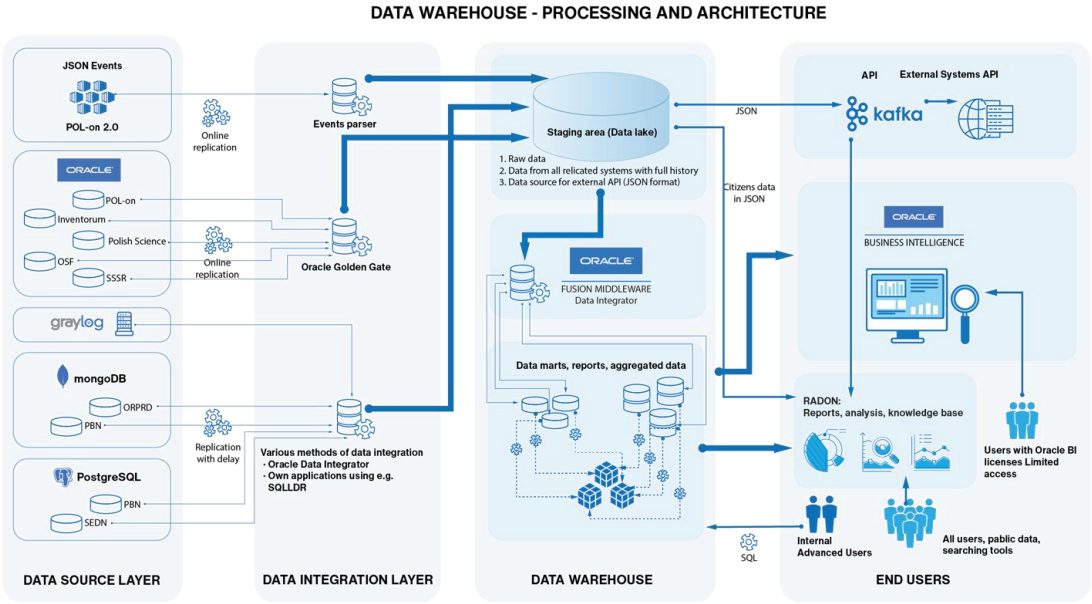

Figure 1. Simplified architecture of RAD-on

Abbreviations and detailed information in the data source layer: POL-on 2.0 – a system of information on science and higher education; Inventorum – a system that facilitates science-business cooperation; Polish Science – a database of Polish scientists and their research; OSF – a system for the management of funding for science; SSSR – a support system for selection of reviewers; ORPPD – a national repository of written diploma theses; PBN – a database of Polish scientific publications; SEDN – the System for Evaluation of Scientific Achievements in Poland.

The dedicated data warehouse combines, deduplicates, and aggregates data from all of the above domain systems. It is organised into logical data processing layers, the first of which is the staging area – a space in which source data from domain systems is stored in an unprocessed form. The second layer comprises tools for integrating the data warehouse with the domain systems – each of which is adapted to the database technology of the domain system. Commercial tools, such as Oracle Golden Gate, Oracle Data Integrator, and DBConvert were used for integration with the domain systems. For integration with systems that store data in non-relational databases, such as MongoDB or ElasticSearch technology, however, it was necessary for the institute to design and implement its own ETL procedures. Next, there is the data mart layer, which contains various data models used by business intelligence tools and other IT applications. We opted for the Data Vault approach in designing the data warehouse, which ensures flexibility and high performance (see, for example, Yessad and Labiod 2016).

The data warehouse is responsible for transforming the source data into useful information through other RAD-on services, including:

1. Dynamic reports for public and non-public use

The reports section of the RAD-on platform is a dashboard that presents statistics in the form of interactive tables and graphs. Users select filters to answer questions such as: ‘How many higher education institutions are there in Poland?’; ‘What courses do they offer?’; ‘How many students are enrolled at each higher education institution?’; ‘How many of them are

foreigners?'; 'Which courses are the most popular among students?'; and 'What salaries might be expected after graduation?'. Each analytical report is complemented by methodological and expert comments. Users can download data in CSV or XLSX formats. They also have the ability to download visualisations as PNG files and reuse them in their own analysis.

For public administration users who are authorised to access non-public data, we offer more advanced BI tools. As a result, the Polish Ministry of Education and Science, as well as other state agencies – including research funding organisations and Statistics Poland – benefit from ready-to-use dashboards tailored to their needs.

2. Complex analysis of the higher education and science sector

Interpretation of statistical data can prove challenging for non-experts, even when only descriptive statistics are presented (Azeroual et al. 2018); therefore, the scientists and experts of NIPI constantly work on more complex analyses, which can be divided into subjects, including science and innovation, higher education, and people of science. Users can download complex reports in PDF format. Currently, RAD-on presents data analyses on artificial intelligence developed in Poland and abroad, in addition to other, more general analyses of the science and higher education system in Poland.

3. Application Program Interface

A uniform programming interface – REST API has enabled the RAD-on databases to be publicly and freely available. It allows users to quickly and efficiently download data that can be useful for conducting analyses and creating statistics, reports, and summaries, as well as developing original solutions and applications that require access to data on higher education.

4. Integrated access to data

In order to facilitate access to data, we have provided RAD-on with an integrated data access point, which comprises:

- a full-text search engine based on the Elasticsearch tool (Andhavarapu 2017) that analyses all of the source databases and is useful in the discovery of semantic correlations between data. Users are able to use the search engine openly, without the need to create user accounts and log in.
- lists of data from different subject categories, such as institutions of higher education and science, scientific and artistic achievements, academic staff, and researchers, as well as promotion procedures, can be viewed using filters and downloaded in CSV or XLSX formats. The data is also available for machine download using the aforementioned API service.

5. A single point of entry for data correction

The service – a single point of access – allows users (who are primarily scientists, R&D experts, or students) to discover whether their personal data is being processed in any of RAD-on's integrated systems. If so, users can download reports containing their data to verify its correctness.

Access to the citizens' data service is provided through a central authentication module provided by the Polish National Electronic Identification Node[*****]. The node verifies the identity of a person; RAD-on then matches this identity with its own data (or more precisely, its source systems). Individuals are able to use the electronic identification means of their choice to confirm their identities – for example, the identification procedures that are used in the banking system. As only identification systems that meet specific requirements are connected to the node, public trust in the system is enhanced. After logging in, users see a list of systems in which their personal data is being

[*****] https://login.gov.pl/login/main (only in Polish).

processed, which differs depending on whether the user is a student, reviewer, or author of publications. In some cases, no system will be available, which means that the user's data has not been found in the RAD-on source systems. Users can order and download individual reports containing their data from every system. The reports include data on courses of studies and grants awarded, projects and scientific publications, employment history, and titles and degrees awarded. The files can be downloaded within 24 hours in two formats: PDF (recommended for presenting data in an organised manner) and JSON (a machine-readable format, recommended for further processing – analysing or transferring data, for instance).

After downloading his/her data, a user can verify it and, if necessary, react to errors so that the time necessary between reporting and updating them is as short as possible. After selecting the 'Submit data request' option, a user is redirected to NIPI's Helpdesk system to specify the scope of the changes requested. The processing of the request varies between systems: not all of them allow the removal of data. Every request pertaining to personal data is handled individually. The updated data feeds into RAD-on's integrated systems, which enables Polish citizens to have a direct impact on the correctness of their data.

5 Impact of the project

In order to evaluate the quality of the services offered by RAD-on, we performed four user-experience studies after each crucial phase of the system's development. Various end-user groups, including those with visual impairments and the elderly, participated in in-depth interviews. We focused on: elements of general user satisfaction with the reports, analyses and integrated data access; satisfaction with the quality of data; and the satisfaction of machine service users. Modifications were made to the system based on the results of these studies.

NIPI constantly monitors the indicators defined for the project and the actual use of its products. A dedicated system for data collection and automatic compilation of statistics has been built and is now accessible via the RAD-on portal[†††††]. From the perspective of the open data policy and the project's goals, the following statistics are of particular relevance:

- amount of public sector data released online: currently 11.3 TB;
- number of downloads of documents containing public sector information per year: 46,841,759[‡‡‡‡‡]. Annual increases in the number of downloads of shared data have been recorded since 2018.

[†††††] Statistics are available at: https://radon.nauka.gov.pl/o-systemie/statystyki.
[‡‡‡‡‡] Value as of 24/08/2021.

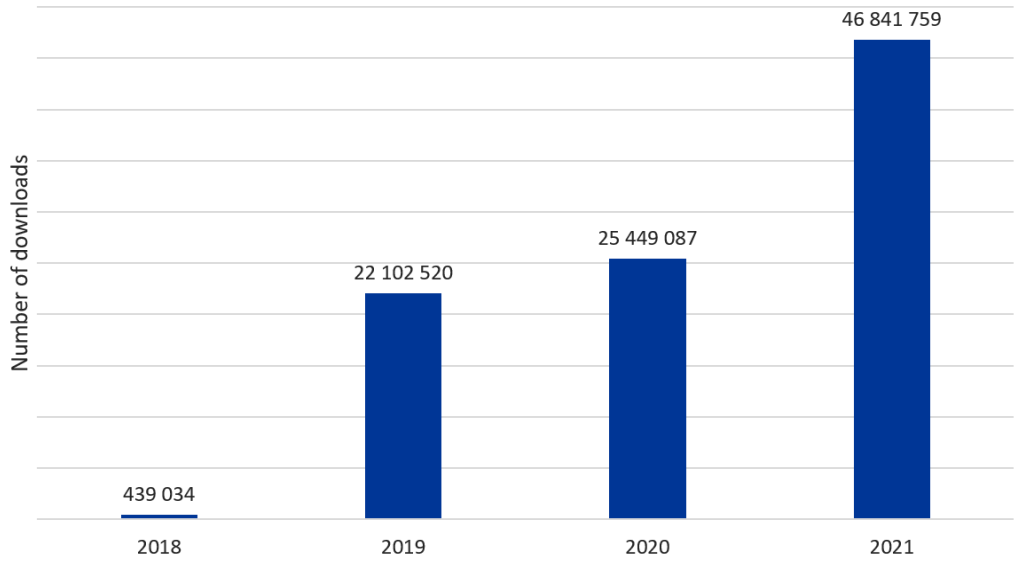

Figure 2: Number of downloads of documents containing public sector information using RAD-on services in 2018-2021

The system was designed with the intention of supporting decision-making processes pertaining to science and higher education. As expected, our analyses and reports are frequently cited by journalists and representatives of governmental bodies. Some of RAD-on's current uses, however, have exceeded our expectations: the platform is used by major universities in Poland to integrate internal systems with the RAD-on Publication API service. The obtained data is processed by these institutions to report on the scientific achievements of their researchers for evaluation purposes. A private foreign company feeds its R&D expert base with data from RAD-on API services on scientists and their publications, projects, and patents. By doing so, the company enhances its product with data on specialists from Poland. Additionally, data from the RAD-on API service on fields of study is used to update The Integrated Qualifications Register (IQR) – a public register of employees' qualifications awarded in Poland.

6 Conclusions

RAD-on is the first fully integrated system for science and higher education in Poland that enables access to governmental and other data from multiple databases. Users can interact with data in a variety of ways, depending on their analytical skills or level of authorisation. Programmers and data scientists benefit from an integrated API, while researchers prefer to download pre-defined tables or visualisations. Decision-makers who are able to access non-public data benefit from more advanced dashboards developed using the BI tool. Journalists and the general public can download comprehensive analyses with in-depth interpretations of data.

The system provides the most up-to-date and credible information available, as the scientists and students whose data is being processed can correct their data through a single point of entry. By implementing this functionality, we were able to fulfil the obligations of the General Data Protection Regulation (GDPR) and lay the groundwork for the further digitalisation of public administration.

Presently, RAD-on comprises approximately 50 reports and analyses, 15 integrated data tables, and a complex API. The institute plans to develop the platform further so that more insight can be offered into complex issues of science policy, such as: the publication patterns of scientists and academics in Poland; the graduates of each specialisation and their incomes; patents and licenses obtained by employees of scientific institutions in Poland; and data on research projects conducted in Poland and financed from national and international funds. In doing so, RAD-on will open the largest set of fully interpreted and integrated data on these subjects to the public. The system will serve as a nationwide, government-supported, reliable source of knowledge on the science sector.

Moreover, the institute has recently offered these resources in English language to ensure the international recognisability of the system. Some components of the system will be reused to create data dashboards presenting international statistics – for example, as part of the GENDERACTION project financed through European Union's Horizon 2020 programme. In order to disseminate know-how acquired through implementation of the project, we took steps to engage in an international initiative that aims to digitalise R&D resources through OGD platforms, such as RAD-on.

References

Andhavarapu, A. (2017). *Learning Elasticsearch*. Birmingham, UK: Packt Publishing Ltd.

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

Dombrowski, U., Mielke, T., & Engel, Ch. (2012). Knowledge Management in Lean Production Systems. *Procedia CIRP, 3*, 436–441. https://doi.org/10.1016/j.procir.2012.07.075.

Duarte, B. B., de Almeida Falbo, R., Guizzardi, G., Guizzardi, R., & Souza, V. E. S. (2021). An ontological analysis of software system anomalies and their associated risks. *Data & Knowledge Engineering, 134*. https://doi.org/10.1016/j.datak.2021.101892.

HESA. (2016). Corporate strategy 2016-2021. Retrieved May 18, 2021, from https://www.hesa.ac.uk/files/HESA_Corporate_Strategy_2016-2021.pdf.

Hiraman, B. R., Viresh M., C., & Abhijeet C., K. (2018). A Study of Apache Kafka in Big Data Stream Processing. *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*, 1-3. https://doi.org/10.1109/ICICET.2018.8533771.

Huston, P., Edge, V. L., & Bernier, E. (2019). Reaping the benefits of Open Data in public health. *Canada communicable disease report = Releve des maladies transmissibles au Canada, 45*(11), 252–256. https://doi.org/10.14745/ccdr.v45i10a01.

Khatibi, V., Keramati, A., & Shirazi, F. (2020). Deployment of a business intelligence model to evaluate Iranian national higher education. *Social Sciences & Humanities Open, 2*(1). https://doi.org/10.1016/j.ssaho.2020.100056.

Le Noac'h, P., Costan, A., & Bougé, L. (2017). A performance evaluation of Apache Kafka in support of big data streaming applications. *2017 IEEE International Conference on Big Data (Big Data),* 4803-4806. https://doi.org/10.1109/BigData.2017.8258548.

Michajłowicz, M., Niemczyk, M., Protasiewicz, J., & Mroczkowska, K. (2018). POL-on: The Information System of Science and Higher Education in Poland. *EUNIS 2018 Congress Book of Proceedings*. Retrieved May 19, 2021, from https://drive.google.com/file/d/1Z-n7ZCJ1FS2r_nF5nCF8-iI4-tQCZbqE/view.

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society, 20*(1), 293–310. https://doi.org/10.1177/1461444816661553.

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. *EUNIS 2019 Congress Book of Proceedings*. Retrieved May 19, 2021, from https://www.eunis.org/download/2019/EUNIS_proceedings_2019.pdf.

Ruijer, E., Grimmelikhuijsen, S., van den Berg, J., & Meijer, A. (2020). Open data work: understanding open data usage from a practice lens. *International Review of Administrative Sciences, 86*(1), 3-19. https://doi.org/10.1177/0020852317753068.

Samuelsen, J., Chen, W., & Wasson, B. (2019). Integrating multiple data sources for learning analytics—review of literature. *RPTEL, 14*, Article 11. https://doi.org/10.1186/s41039-019-0105-4.

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, J. K. (2017). Open data and its usability: an empirical view from the Citizen's perspective. *Information Systems Frontiers, 19*, 285–300. https://doi.org/10.1007/s10796-016-9679-1.

Williamson, B. (2018). The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education, 15*, Article 12. https://doi.org/10.1186/s41239-018-0094-1.

Wu, H., Shang, Z., & Wolter, K. (2019). Performance Prediction for the Apache Kafka Messaging System. *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 154-161. https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00036.

Vohra D. (2016) *Apache Kafka. Practical Hadoop Ecosystem*. Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-2199-0_9.

Yessad, L., & Labiod A. (2016). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. *2016 International Conference on System Reliability and Science (ICSRS)*, 95-99. https://doi.org/10.1109/ICSRS.2016.7815845.

Authors' biographies

Jarosław Protasiewicz (PhD) is the head of the National Information Processing Institute, where he also serves as an assistant professor. He obtained 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

Emil Podwysocki (MSc) obtained a master's degree in telecommunications systems at the Technical University of Lodz. He has ten years' professional experience related to ETL/ELT, data warehouses, and business intelligence. His areas of interest include Oracle technology, big data, business intelligence, and data visualisation. Currently, he works as the head of the Laboratory of Databases and Business Analytical Systems at the National Information Processing Institute.

Email: emil.podwysocki@opi.org.pl

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

Sylwia Ostrowska (MSc): is a business analyst and project manager at the National Information Processing Institute. She obtained a master's degree in Management at the University of Warsaw and completed postgraduate studies in designing information systems at Warsaw University of Technology (2018). She has professional experience in business analysis and project management in an agile environment. Currently, she is the project manager of RAD-on. She holds PRINCE2 Foundation and Professional Scrum Product Owner certificates.

Email: sylwia.rosiak@opi.org.pl

LinkedIn: https://www.linkedin.com/in/sylwia-ostrowska-06749b23

Aldona Tomczyńska (PhD): is a doctor of philosophy in international political economy. She was awarded her doctorate degree with distinction from the University of Warsaw in 2017. She has worked as an assistant professor and data science team leader at the National Information Processing Institute for more than eight years. She conducts research in political economy, innovation, and data science. She also serves as an expert in the field of science and higher education policy. She has coordinated multiple scientific projects, in addition to participating in evaluation studies and research and support actions within the Horizon 2020 programme.

Email: aldona.tomczynska@opi.org.pl

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