Beyond CRIS: A research and higher education information system in Poland

Aldona Tomczyńska, PhD
Emil Podwysocki
Sylwia Ostrowska
Jarosław Protasiewicz, PhD
Reports, Analyses and Data on the science and higher education system in Poland – RAD-on

The National Information Processing Institute and the Ministry of Education and Science in Poland

www.radon.nauka.gov.pl
Project goals
Goal 1. Open access to data on science and higher education

Since 2011, OPI PIB has been engaged in the development of various data collection systems, which operate according to official legal regulations.

Due to the majority of research institutions being obligated to upload data to the systems on a regular basis, the information stored by such systems is the most up-to-date and reliable.
Goal 2. To support decision-making processes by providing IT tools for open government data (FAIR data: findable, accessible, interoperable, reusable) in data-driven policy making.
Public system
RAD-ON
RELIABLE SOURCE OF REPORTS, ANALYSES AND DATA ON HIGHER EDUCATION AND SCIENCE IN POLAND

Search in the system

REPORTS

Reports are based on data imported, among other sources, from the POL-on System of Information on Science and Higher Education as well as from databases, studies and publications put out by the National Information Processing Institute (OPI-PIB).

The user gains access to:
- data on Polish science and results of research in the area of new technologies conducted at OPI-PIB
- interactive maps and charts
- tools which can be used to create their own data summaries

GO TO REPORTS
BI tools for institutions
Business Intelligence Tools for users with additional credentials.

Governmental organizations (ministries, national scientific councils, research funding organizations) who can access non-public data benefit from more advanced dashboards created using the BI tool.
Technology
Overall schema of internal architecture RAD-on
<table>
<thead>
<tr>
<th>Technology</th>
<th>Main task</th>
<th>Justification of the chosen technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java Enterprise Edition</td>
<td>Programming tools</td>
<td>A classic tech stack for developing enterprise-class systems</td>
</tr>
<tr>
<td>JDK 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hibernate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle Enterprise Edition database</td>
<td>Data gathering, aggregating, and sharing</td>
<td>A reliable dataset technology in enterprise-class systems used by the key OPI PIB systems to facilitate data integration</td>
</tr>
<tr>
<td>Oracle Data Integrator</td>
<td>Data integration</td>
<td>ETL process management</td>
</tr>
<tr>
<td>Oracle Golden Gate</td>
<td></td>
<td>Real-mode data replication from source systems (Oracle)</td>
</tr>
<tr>
<td>Oracle Analytics Server</td>
<td>BI tool</td>
<td>Information desktops, self-service reporting, email report distribution</td>
</tr>
<tr>
<td>Oracle APEX</td>
<td>Low-code platform</td>
<td>Fast implementation of web applications and RestAPI interfaces</td>
</tr>
<tr>
<td>Apache Kafka</td>
<td>Message broker</td>
<td>A free, extremely efficient, and distributed platform to manage the exchange of messages between systems.</td>
</tr>
<tr>
<td>Elasticsearch</td>
<td>Data indexing and searching</td>
<td>Implementation of full-text search</td>
</tr>
<tr>
<td>Tomcat</td>
<td>Application server</td>
<td>A free application server</td>
</tr>
<tr>
<td>Graylog</td>
<td>Application monitoring</td>
<td>A free system that analyses application logs</td>
</tr>
<tr>
<td>Swagger</td>
<td>Documentation</td>
<td>Self-documenting Rest API services</td>
</tr>
<tr>
<td>Kubernetes</td>
<td>Orchestration</td>
<td>The most popular tool that manages, automates, and scales container applications</td>
</tr>
</tbody>
</table>
Project impact
Statistics 2022–2023: public system

- 125,000 users
- 78% new users
- 7.3 webpages on average by a user
- 250 items about or references to RAD-on or its data appeared in online media
- 1 item monthly in traditional press
Statistics 2022–2023: public system

11.45 TB of data from 9 systems on science and higher education were integrated and shared.

API was most commonly used to download data on:
- research institutions (105,004,023 downloads)
- scholars' publications (455,962 downloads)
- scholarly staff (193,142 downloads)
- university programmes (106,414 downloads)

150 million downloads of documents (five-fold increase between 2020 and 2021)

Statistics 2022–2023: public system
Statistics 2022–2023: Internal BI tools

- 104 users from 8 public institutions (10 unique users every day)
- 24 thematic dashboards
- 82 unique ETL processes that power the data warehouses
- 2.2 TB of data stored in the data warehouse
- 255 million lines of text in reports downloaded
- 61 thousand reports generated
- 191 REST API interfaces shared
Open Data: API
Application Programming Interface

A uniform programming interface – REST API enables free and public access to the RAD-on databases.

www.radon.nauka.gov.pl
Application Programming Interface

REST API allows users to quickly and efficiently download data that is useful in conducting analyses and in creating statistics, reports and summaries.

www.radon.nauka.gov.pl
Application Programming Interface

Users can develop original solutions and applications that require access to data on higher education.

www.radon.nauka.gov.pl
ABOUT THE SYSTEM

The RAD-on system is a part of the Integrated Network of Information on Science and Higher Education, which is the biggest, in terms of the scope of collected data, public system in Poland supporting the Ministry of Education and Science and other state agencies in shaping the scientific policy.

Statistics include information on the quantity of data that are made available in the system and on how many times every service and every document containing public sector information has been downloaded.

Quantitative indicators for data shared in 2018

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of downloads/views for documents containing public sector information</td>
<td>12345678</td>
</tr>
<tr>
<td>Number of entities that have shared public sector information online</td>
<td>2</td>
</tr>
<tr>
<td>Number of documents containing public sector information shared online</td>
<td>20,977,035</td>
</tr>
<tr>
<td>Number of APIs created</td>
<td>8</td>
</tr>
<tr>
<td>Number of datasets shared online through APIs</td>
<td>14</td>
</tr>
<tr>
<td>Size of public sector information shared online</td>
<td>2,467TB</td>
</tr>
</tbody>
</table>
RAD-on offers the most up-to-date and credible information, as scientists or students can correct their own data through a single point-of-entry.

Journalists and the general public can download comprehensive analyses with in-depth interpretation of data.

Decision-makers who can access non-public data benefit from more advanced dashboards created using the BI tool.

RAD-on is the first fully integrated system for science and higher education 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 and level of authorisation.

Programmers and data scientists benefit from an integrated API; whereas researchers prefer to download pre-defined tables and visualisations.
More about RAD-on
in a book available at:
www opi org pl/wydawnictwo
National Information Processing Institute
al. Niepodległości 188 b
00-608 Warsaw, POLAND

Contact us:
Aldona Tomczyńska atomczynska@opi.org.pl
Emil Podwysocki epodwysocki@opi.org.pl