POL-on: The Information System of Science and Higher Education in Poland

Marek Michajłowicz, Marta Niemczyk, Jarosław Protasiewicz, Karolina Mroczkowska
National Information Processing Institute, Warsaw, Poland

Keywords
information system, higher education and science, decision support system

1. ABSTRACT

The aim of this paper is to introduce the Information System of Science and Higher Education in Poland, POL-on. It is an integrated and centralised information system which supports the Ministry of Science and Higher Education, as well as other ministries and institutions related to science and higher education. The system acquires data from universities and research institutions, analyses these data, and supports the decision-making processes of governmental bodies such as funding, evaluating, controlling, monitoring, and informing. The POL-on system has been continuously developed since 2011. Nowadays, it is a mature and fully deployed system. However, there are still ongoing modifications and improvements that result from the user needs or the law changes. Moreover, POL-on is going to be the essential part of a new integrated system of services for science, which should increase interoperability of the whole information ecosystem with domestic or European institutions or entrepreneurs. As a result, it should increase innovativeness.

2. INTRODUCTION

Both higher education and science have a crucial influence on the modern economy. Universities educate creative and highly skilled people who very often lead organisations and stand behind significant breakthroughs. One may consider innovativeness as a principal initiator of competitiveness and consequently, the wealth of nations. Therefore, each government pays close attention to the condition of higher education and science, which are interwoven and should be considered as inseparable. Governmental institutions have various tools to implement these policies mainly by introducing proper law and regulations. However, policymakers require reliable information. Due to the complexity of the modern world, this can be only delivered by an information system. We have to note that the stakeholders of a system are not only limited to policymakers, but also include research managers, administrators, research councils, technology transfer organisations, media, and the general public. All of them need easy and open access to broad scientific information, which may be gathered in a central system (Blümel et al. 2014). Such a system can also act as a solution to growing reporting requirement, for researchers who often have to enter the same data many times (Biesenbender & Hornbostel, 2016).

Most of the scientific institutions and organisations have some local systems or databases for collecting data. There are also publication repositories and many other sources of scientific information (Ribeiro et al. 2016; Blümel et al. 2014). Thus, the central national information system has to deal with an uneasy problem of gathering data from many different sources. The data are often unstandardized, have different classifications or terminology, and their duplication may occur (Biesenbender & Hornbostel, 2016). Although each country applies its own strategy regarding a central information system of science, those based on Common European Research Information Format (CERIF) by Current Research Information System[1] (CRIS) are worth noting. For example, the Current Research Information System in Norway (CRIStin) supports Ministries in a research-based funding system by providing information about scientific excellence. In addition, the Norwegian Research Council uses the system for discipline evaluation and follow-up projects (Karlstrøm, 2016). The Dutch system NARCIS[2] also uses CERIF (Dijk, 2006) likewise Research Portal of Catalonia[3], which

[1] https://www.eurocris.org
[2] https://www.narcis.nl

additionally utilizes Open Researcher and Contributor ID[4] (ORCID). On the other hand, Flanders Research Information Space[5] (FRIS) uncovered some difficulties coming from the inconsistent semantics of CERIF, which was concurrently applied by different institutions. Thus, an additional semantic layer had to be added to the system to cope with CERIF translations (Vancauwenbergh et al. 2016).

Some systems are more of repositories offering infrastructure and communication interfaces for cooperating scientific institutions. The Digital academic archives and repositories[6] (DABAR) (Celjak et al. 2017) and The National Repository of Theses in Poland (Protasiewicz et al. 2017) utilize the Open Archives Initiative Protocol for Metadata Harvesting[7] (OAI-PMH), which is a well-established technology dedicated for harvesting data to repositories (Ramirez et al. 2013). Yet another architecture of an information system for science is a cloud, in which infrastructure of cooperating institutions is integrated by a software platform. An example of such an approach is Sciebo[8] – the campus cloud for North Rhine-Westfalia in Germany (Vogl et al. 2015) or an interoperability platform, Higher Education Institutions Closer introduced in Portugal (Ribeiro et al. 2016).

In this study, we focus on architecture, business processes, and the environment of an information system of higher education and science, POL-on[9] which has been relatively introduced in Poland recently (Protasiewicz et al. 2017). More specifically, the goals of the study are the following:

(i) to show underlying aims of the central information system of higher information system in Poland, and concurrently discuss its environmental factors and constraints;

(ii) to disseminate the project's results and its impact on the scientific, educational and business environments;

(iii) to discuss the further development plans of the system and its possible impact in the European context.

The Information System of Science and Higher Education, POL-on is the integrated and centralized information system, which supports the Ministry of Science and Higher Education as well as other Ministries and institutions related to science and higher education. Its primary task is to create a global database of scientific institutions, universities, and Polish science. Collected information supports the decision-making processes. Moreover, some data stored in the system are publicly available. Although the system is designed only for Poland, cooperation with other countries may be established. The novelty of this system, as well as this paper, rely on the following aspects:

(i) it supports both higher education and science concurrently as their processes are interwoven;

(ii) it helps to introduce policies on the governmental level regarding financial aspects as well as scientific and educational issues;

(iii) it acts as a central repository of theses, publications, and various information which are utilized to assess research units.

The remainder of this paper is structured as follows. Section 3 explains the goals of the project. Subsequently, Section 4 introduces the environment in which the project takes place, including technical, administrative and financial constraints. Then, Section 5 shows the most important results and their impact on the institutions in Poland. Section 6 presents the present development stage of the project; whereas, Section 7 informs about its possible further evolution, and then, speculates about the applicability of the project to other institutions. Finally, conclusions, references, authors' biographies are presented.

---

[3] https://portalrecerca.csuc.cat
[4] https://orcid.org
[5] https://researchportal.be
[6] https://dabar.srce.hr
[7] https://www.openarchives.org/pmh
[8] https://www.sciebo.de
[9] https://polon.nauka.gov.pl (only in Polish)

3. GOALS OF THE PROJECT

The system acquires and processes data related to almost all aspects of higher education and science in Poland. These data help the government to create and apply more effective policies as well as to provide reliable information required by domestic and European statistical institutions. More specifically, the system supports many aspects of managing science and higher education such as funding, evaluating, controlling, monitoring, and informing (Figure 1)

Firstly, having acquired comprehensive data, the system is able to support the division of various funds. It contains algorithms helping to determine funds which ought to be transferred to each university for educational purposes. It also helps to take decisions on funds for scientific activities of each research unit.

Secondly, it is essential to evaluate the scientific excellence of research units after they consumed the granted funds. The evaluation is based on data describing scientific and deployment achievements such as publications, patents, and alike. These data are collected continuously by the system.

Thirdly, it might be advisable to influence academia to introduce corrective actions based on controlling indicators provided by the system. In detail, it (i) helps to erase the plagiarism cases of diploma theses; (ii) checks the quality of higher education and particularly suitable employment structure of academic teachers and researchers; and (iii) offers the reliable register of scientific diplomas like PhD, professor as well as authenticates diplomas for international academic exchange.

Fourthly, the system monitors students and graduates. During studies, it checks the correctness of financial aid for students. Then, it tracks the graduates' careers to monitor their successes or troubles on the market in respect of their studying fields.

Finally, the system is aimed to inform policymakers, academia workers, students, and the society. For example, it provides reliable educational offers to candidates for universities or implements remote processes of financial and statistical reporting. Moreover, it shares public data, e.g. national repositories of scientific publications, patents, scientific equipment and laboratories, research units, and alike. We can say that the above examples are the practical implementation of the policy of open data.

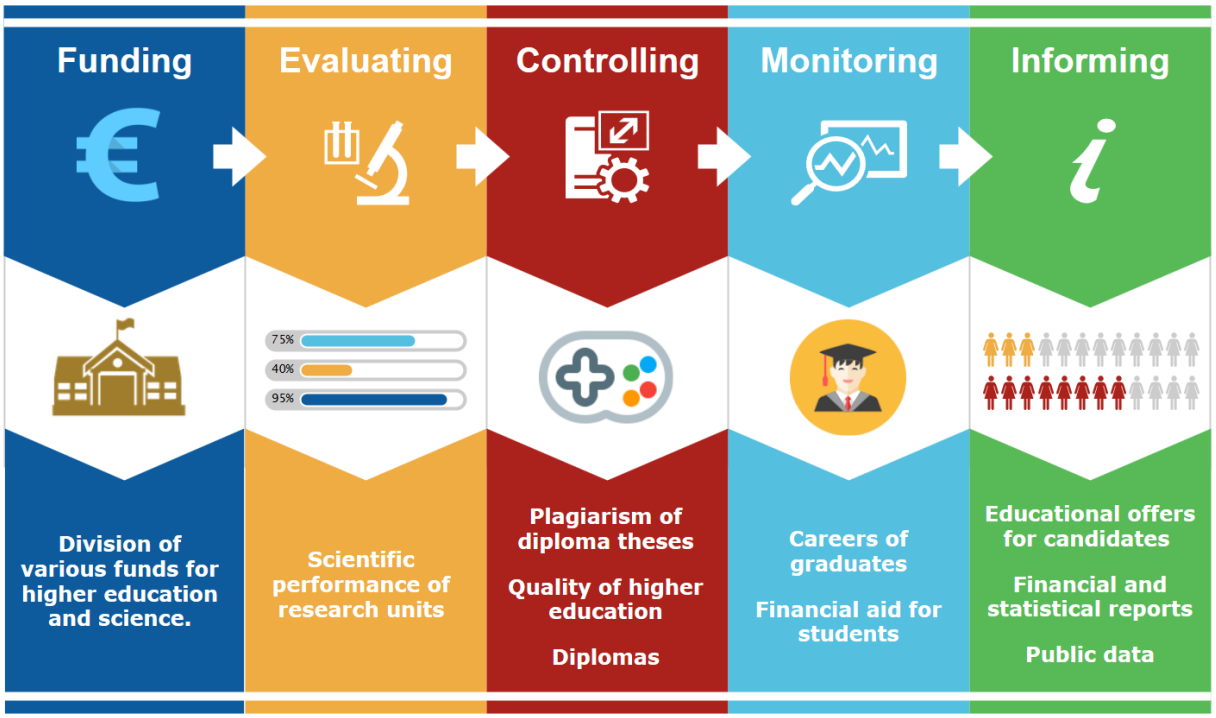

Figure 1. The main goals of the project.

4. ENVIRONMENT OF THE PROJECT AND CONSTRAINS

The project was launched at the beginning of 2011 as a partnership project co-financed by the European Union under the European Social Fund. As a result, the system, POL-on has been developed under the leadership of the Ministry of Higher Education and cooperation of three entities, namely: National Information Processing Institute, Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw, and Index Copernicus Ltd. Since 2014, the system has been maintained and developed only by the National Information Processing Institute. Although the system properly fulfils all goals, it is going to be redesigned due to reforms in Polish science and higher education which are currently taking place. The critical problem is high changeability of the law in Poland. In the last seven years, there have been three amendments to the bill about higher education and science in Poland.

The Polish universities and research units are obliged by the law to provide various information and store it in the central system, i.e. POL-on. The information concerns their educational and scientific activities, namely students, researchers and academic teachers, academic degrees and titles, faculties and studies, diploma theses, projects, patents, publications, conferences, awards, laboratories and their equipment, properties, investments, etc. The data can be transferred to the system by using tools of mass import of XML files as well as RESTfull API (Representational State Transfer, Application Programming Interface) or they may be manually entered via the web interfaces. However, the automatic import is available only for limited kinds of data, mainly information regarding students, researchers, and publications. Regardless of the data delivery method, all records are validated according to the rules based on the law and the best practices. These rules are the key element of the system architecture because they protect and verify in real time the compliance of independent policies of universities and research institutes with applicable law, especially in the area of organization of studies and employment of researchers.

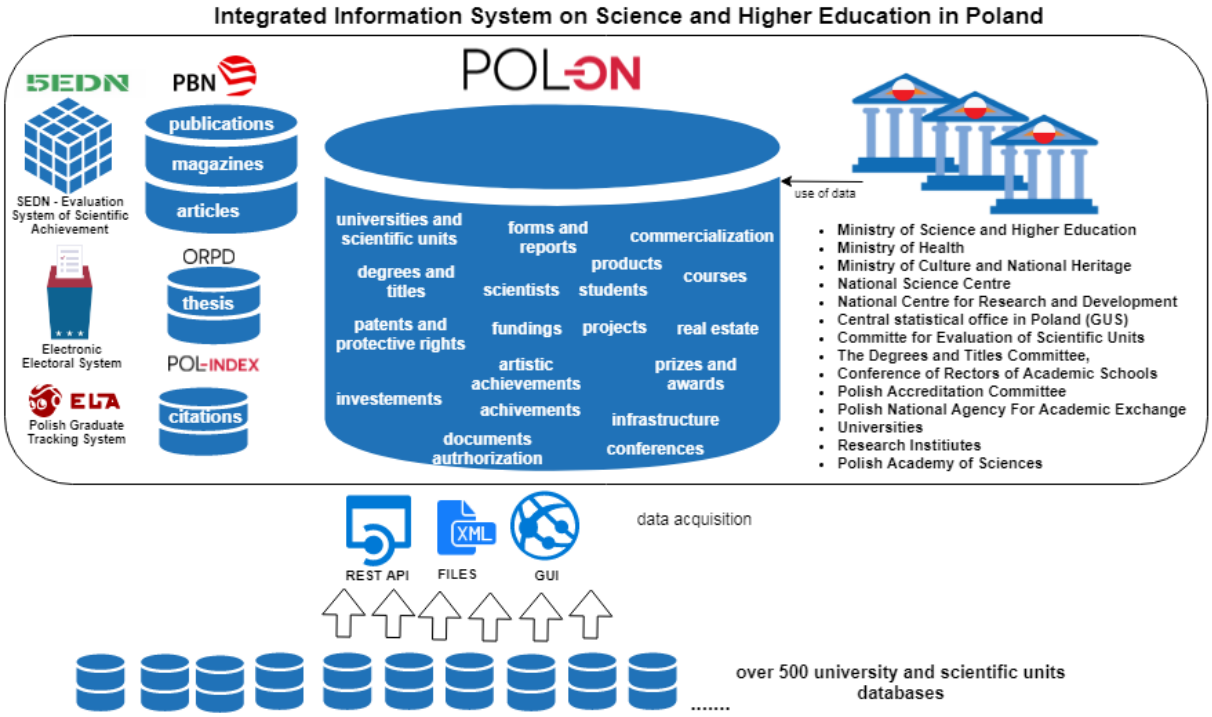

Figure 2. The heterogeneous information environment in which POL-on works.

POL-on works in a heterogeneous information environment which hampers integration in one datacentre. The overall idea of the environment is presented in Figure 2. The core system covers roughly forty modules processing data and accomplishing business tasks. The most important modules are enumerated in the middle of the figure. At the bottom, we see universities and research units, which mostly transfers their data to POL-on. Usually, they have unique information systems in various technologies and architectures such as ERP, antiplagiarism systems, repositories, CRIS systems and alike. Thus, they integrate through interfaces discussed above. Additionally, POL-

on is tightly coupled with six cooperating systems constituting the whole information ecosystem for science and higher education in Poland. These systems are shown in Figure 2 on the left. They are as follows: (i) the Evaluation System for Scientific Achievements (SEND), (ii) Polish Scientific Bibliography (PBN), (iii) Polish Citation System (POL-Index), (iv) Nationwide Repository of Theses (ORPPD), (v) Electronic Electoral System, and (vi) Polish Graduates Tracking System (ELA). On the right side of the figure, we see governmental institutions which utilise provided data, information, and tools supporting decision-making processes. We must note that we pay close attention to supporting users by providing manuals, FAQs, glossaries, e-learning movies, and direct help.

5. RESULTS OF THE PROJECT AND ITS IMPACT ON THE INSTITUTION

The system has an impact on almost all institutions somehow connected with either science or higher education in Poland. Currently, there are nearly 20 thousand registered users originating from universities and governmental institutions, such us the Ministry of Science and Higher Education, the Ministry of Health, the Ministry of Culture and National Heritage,the National Science Centre, the National Centre for Research and Development, as well as the Central Statistical Office in Poland, etc. (see Figure 3).

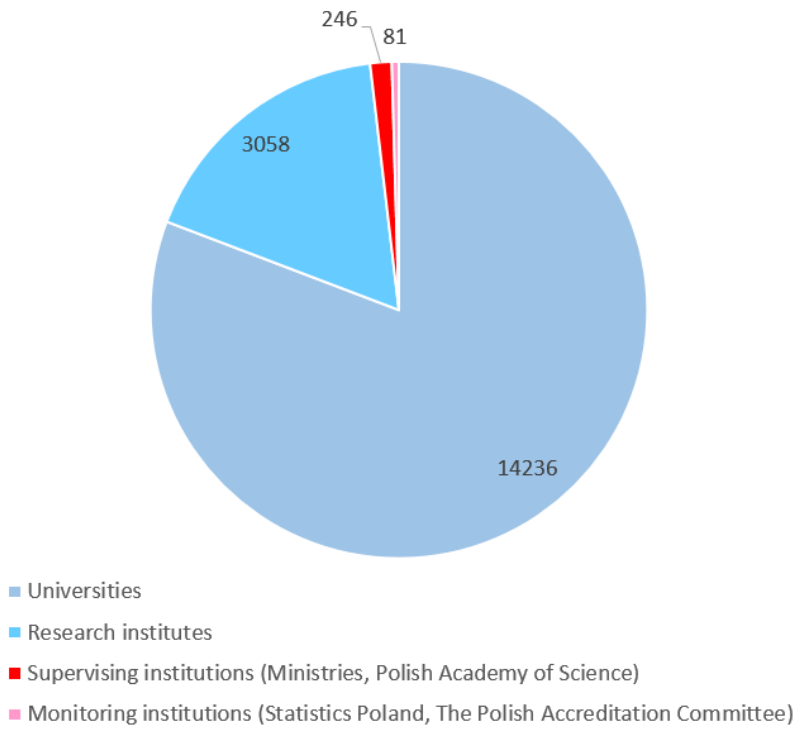

Figure 3. The number of users by an institution type.

Figure 4 covers selected statistics of the system's data. Although it is merely a glimpse, it illustrates the richness and complexity of the system. According to the Central Repository of Public Information, POL-on covers the largest public registers of data. The below statistics reveal:

- number of institutions stored in the system divided by type (universities and scientific institutions)
- number of scientific employees registered in the system, divided by academic teachers and research as well as engineering-technical employees,
- information regarding data stored in the system concerning first and second degree studies, including the number of carried out majors, average number of students in an academic year as well as number of the diploma theses gathered in the system to date,
- information regarding data stored in the system concerning doctoral studies, including the number of such studies and the number of present students of such studies,

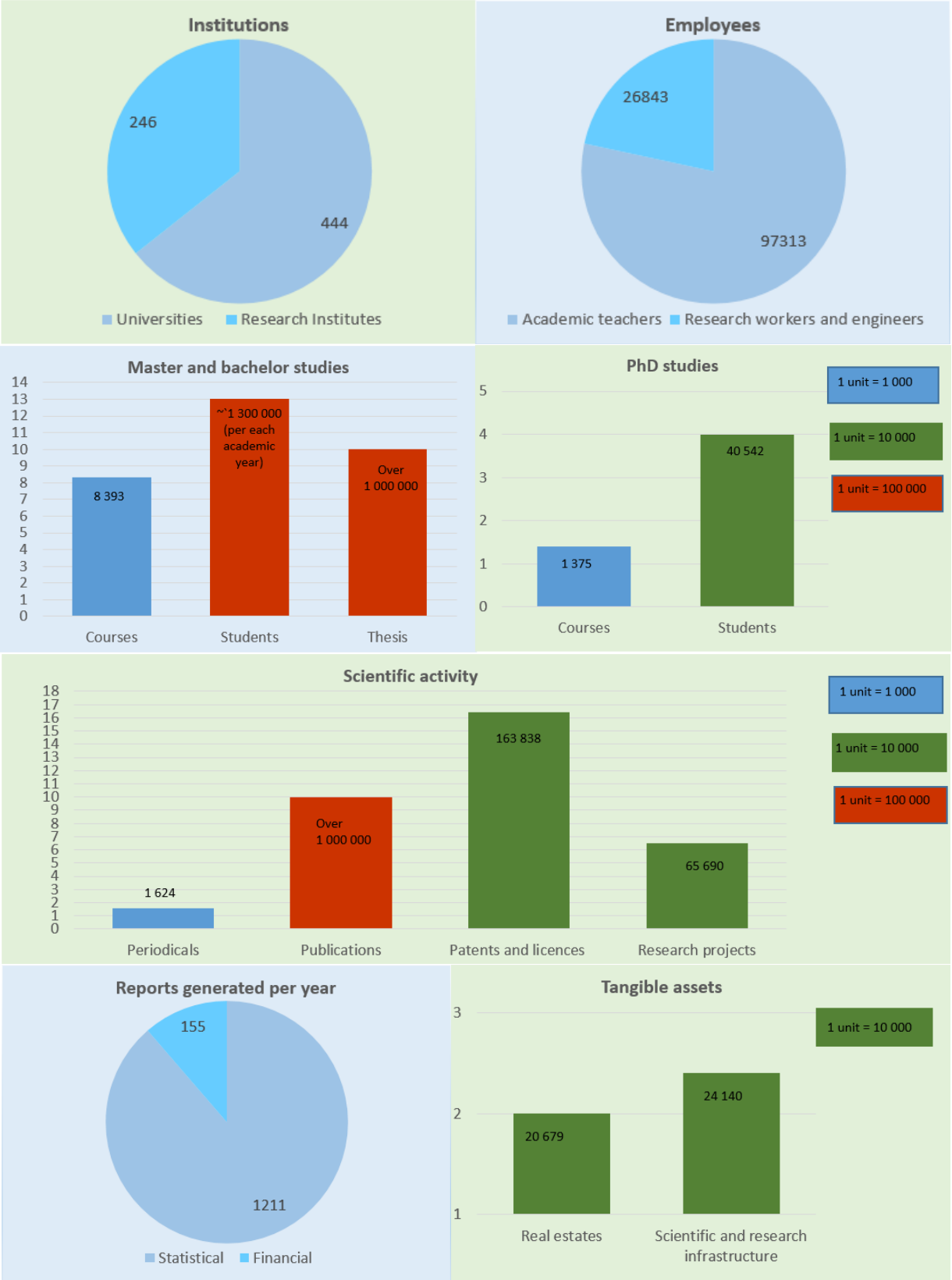

Figure 4. Selected statistics of POL-on as for February of 2018

- information regarding data stored in the system concerning scientific activities of universities as well as scientific institutions: number of currently registered scientific

journals, scientific publications, patents and licenses as well as research projects in the system,

- number of reports generated each year in the system by universities and scientific institutions on the basis of data gathered there, divided by statistical reports (submitted to the Central Statistical Office) and financial reports (submitted to the respective ministries supervising academic and scientific activities),
- information regarding data stored in the system concerning tangible assets held by universities and scientific institutions, divided by real estate and research-scientific infrastructure.

Usage of the system strongly depends on the academic year cycle. In its peaks, it reaches nearly 6 million requests per month (see Figure 5). Intensification of demands sent out by the system is related, above all, to the updates of data on students. For this reason, there is a place (as visible on Figure 5) at the beginning of each new semester.

Figure 5. The number of requests

Pol-on is one of the biggest functioning governmental systems in Poland in terms of the scope of acquired data. It influences the Polish science and education in the following ways:

(i) It reduces bureaucracy by supporting over 20 reporting obligations of universities and research institutes.

(ii) Since 2015, it has been monitoring graduates' careers nationwide using the data system in conjunction with the insurance state register.

(iii) It supports the creation of financial policies by providing strategic data about the Polish science and higher education system.

(iv) It checks the employment quality in higher education, i.e. it automatically verifies multiple employment of researchers and science employees.

(v) It gives the possibility to check compliance with legal requirements of courses.

(vi) It automatically checks the possibility to grant financial aid to students.

(vii) It evaluates all national research units – assesses scientific values that help to establish relevant policies on fund spendings in the scientific field.

(viii) It shares data for public and free use.

(ix) It is going to check plagiarism and frauds by supporting verification of the correctness of students' scholarships, as well as public funds spending.

(x) It is going to be used for the authentication of diplomas for international purposes.

6. STATE OF DEVELOPMENT OF THE PROJECT

The POL -on is implemented based on the agile and iterative AGILE-SCRUM methodology. Each iteration lasts two weeks (see Figure 6). In most cases, a new software version is released after the end of each iteration. Changes in the law determine the system functionalities. By the provisions of the law, system requirements are created as user stories with the particular emphasis on acceptance criteria. Analysts play the crucial role in this process, acting as product owners, and concurrently as designers of detailed technical solutions. In addition, we utilise a methodology, PRINCE2 in our projects to manage them strategically. A flagship example is a project of adjusting the POL-on system to support the process of data collection for the requirements of official statistics in Poland in the years 2015-2017 and, consequently, to support the division of public subsidies.

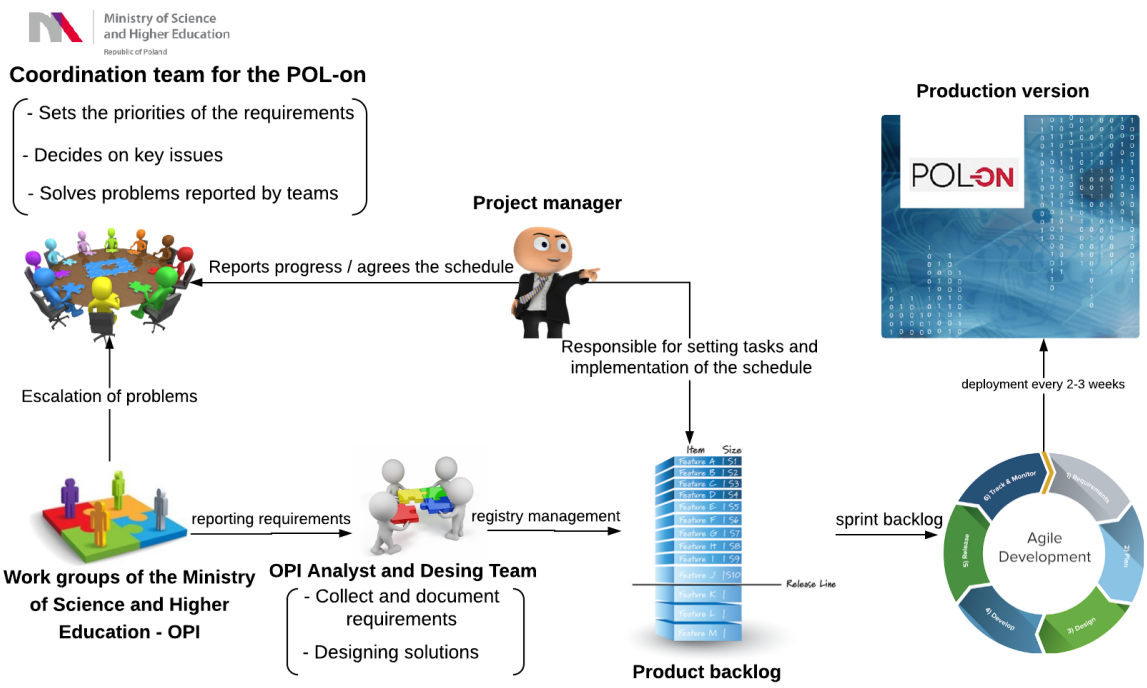

Figure 6. Organization of the project.

The POL-on, in its core version, was developed using a three-tier client-server architecture (see Figure 7). The data are stored in a database, a lightweight container of web applications processes the data, whereas a browser presents data and allows to send requests to an application server. The application is implemented in Java using Spring framework, Hibernate and Java Server Faces. Current work is focused on design and implementation of RESTful API to enable integration with IT systems of universities and research units. Currently, over 20 resources are available[10].

The software development process is very much based on the approach of continuous integration with the use of components such as Maeven, Flyway, Jenkins 2.0 and functional automatic tests Cucumber / Selenium[11]. Since the domains under consideration are complicated, we rely on the Domain Driven Design (DDD) approach.

---

[10] https://polon.nauka.gov.pl/api-rest/

[11] Specification by example approach https://gojko.net/books/specification-by-example/

A relational database plays the crucial role in the architecture of the system, POL-on. The database persists data in a highly standardised structure for convenient handling of critical processes. Currently, the database has reached the size of about 2 TB, without taking into account file data such as theses, publication data, and other files. To control the quality of data at the application layer, numerous validation rules have been implemented, some of which are a direct implementation of the provisions of laws and regulations. The quality control also requires verification of different versions of objects over time and confirmation of historical states. To realise this goal, we utilise an events-oriented hibernate library[12], which covers our modifications.The overall size of all databases, including file data of theses and diploma templates, is over 7 TB.

The existing architecture of the system was based on the approach to the application as a monolith with separate domain modules. Due to the increase in the scale and scope of the system, the process of migration to the micro-oriented model oriented architecture, controlled by the DDD approach, was undertaken.

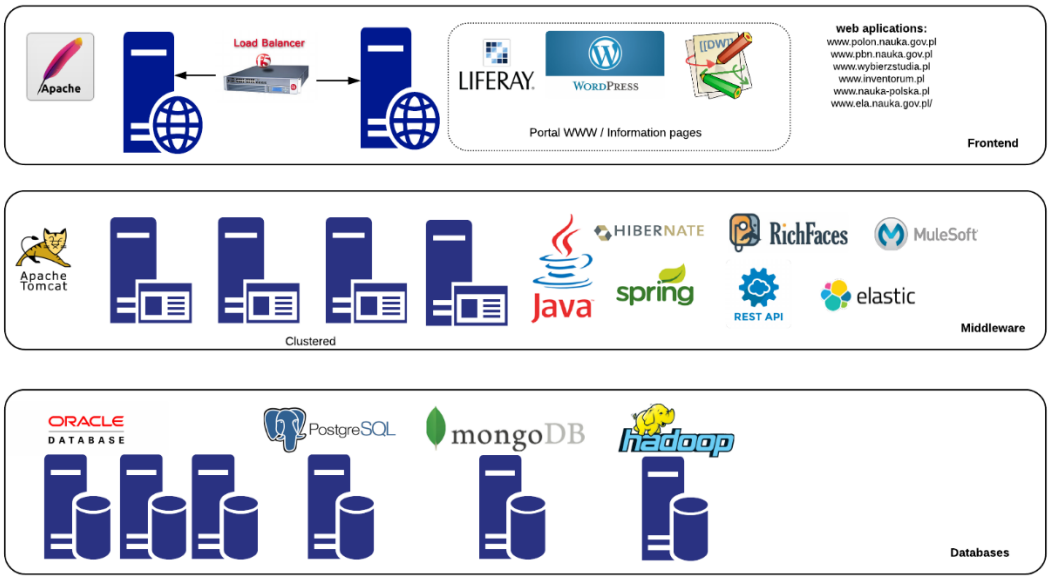

Figure 7. Overall schema of internal architecture of the system.

7. FURTHER DEVELOPMENTS AND APPLICABILITY OF THE PROJECT TO OTHER INSTITUTIONS

The Polish government is going to introduce a new law, named "The Constitution for Science" in the near future. The forthcoming changes in the law create new challenges to the system. The scope of gathered data is going to increase. New documentation repositories are going to be created for the assessment of scientists' advancements and the verification of university documents. The system will also be extended with a wide range of information related to the international exchange of scientists and students. A newly created organisation, the Council for Scientific Excellence requires an information system to support the electoral processes. However, the most significant challenge will be an adaptation of the system, POL-on to a new scientific evaluation procedure using a new classification of scientific disciplines, and concurrently integrating POL-on with ORCID, Crossref[13], and Digital Object Identifier[14] (DOI). The above requirements influence the system so much that it should be redesigned into a new system, POl-on 2.0. Figure 8 depicts the main phases of the project.

---

[12] http://hibernate.org/orm/envers
[13] https://www.crossref.org
[14] https://www.doi.org

In addition, there are development plans that result from the users' needs. Firstly, the most crucial requirement is to establish a warehouse, which will help us to produce sophisticated analyses. Secondly, we have to increase interoperability with other governmental systems in Poland and across Europe. Thirdly, it should share more data in open access. These needs constitute a new project, "The integrated system of services for science" (Figure 9), which has been launched recently.

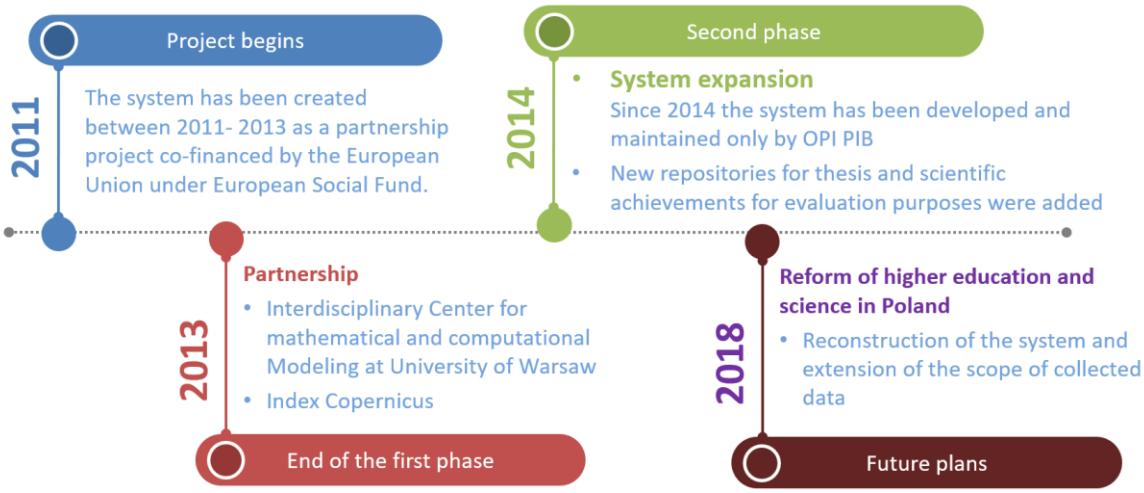

Figure 8. The phases of the project.

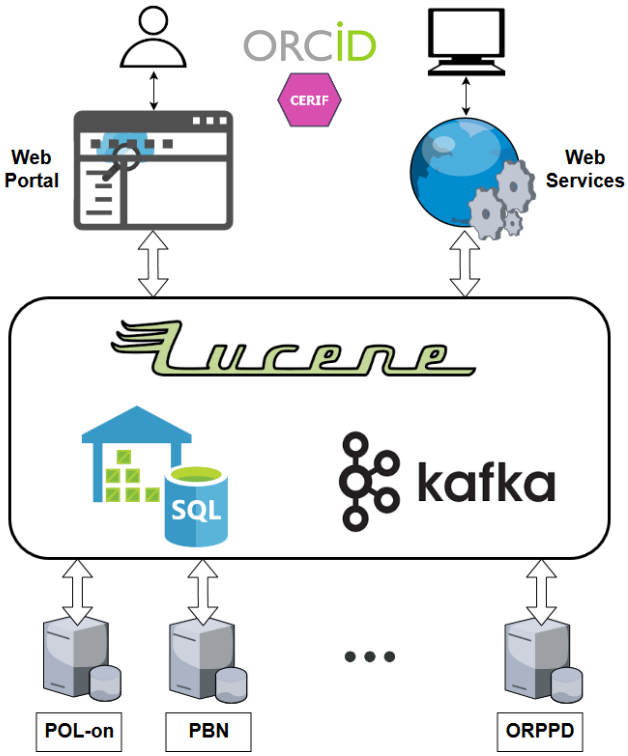

Figure 9. The concept of an integrated system of services for science. The abbreviations POL-on, PBN, and ORPPD, stand for the name of domestic information systems.

Figure 9 outlines the main parts of a new information ecosystem, which should be fully in operation by 2020. Its crucial features are as follows. Firstly, several domestic information systems are going to be integrated through a data exchange model. An open source software, Apache Kafka[15] as an

[15] https://kafka.apache.org

integration platform has been used in this regard. It provides tools for sharing data between these systems. Secondly, the systems' data will be integrated into a warehouse and then transformed to reports or more sophisticated analyses. The warehouse is going to be built with the use of Oracle[16] technology. Thirdly, the access to the system will be twofold. A web portal is dedicated to users; whereas, web services to other information systems. The web portal will allow querying by using a natural language thanks to using a full search engine. We have chosen Apache Lucene[17] was chosen for this purpose. The web services enable automatic storing and retrieving data by authorised machines from the Internet.

The developments described above should increase the applicability of the whole information ecosystem to other institutions. This is understood as the possibility to integrate easily with domestic and European research units, universities, governmental institutions dealing with science and higher education, and entrepreneurs. Technically, the integration could be done by using web services and common standards like the Common European Research Information Format (CERIF) supported by Current Research Information Systems (euroCRIS[18]) and a common researchers' identifier, ORCID.

8. CONCLUSIONS

In this study, we have presented goals, environment, and results of introducing the Information System of Science and Higher Education in Poland, POL-on. Moreover, we have discussed its development issues and possible further evolution.

The system acquires data related to almost all aspects of higher education and science in Poland. These data help the government to fund, evaluate, control, and monitor several management processes in this area. In addition, it supports information policies and shares data.

We have to underline that POL-on is not only a single database, but rather a whole information ecosystem composed of several modules or separate systems. In detail, it acquires data from about five hundred universities or research units. The system itself covers nearly forty modules, where each of them realises one or more business processes. In addition, there are six information systems which closely cooperate with POL-on. Finally, many governmental institutions make use of information processed by this ecosystem.

The results indicate that the system is extensively used both by academia and governmental institutions. Since it covers comprehensive information about science and higher education, it is the valuable tool supporting decision-making processes.

The project POL-on has been developed since 2011. Nowadays, it is the mature and fully deployed system; however, there are still ongoing modifications and improvements. The further developments may include integration with other Polish governmental systems as well as some European systems. As a result, it would not only be a tool for academia but it also should lead to better cooperation between entrepreneurs and researchers, which together with open data may increase innovativeness, and consequently wealthness.

9. REFERENCES

Biesenbender, S., & Hornbostel, S. (2016). The Research Core Dataset for the German science system: challenges, processes and principles of a contested standardization project. *Scientometrics*, *106*(2), 837-847.

Blümel, I., Dietze, S., Heller, L., Jäschke, R., & Mehlberg, M. (2014). The quest for research information. *Procedia Computer Science*, *33*, 253-260.

Celjak, D., Bekić, Z., Cundeković, M., Jertec, L., Milinović, M., & Zubić, A. (2017). DABAR - the national infrastructure for digital repositories, In *EUNIS 23nd Annual Congress Book of Proceedings*

Dijk, E., Baars, C., Hogenaar, A., & van Meel, M. (2006). NARCIS: The Gateway to Dutch Scientific Information. In *ELPUB* (pp. 49-58).

[16] www.oracle.com
[17] https://lucene.apache.org
[18] https://www.eurocris.org

Karlstrøm, N. (2016). More than just a CRIS-How CRIStin is building a national infrastructure for Open Access.

Protasiewicz, J., Stefańczuk, M., & Sadłowski, A. (2017). The National Repository of Theses: A Short Polish Case Study, In *EUNIS 23nd Annual Congress Book of Proceedings*

Ramirez, M. L., Dalton, J. T., McMillan, G., Read, M., & Seamans, N. H. (2013). Do open access electronic theses and dissertations diminish publishing opportunities in the social sciences and humanities? *College & Research Libraries*, 74(4), 368–380.

Ribeiro, L., de Castro, P., & Mennielli, M. (2016). EUNIS-EUROCRIS Joint Survey on CRIS and IR.

Ribeiro, L. M., Pereira, R. H., Pacheco, O., Bernardes, M., & Martins, R. T. (2016). Interoperability between Information Systems of Portuguese Higher Education Institutions. In *EUNIS 22nd Annual Congress Book of Proceedings*

Vancauwenbergh, S., De Leenheer, P., & Van Grootel, G. (2016). On research information and classification governance in an inter-organizational context: the Flanders Research Information Space. *Scientometrics*, *108*(1), 425-439.

Vogl, R., Angenent, H., Rudolph, D., Thoring, A., Schild, C., Stieglitz, S., & Meske, C. (2015). "sciebo—theCampuscloud" for NRW. *EUNIS Journal of Higher Education*.

10. AUTHORS' BIOGRAPHIES

Marek Michajłowicz (MSc) – He is a Deputy Head of the Laboratory of Intelligent Systems. He received a Master degree in Sociology at Cardinal Stefan Wyszyński University in Warsaw and a Bachelor of Engineering (B.E.) in computer science at Warsaw School of Information Technology under the auspices of the Polish Academy of Sciences. He has several years of professional experience related to business and system analysis. Since 2014 he has worked as the project manager of the POL-on system in the National Information Processing Institute. His areas of interest include agile project management, software design and development, big data and warehouses.

E- mail:marek.michajlowicz@opi.org.pl;
Profile: www.linkedin.com/in/marek-michajlowicz

Marta Niemczyk (MSc) – she received a Master degree in Mathematics at University of Warsaw. She has many years of professional experience as a business and system analyst, endorsed by the certificates in methods of IT systems modelling. Currently, she works as software designer and business analyst in the National Information Processing Institute.

E-mail: marta.niemczyk@opi.org.pl

Jarosław Protasiewicz (Ph.D.) – he is an assistant professor at the National Information Processing Institute and the head of the Laboratory of Intelligent Systems. He received a Ph.D. degree 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 and map reduce, machine learning, bio-inspired algorithms. In the period 2011-2013 he was the project manager of the POL-on project.

E-mail: jaroslaw.protasiewicz@opi.org.pl

Karolina Mroczkowska (Ph.D.) - she received a Ph.D. degree in Agricultural Science at the Warsaw University of Life Sciences and ostgraduate diploma in Information Technology - Database Systems Design at Warsaw University of Technology. She has professional experience as a software tester, endorsed by the ISTQB Certificate. Currently, she works as an IT consultant for the global technology company, conducting application support according to the ITIL process.