

Omega-Psir – Institutional CRIS at Polish Universities

Henryk Rybinski

Weronika Kubrak

Łukasz Skonieczny

Jakub Koperwas

Wacław Struk

Abstract:

OMEGA-PSIR is an institutional Current Research Information System, originally developed at the Warsaw University of Technology (WUT). The system is used for presenting the University research in various aspects. It is also used for internal research assessment purposes and is able to report to the Polish Ministry. Since 2013 OMEGA-PSIR has been used as University Knowledge Base¹. The system has been adopted by 12 Polish institutions. A special User Group has been launched for coordinating the system development. At the universities OMEGA-PSIR operates as institutional repository thus being able to support the implementation of Open Access and Open Science. In addition, it is used as institutional CRIS. In the paper we briefly present the main system architecture and its features. We also focus on the experience of WUT, discussing the organizational matters that have to be solved for heterogeneous sources of information, such as distributed data entry for the bibliographic materials, staff activities and achievements, and theses on one side, and information about the research projects on the other side.

Keywords: Research management, Knowledge base, CRIS, Digital library, Institutional repository

Introduction

In 2010, a dedicated project (SYNAT) was launched in Poland in order to address deficiencies in the national scientific information infrastructure. The main SYNAT construction is based on two levels of distributed knowledge bases - with a central database at the highest level, and university databases at lower levels. The ultimate goal of the research knowledge network was to ensure the nationwide dissemination of Polish scientific achievements, and to improve the integration and communication of the scientific community, while also leveraging existing infrastructure assets and distributed resources. The task of finding a solution for the university level in Poland was given to the team of the Warsaw University of Technology (WUT). One of the outcomes of SYNAT was the software OMEGA-PSIR (in the sequel Ω - Ψ^R), designed and implemented by a team of Warsaw University of Technology.

In 2013 the system started working at WUT. In addition, it has been deployed at other eleven academic institutions in Poland². In the paper we briefly present the main system architecture and its features. We also focus on the experience of WUT, discussing the organizational matters that have to be solved for heterogeneous sources of information, such as distributed data entry for the bibliographic materials, staff activities and achievements, and theses on one side, and information about the research projects on the other side. We also present how the organizational problems are solved at other Polish universities with running and developing the system.

Motivation

One of the first steps of the WUT team was to analyze what kind of the system, in terms of functionality, and then software solutions, should be taken into account. Observing contemporary information systems dedicated for institutional research knowledge bases, one can see two main different solutions: (1) on one side, an approach represented by systems like Fedora-Commons or D-space (see e.g. Berman (2008)), which focus mainly on the repository functions, such as storage and indexing of research-related documents, including also aspects of long term durability; (2) on the other side, an approach represented by systems, which are very well suited to the research-centered model of Current Research Information Systems (CRIS).

Institutional repositories are actually a dominating way for building institutional research knowledge bases. The systems within this approach provide rather simple end-user functionality, mainly limited to browsing and

¹ <http://repo.bg.pw.edu.pl/index.php/en/>

² Additional four universities start deploying the system, planning the regular exploitation for beginning of 2019 till mid of 2019.

querying the repositories. They are bibliography oriented, usually document-centric ones, and do not provide end users with any analytical functionalities, or with sophisticated presentation capabilities. Additionally, the data acquisition procedures are rather straightforward, based on human work, or harvesting data from well-defined resources. According to Russell and Day (2010) the main motivations for building institutional repositories are *inter alia*:

1. to provide a showcase for scholarly output from an institution (e.g. facilitating increased visibility; generating indicators of academic quality);
2. to improve dissemination of research output;
3. to ensure the long-term preservation of resources; and
4. to break down access barriers to content (i.e. reforming the scholarly communication system).

However, many authors indicate that achieving these goals with the current repository systems is very problematic. Aschenbrenner et al. (2008) claim that dissemination fails, mainly because the current journal system is seen as the most reliable way towards the academic career. Similar problems with the acceptance level of institutional repositories among researchers have been reported by Davis and Connolly (2007). Salo (2008) has noticed that an IR is *"like a roach motel – data gets in but never gets out."*

A lot of attention has been put recently to the institutional CRIS systems. Nabavi et al. (2016) define CRIS as Research Management Information System, which main goal is to provide a tool for supporting research decision makers (institutional or national, respectively) to carry out the following tasks:

1. Defining research priorities;
2. supporting and automating the process of assigning budgets for research groups, teams, faculties (at the institutional level);
3. evaluating and ranking research institutes based on defined criteria.

This view of institutional CRIS requires that the research be assessed, so, as noted by Bittner and Müller (2011), CRIS systems are very formal – "bureaucratic".

Yet another group of scientific information systems can be observed. Since 2008 one can development of the systems like ArnetMiner, Microsoft Academia, Vivo, or a bit later ResearchGate or Academia.edu. Among those systems, only Vivo can be seen as an institutional one, nevertheless all of them have something common – namely they are researcher-centric, rather than research-centric. In this context one should accept the view of Nabavi et al. (2016), which distinguishes Research Profiling Systems (RPS) from CRIS.

Having in mind all these tendencies, we have decided to integrate within $\Omega\text{-}\Psi^R$ as many functionalities of all the kinds of the above considered system types, as possible. Fig. 1 shows how the system types evolved into the nowadays scientific information systems. Our motivation has resulted from the first goal, which was to provide a functionality of an institutional system that would cover needs of as wide group of users as possible.

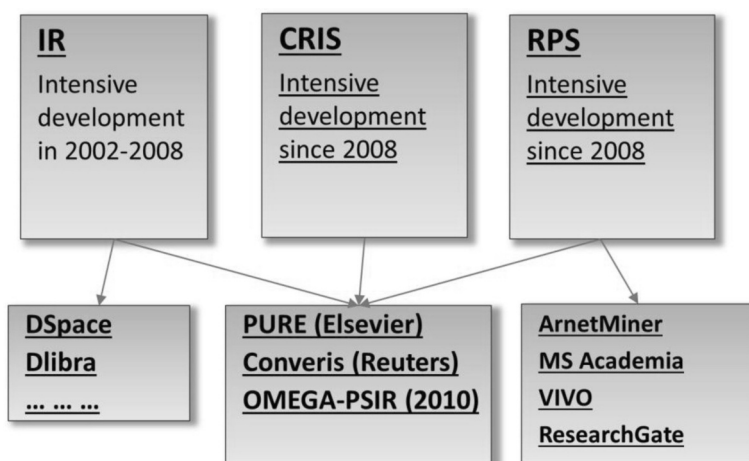


Figure 1 The ways of integrating various types of scientific information systems

From the point of view of data structures, the system is thought to store all the data representing various outcomes of research performed by the researchers. It is compatible with CERIF. The main objects covered

currently by the system are:

1. information about researchers and their affiliations
2. publications (books, articles, etc.),
3. patents,
4. diplomas (B.Sc., M.Sc., Ph.D.),
5. research projects, along with the project documents,
6. multimedia presentations, as well as research data, benchmarks, etc.,
7. other scientific documents (reports, reviews, etc.),
8. other external and/or internal activities),

Functionality of Ω - Ψ^R

The planned functionality was focused to cover needs of all the actors within the university, but also to take into account the needs of external users. For internal users, we took into account the following groups:

1. researchers;
2. students (graduates, undergraduates);
3. university administration;
4. scientific bodies (faculty councils, senate, promotion commissions, etc.); and
5. university managers responsible for research strategies.

Obviously, the needs of these groups vary essentially, and in many cases are even contradictory. For a researcher, one of the important functionalities is to provide access to the published materials (as IR does), but also

to support preparing bibliometric analysis of his/her research activities for promotion (as CRIS does), or for a grant;

to promote his/her research activities to the peers (as RPS does)

Students mainly search for *open access* manuals, whereas the administration is concerned about the completeness of the database and the production of reliable evaluation reports of the researchers, and/or university units, like faculties or institutes.

Yet another functionality is needed for the top management of the university, especially:

1. financial reports concerning research projects;
2. looking for most successful research domains at the universities, best teams, rising stars, etc., but also the weaker research domains, so that a proper development strategy can be worked out for the university (CRIS);
3. visualization of trends in research at the university (CRIS, RPS).

For external users, the role of the system is also multifold, though slightly different. We have taken into account not only academia users, but also:

1. entrepreneurs
2. regional and governmental administration,
3. judiciary
4. regional

To this end, the system should integrate various functions, but special importance should be given to the functions promoting researchers, university units and informal research teams. In this context, the idea researcher-centric approach, especially with the functionality of profiling researchers and university units is of very high importance. In the figure below the profiling functionality is illustrated.

As one can see, various aspects of the researcher activities are presented on the researcher profile. In addition, in order to strengthen cooperation of the university with external world we have implemented the following functions:

1. discovering experts in a given domain, based on their research achievements registered in the knowledge base;
2. finding networks of cooperating researchers.

Technical details concerning the algorithms implementing the above functionalities are presented in more detail by Rybinski et al (2017).

There are also other functionalities that we have taken into account, which aim at reducing costs of maintaining the system exploitation. These are *inter alia*:

1. Interoperability with other systems (Crossref, Scopus, Google Scholar, JCR, OA directories, etc.)
2. Acceptance of various import formats (OAIPMH, BibTeX, XML based formats, etc.)
3. Simple data entry procedures, including autoarchivin
4. Simple data control procedures

One of the main features of the software is its flexibility, so that so that demands for new object types and new relationships can be fairly easy to implement. The flexibility is reflected in providing administrative tools that enable defining new data structures, and then for the new objects make it easy to define new data entry worksheets and new search screens. As a result of a thorough evaluation of the needs of various levels of the academic community, the fundamental features of the implemented software are as follows:

1. the system provides an easy way for:
 - a. storing any typed metadata along with digital content
 - b. defining custom types of stored metadata
 - c. defining relationships between records of different types
2. with the defined data structures (rich in relationships between various objects), there are simple yet powerful means for the maintenance and control of the semantic data network stored in the knowledge database,
3. the system is able to preserve 'historical values' of linked objects in the course of changes,
4. the system provides multiversioning of the data,
5. an efficient full-text search capabilities in both metadata and digital objects is available,
6. the system provides means for automatic generation of highly ergonomic and customizable GUI,
7. the system provides means for an easy integration with external systems, such as SherpaRomeo, Scopus, Web of Science, Scholar and exposure for external search engines,
8. the system provides means for extensive access control mechanisms,

With all the features above, the last but not least, is the multilinguality of the system.

Implementation of OMEGA-PSIR in Poland

The Ω - Ψ^R system has been built iteratively over the period of three years. It was installed in production environment in its early development stages – at the beginning only at Institute of Computer Science, with functionalities limited to the basic repository functions. In the course of the development of new functionalities the system was providing more functions, and was delivered to wider range of users, first, in 2011 at Faculty of Electronics and Information Technology, and then in 2013 to the whole Warsaw University of Technology, still being subject of further development.

Such approach caused that the system was confronted with its users form the very beginning, and the developers were confronted with real user needs, so that when the system was finally ready to be shared with other universities in the form of a complete Ω - Ψ^R package, it was already mature, well-tested and well-documented.

Already in 2013 the functionality of Ω - Ψ^R went beyond the typical functionality of institutional repository towards the functionality of CRIS, additionally the functionalities of RPS have been also implemented. Due to applied intelligent tools (acquisition tools, reporting functionalities), the maintenance efforts of Knowledge Base are essentially reduced compared to the typical solutions. The process of moving the system to the University level was already simpler, as the team had experience with organizational and training issues at the faculty level.

As a matter of fact, at the beginning there was a skepticism coming from the researchers group. In the course of the knowledge base development the researchers could immediately observe their profiles, so that they gradually turned to be more and more keen to contribute to the process of the database maintenance. The main trigger for the staff involvement was the fact that they have noticed correlation between the way the system was presenting their own profiles, and their achievements in the knowledge base. Getting familiar with the integration of repository functions, visualization of research, and reporting the university staff became the first beneficiary of the research knowledge base.

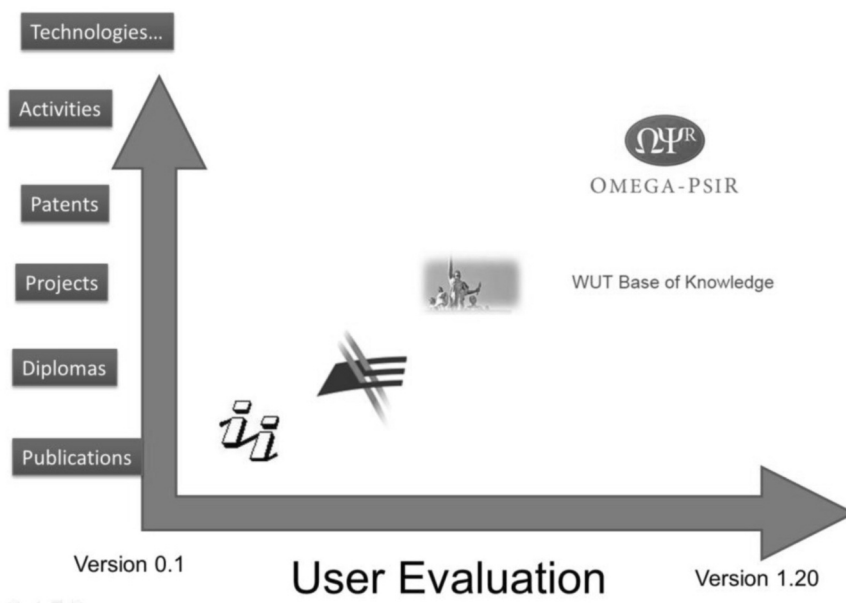


Figure 4 User-driven Development of OMEGA-PSIR

What is important to notice, the cost of the system running is not much higher than the cost of a typical IR. This is mainly because of simplifying the data entry procedures. The data entry processes are performed at the locations where the new information emerges. So, the new publications are registered either by scientists or by faculty editors, the projects are maintained by the relevant university offices.

On the other hand, the effects resulting from using the system just for reporting purposes are quite noticeable. The time gained by using the system reporting features can be efficiently used by the system users for more valuable activities.

In addition, now the system works at 11 other universities in Poland. The main functionalities that have been implemented refer to running the bibliographic data and theses. The database maintenance processes are usually organized in a different way than at WUT, namely the tasks for data entry are performed centrally by the main libraries. In most of the cases, the system is used for reporting to the national evaluation system. There are however steps undertaken towards implementing other "system modules", especially the research projects module.

System usability

Having experienced three years of running the system at WUT, i.e. at the end of 2016, we have decided to verify the system usability thoroughly. To this end, we have performed two types of research:

1. we have prepared a survey research among the active and potential users;
2. in addition, taking advantage of using Google Analytics since May 2013 we have analyzed the Google Analytics data.

Survey research

The background for the research has been provided by Kedzierska (2013). One of the main defined group of the system recipients are research workers from scientific units, research and development units and universities. That is why the questionnaires were distributed among Polish scientific institutions (156 scientific and research and development units, 73 scientific institutes of the Polish Academy of Sciences and 411 public and private higher education institutions). 197 completed questionnaires were sent back (60% from public higher education institutions, 30% from scientific institutes, 4% from private higher education institutions, 4% from research and development units, 1% - others).

Other groups of the academic community were also asked about their opinion. 500 students and 180 doctoral students of the Warsaw University of Technology received the questionnaires and 45 students and 19 PhD students sent back their completed questionnaires. Entrepreneurs, identified at the beginning of the study as an important separate group of stakeholders, were very difficult to analyse. The list of companies cooperating

with the WUT Careers Office was used to create the mailing list and it was possible to collect 20 opinions of entrepreneurs.

Public administration employees were also an important surveyed group. The questionnaires were sent out to representatives of various Polish ministries and municipalities, 314 addressees in total. 22 public administration employees filled in their questionnaires.

In total, 424 opinions were collected. When asked which group will find the data provided by the OMEGA-PSIR system most useful, the respondents indicated that it is mainly employees of Polish and foreign scientific units and universities.

A high level of relevance and usefulness of such data to Polish and foreign entrepreneurs was also indicated. It is worth noting that each group of surveyed beneficiaries declared that the resources would be very useful for them.

All target groups selected in the analysis differ in the way they would use the provided data. Particular attention should be paid to the following groups: universities and scientific units, students and PhD students, media representatives and university management staff.

The results of the survey can be summarized as follows:

1. The largest group of the system recipients are academic staff of Polish and foreign universities and scientific institutions. Open access to full-text scientific publications brings many benefits to both scientific entities and to the whole scientific community. It is important to mention the possibility of multi-criteria search and automatic generation of information about the achievements, areas of interest of individual scientists, units, and teams. All these criteria have a positive effect on intensifying the exchange of scientific ideas and establishing new cooperation between institutes.
2. Another group of recipients are entrepreneurs who, with help of the OMEGA-PSIR system, can cope with the existing barriers impeding the establishment and effective cooperation of science and business. The biggest problem in this area is insufficient knowledge of the entrepreneurial sector related to the areas of expertise of Polish universities, which is also related to the lack of experience and good practices in effective presentation of the university academic achievements and research.
3. PhD students, undergraduate and graduate students, but also and candidates to study, are also an important group. The system is particularly helpful in recognizing the character of the unit and the quality of its services.
4. The system is also helpful for media representatives who can have access to information about experts in various fields, which is necessary to get professional comments on current events.
5. And last but not least important group is the university management staff. Having the OMEGA-PSIR system with the CRIS elements, they have an important tool for assessing the work of individual scientists and teams, and also a tool for developing a scientific policy of the unit.

In addition to identifying the potential stakeholders, the survey also enabled identification of the key needs of potential recipient groups and identification of data types and analytical services in the systems with open access to scientific information. The most important issue for the respondents using these types of systems is free access to the data. They also pointed out that in this context, it is important for them to have open access to as many scientific publications as possible as well as to raw research data. In addition, we have asked for the opinions concerning especially those non-typical ones:

1. graphical visualization of cooperating teams;
2. graphical visualization of profiles of individual researchers or university units;
3. search for experts.

It turns out that researchers evaluate equally high the functionalities of visualizing cooperation between units, researchers, as well as, visualization of the research areas of units and individuals, which is provided in the form of terms clouds.

An overall view of the users preferences is summarized in Table 1:

Google Analytics statistics

The GA data for the WUT Knowledge Base are available since 2013. They show a growing interest for Data in the WUT Knowledge Base. In 2016, some 1,000 sessions per month (10 % of total traffic) were started by users out of Poland, now (2018) the number of sessions started out of Poland reaches 3,000. The geographic distribution shows that the traffic from USA, and EU countries is significant (now more than 15 %).

Services	Interactive graph of cooperation between faculties	Interactive graph of cooperation between researchers	Visualization of research areas of the university and its units	Visualization of research areas of researchers	Multicriterial search for experts and teams in a given domain
Researchers	7,32	7,43	7,41	7,57	8,57
Business	5,15	4,8	6,05	6,2	8,15
Public admin.	4,00	3,73	3,95	4,00	5,05
PhD students	6,84	6,89	6,84	7,37	8,21
Students	6,93	6,67	6,82	6,82	8,09

Table 1 Preferences of Ω - Ψ^R services by various user groups

With GA we can assess the statistics for the users' behavior. In particular, we have analyzed the types of the starting screens, and the distribution of the functions used by the users. It turns out that quite a meaningful number of users start their session with the screens that are not available within typical repositories, such as viewing researcher profile (17 percent), the screen for searching experts by domain (10 percent), or the screens with profiles of the university units, such as faculties, institutes, etc. (7 percent).

As an example, Figure 5 below shows behavior of the users from USA in 2013-2016. One can see that in addition to the standard repository functions, users often refer to the function of looking for experts.

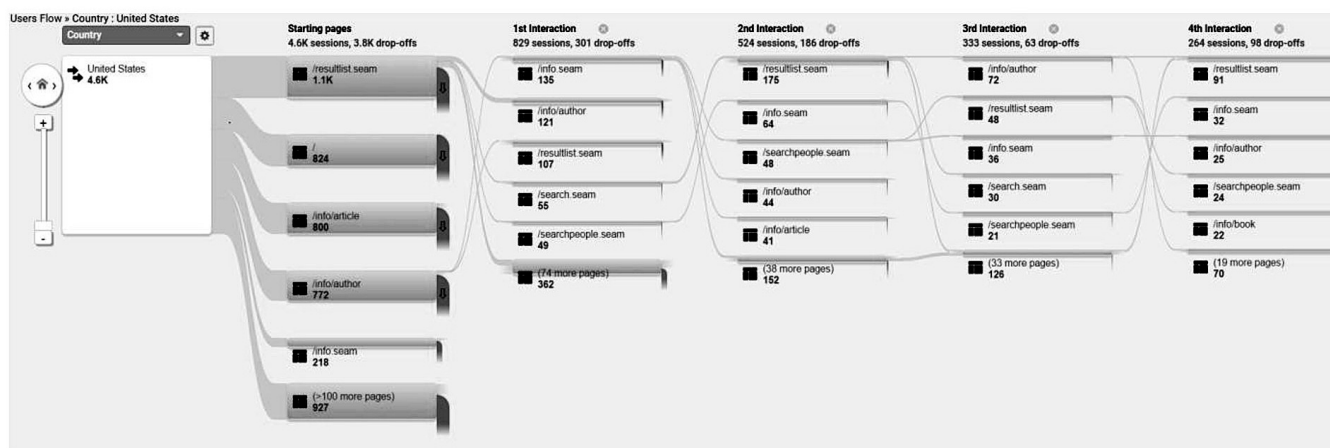


Figure 5 Site-flow for OMEGA-PSIR users from USA

Unfortunately, with the GA data we cannot decide, which groups of users generate requests for these functionalities, but the analysis confirms the results of the survey research concerning high interest in those functions that are not typical for institutional repositories, and result from the researcher-based approach.

Conclusions

One of the lessons learned was that with building an information system, for the first glance looking as a fairly typical one, we have encountered many interesting real-life research problems in such areas like knowledge acquisition and discovery, text mining, or information retrieval.

The Ω - Ψ^R platform has been successfully implemented at WUT as the research knowledge base, and its acceptance level is now very high among various users groups at the University.

A positive evaluation of the system resulted in a high interest of deploying Ω - Ψ^R at other universities. In addition to WUT, the system has been deployed at 11 Polish universities.

What is important, we continue developing the system. Recently we have developed interoperability with Scopus. Also the reporting functionality is subject of further extensions – we aim at providing such tools that administration staff is able to define statistical tables without any intervention from the IT people.

Also, some new functions for the University management are planned. In particular, we plan adding the functionality for looking for "rising stars" (some research in this direction has been already performed).

Acknowledgement

The authors would like to thank Dr Keith G. Jeffery for his constructive comments.

References

- Aschenbrenner, A., Blanke, T., Flanders, D., Hedges, M., & O'Steen, B. (2008). „The Future of Repositories?“ D-Lib Magazine, 14(11/12), 1082-9873.
- Berman, F. (2008). „Got data?: A Guide to Data Preservation in the Information Age“, Comm. of the ACM, Vol. 51 No. 12, pp. 50-56
- Bittner, S., & Müller A. "Social networking tools and research information systems: Do they compete?" (2011): 1-4.: WebSci '11, June 14–17, 2011, Koblenz, Germany
- Davis, P.M., & Connolly, M.J.L. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's Installation of DSpace. D-lib Magazine, Vol. 13, No. 3/4
- Kędzierska, E. (2015). Wzmocnione publikacje: nowy model wsparcia komunikacji naukowej. Zagadnienia Informatyki Naukowej, (1).
- Nabavi Majid, Jeffery Keith, Jamali Hamid R., (2016), "Added value in the context of research information systems", Program, Vol. 50 Iss 3
- Rybinski, H., Skonieczny, L., Koperwas, J., Struk, W., Stepniak, J., & Kubrak, W. (2017). Integrating IR with CRIS—a novel researcher-centric approach. Program, 51(3), 298-321.
- Russell, R., & Day, M. (2010). Institutional repository interaction with research users: a review of current practice. New review of academic librarianship, 16(S1), 116-131.
- Salo, D. (2008). Innkeeper at the Roach Motel. Library Trends, Vol. 57, No. 2, pp. 98-123

H Rybinski, W Kubrak, J Koperwas, Ł Skonieczny, W Struk (2018), "Omega-Psir – Institutional CrIS at Polish Universities". *ITlib. Informačné technológie a knižnice* Special Issue 2018: pp 36 – 44
<https://dx.doi.org/10.25610/itlib-2018-0005>

Henryk Rybinski, Weronika Kubrak, Łukasz Skonieczny, Jakub Koperwas, Waław Struk

(Warsaw University of Technology · Faculty of Electronics and Information Technology, Poland) ■