

Marzena Błaszczczyńska, Poznań Supercomputing and Networking Center,
marzena@man.poznan.pl

Karolina Bohdanowicz, Poznań Supercomputing and Networking Center,
karolina@man.poznan.pl

Natalia Jeszke, Poznań Supercomputing and Networking Center, njeszke@man.poznan.pl

Tomasz Parkoła, Poznań Supercomputing and Networking Center,
tparkola@man.poznan.pl

Simplicity, quality and visibility as key drivers to build SINUS system

Abstract

Universities and research institutes in Poland use digital libraries, repositories, bibliographic databases or Current Research Information Systems to promote or give access to their scientific achievements.

In 2015, the Poznań Supercomputing and Networking Center started cooperation with the Library of the Poznań University of Technology to optimise the process of managing the scientific achievements of the Technical University. The envisioned solution was to replace four legacy systems used at the Technical University, provide comprehensive information for end users, automate the work of collecting and sharing resources, and support flexible data export.

In course of the work we executed in-depth analysis of existing software solutions and as a result decided to develop a new system adjusted to the needs of the Technical University. Our approach in the design & development work was to deliver system which is simple to use, has built-in features to improve/ensure quality of provided data and has good visibility in search engines.

After several years of work we have strong believe that detailed analysis of the needs, flexible solution as well as simplicity, quality and visibility are the most important factors of success when building comprehensive solution for promotion and access to academic research information/outcomes.

Keywords

Open repository, digital library, system for scientific information, Current Research Information System

Background

SINUS is an open repository, a digital library and an academic staff evaluation tool in one. We argue that simplicity, quality and visibility are the most important characteristics of such a solution and present our experiences as well as lessons learned in this regard. New generation digital repositories or research information systems should be developed with these factors in mind.

Content

Many universities and institutes in Poland use digital libraries or repositories to share and promote information of their scientific achievements. This openness creates opportunities not only to re-use scientific information at the domestic branch of Polish science, but also worldwide. Additionally, many institutions share their scientific achievements in bibliographic databases or Current Research Information System (CRIS), where only metadata and sometimes links to content are available.

Bibliographic databases are used in most cases to gather, prepare and present statistical data for their own needs or for the needs of higher education institutions, such as the Ministry of Science and Higher Education. Mentioned systems operate independently, which leads to duplication of information, and therefore to difficulties for end users to access/discover information. In addition, updating information in several systems significantly increases the amount of work spent by editors. Working with multiple tools is usually non-intuitive, and sometimes the quality of the data is questionable. On top of that, mentioned systems usually do not allow to introduce untypical resources such as architectural designs, technical documentation or software.

In 2015, the Poznań Supercomputing and Networking Center started the cooperation with the Library of the Poznań University of Technology in order to develop and implement a tool to optimise the process of managing the scientific achievements of the Technical University. This optimization is on both levels: the metadata level and digital content level. According to the assumptions, the new system was to: replace four other systems which were used at the University at the same time for a similar purpose, provide comprehensive information for end users, automate the part of the work of people responsible for collecting and sharing resources, and support connection with external systems (in particular in connection with the national system of the Polish Scientific Bibliography that obliges Polish universities to report scientific achievements).

The first stage of work was the analysis of existing domestic and foreign IT systems, which had features and functionalities helpful in building a tool for the University of Technology. In case of repository systems DSpace and E-Prints were verified. In the case of these systems, the data models excluded the possibility of using them as the base for the construction of a CRIS-like system, i.e. the data model was too poor. In case of CRIS-like systems the VIVO solution was verified. This semantic-based system met requirements for the institutional repository and the CRIS system itself. In addition, it had a graphical interface available through a web browser. However, during the final internal verification focused on usability, performance and functionality, it turned out that the system did not meet the University's expectations. After several months of testing, it turned out that the performance of the system significantly drops with the introduction of thousands of resources with an extensive description. The lack of creators to introduce complex bibliographic descriptions, and the lack of duplicate detection mechanisms, also turned out to be a cumbersome. The elimination of the observed problems in the context of the VIVO system was considered as excessively time-consuming. Therefore, at the end of 2015 it was decided to design an individual solution that would meet all the University's requirements (repository and CRIS features) and ensure ergonomics of data entry.

The second stage was based on the previous experience. We designed from scratch a software tool named SINUS with the strategy focused on three main characteristics: **simplicity**, **quality** and **visibility**. Our top priorities related to **simplicity** are clarity and

transparency of the web portal interface, a wide range of search options and a convenient way to browse the available resources. Users of modern websites expect a well-designed architecture of information and friendly, engaging, but no-intrusive interactions. Therefore, in search of inspiration and good practices in the area of interface usability, we have reached out to many websites outside the science field. We are aware that simplicity is a feature that supports navigation, emphasizes the hierarchy of individual content, and facilitates browsing. We also know that the measure of success of websites is the **visibility** of their content in search engines. Therefore in SINUS we use best practices of Search Engine Optimization (SEO) to improve our site's appearance on Google Search and others. We organized site hierarchy into distinct sections (people, organizations, publications, journals, conferences) and each item has its own page. We created unique titles for each page to display brief, but descriptive information (e.g. site prefix with person/organization/conference name or publication/journal title). Where applicable and reasonable we use URLs with words that are relevant to our site's content and structure, not just identifiers which are hard to read/understand by humans. To make our site more visible we also created a sitemap file and submitted it for indexing. Our sitemap file contains URLs to main page, profiles of all publications, all journals where publications were published, all conferences, where publications were presented, all authors of publications and all organization units of the Poznań University of Technology. This file is updated on a daily basis. In addition, we made our site mobile-friendly. In production environment we used Google Search Console to monitor site's presence in search results, to optimize site for search engines and to fix indexing problems.

The last stage was to improve and facilitate the data entry process (**quality**). We divided the forms into steps and introduced a number of improvements in the ergonomics of their completion (e.g. lists with dictionaries, error detection). We also applied a duplicate detection mechanism for publications, profiles and conferences. SINUS has two types of duplicate detection, first one is based on the mechanism of assigning unique numbers, and second one is based on external identifiers such as: ORCID, ResearcherID, Google Scholar, PBN ID (which is the database of Polish Scientific Bibliography) and DOI for publications. In SINUS authors can submit their publications on their own (self-archiving). Each time, these data are verified and updated by staff working at the Library of The University. The described functions mean more effective work of the librarians, faster process of collecting and sharing publications, but most of all - higher data quality.

SINUS is also equipped with the evaluation feature. It can be used for evaluating achievements of researchers and, in a broader perspective, entire institutions conducting scientific and research activities. The assessment of the quality of research results and the characteristics of institutions is based on guidelines from the Polish Ministry of Science and Higher Education. The purpose of the guidelines is to select the best researcher and teaching staff, as well as the best institutes and faculties and even broader - the leading disciplines of research. In the long run, it means better financing of units and support for specific areas of research. In the new system it is easier to generate data results, which is obligatory to report on every polish university.

There are more possibilities of generating statistics in the system. The university authorities have a quick insight into the research work of the entire scientific and didactic unit by generating data to an XLS file. Excel forms allow for easy processing of statistical data, which can be done by qualified employee. In this way, SINUS makes it possible to

delegate the tasks of evaluating the achievements of institutes, to administration staff and librarians.

The evaluation of achievements is based on bibliometric parameters as well as criteria specified by the Act of the Ministry of Science and Higher Education. SINUS processes the following parameters: Impact Factor, information on indexing in the Web of Science, Altmetrics with citation based metrics and scientific publication influence. Aside the mentioned external indicators, the Ministry introduced its own criteria for evaluating achievements, in the form of scoring list of journals, rules concerning the scoring of monographs, scientific conferences or patents. What is also worth mentioning, SINUS allows for the implementation of all types of indicators and criteria for the evaluation of researchers' achievements.

Further development of the SINUS system will be focused on the needs of end users such as university authorities and staff. Because the legal regulation on parameterization is not limited only to scientific publication, but also includes such work as patents, inventions forms, national and international conferences' papers, it is planned to expand the existing functionalities of the system. Adding new features, especially in the context of preservation of scientific data, will broaden the spectrum of its possibilities and result in the complex solution providing features of digital repository, CRIS as well as DAM system.

Conclusion

Our experiences from past years give us a clear message to keep in mind the following:

- thoroughly test existing solutions and validate them against your requirements before deciding to use one or the other software solution, do not base on common approaches as "it's open source, it must be good for me" or "it is very popular in this country, it must be good for me".
- in the current digital ecosystem of world wide web it is a must to provide simple, good quality and visible solutions, so that the users are engaged, satisfied and guided to the resources they require; wherever possible use User Centered Design and test your solutions with real users.
- technical foundations of the system should be build in such a way that the changes, which are inevitable, can be introduced as smoothly as possible, e.g. by using flexible technical framework of appropriate programming language.

References

Smith, M.; Barton, M.; Bass, M.; Branschofsky, M.; McClellan, G.; Stuve, D.; Tansley, R.; Walker, J. "DSpace: An Open Source Dynamic Digital Repository", *D-Lib Magazine*, Volume 9 Number 1, January 2003.

Tansley, R.; Bass, M.; Stuve, D.; Branschofsky, M.; Chudnov, D.; McClellan, G.; Smith, M. "The DSpace institutional digital repository system: current functionality", *2003 Joint Conference on Digital Libraries, 2003. Proceedings.*, 20 June 2003. DOI: 10.1109/JCDL.2003.1204846

Beazley, M. "Eprints Institutional Repository Software: A Review", *Partnership: the Canadian Journal of Library and Information Practice and Research*, Volume 5 Number 2, 2010. DOI: 10.21083/partnership.v5i2.1234

André Rosa, C.; Craveiro, O.; Domingues, P. "Open Source Software for Digital Preservation Repositories: a Survey", *International Journal of Computer Science & Engineering Survey (IJCSES)*, Volume 8 Number 3, June 2017. DOI: 10.5121/ijcses.2017.8302

Błaszczńska, M.; Kozak, M.; Mazurek, C.; Szymczak, M.; Werla, M. "System SINUS – otwarte narzędzie do budowy bibliograficznych baz danych", *Bibliograficzne bazy danych : perspektywy i problemy rozwoju. Kraków, 26-27 czerwca 2017 r.*, Number 25, 2017.