PBN - Polish Scientific Bibliography - its roots, usage, and mapping with Dublin Core and CERIF – case study of the Jagiellonian University

Introduction

In 1990-2005 the number of Polish Higher Education students increased massively from 403 thousand to 1 million 954 thousand. In parallel with this quantitative increase in teaching, the academic institutions undergo significant qualitative improvement in scientific activities. The Polish government supported this transformation by the introduction in 1991 of KBN – State Committee for Scientific Research to administer a performance-based research funding system. After a decade of expert-based evaluation run by KBN, a new body KEJN - Committee for Evaluation of Scientific Entities replaced KBN. It oversaw two assessments for 2009-2012 and 2013-2016, awarding categories for organisational units of scientific institutions, which impact their standing and funding. A new regulation from 2018, the Constitution for Science, changed the scope of categorisations to scientific disciplines within institutions and established rules for the assessment for 2017-2020. This law also directly connected the rights of institutions to award scientific titles with obtained categories. Last but not least, the Constitution for Science replaced KEJN with another body KEN - Commission of Evaluation of Science. COVID-response regulations extended the third evaluation to include 2021.

Bibliographical data in the evaluation of scientific activities

Under the current law, the evaluation of publications by authors affiliated with a particular entity is a significant part of its categorisation. The assessment uses the data from POL-on - an integrated information system on Polish Higher Education. One of its parts is PBN - the Polish Scientific Bibliography, where authors or institutions submit bibliographical data. SEDN – the system for evaluating scientific activities uses the PBN data to select the most valuable publications and assess the quality of research at institutions where contributors affiliate to.

The law requires bibliographical data for the evaluation to be submitted via PBN. The universities participating in the evaluation sent over 700 000 publications issued in 2017-2021 to the PBN. The largest collaborator of the PBN is the Jagiellonian University in Kraków, which submitted over 36,000 publications for evaluation.

PBN Bibliographical data

PBN differentiates three types of publications:

- Article
- Monography

---


3. Dz.U. 2010 Nr 96 poz. 615
7. Dz.U. 2015 poz. 944 + modyfikacje …
Initially, the PBN structure was quite simple, with only a few data elements outside a typical bibliographical system. Nevertheless, it sufficed for the first two, mainly quantitative, evaluations undertaken for 2009-2012 and 2013-2016. The additional data primarily dealt with authors' affiliations to scientific institutions. Universities used the online interface or the XML import to submit their data.

The current evaluation, focusing on the quality of publications introduces further data set extensions. For example, the law requires information on points assigned to publications, including the contribution of individual authors, more detailed information about conferences and more specific information on scope of authors contribution. PBN system has been modified accordingly also an API for data submission has been introduced.

**Preparation of PBN data at the Jagiellonian University**

Jagiellonian University has been using DSpace as a repository of its publications since 2014 using the Dublin Core format to save publication metadata. For the first evaluations, the Jagiellonian Library team added additional fields required in a separate tab and later used a dedicated ETL tool to export DSpace data into PBN compatible XML. The metadata required by the PBN has been added to the basic Dublin Core schema in DSpace. However, in the evaluation carried out for 2017-2021, the set of metadata was expanded once again. The team of the Jagiellonian Library decided that further expansion of the Dublin Core scheme in the Jagiellonian University Repository operating on the DSpace system would not be the best solution.

The University has deployed Sciencecloud\(^9\) a research information system in preparation for the current evaluation. The system mediates in the export of metadata between the DSpace and PBN system, and at the same time allows you to add metadata necessary for the ongoing evaluation to individual bibliographic descriptions. Metadata sent to Sciencecloud are confirmed by authorized employees of faculties and departments, and after final checking by librarians, exported to PBN using its REST API. This allows for unlimited changes to the metadata format according to ministerial guidelines without the need to constantly expand the Dublin Core schema at the DSpace level.

**Mapping of PBN data into CERIF structure**

Storing complete information on a wide range of research activities of a university requires an appropriate methodology for their description that is much more complicated than the DC model that was designed as a simple metalanguage appropriate for general purposes. The CERIF (Common European Research Information Format)\(^10\) offers more complex idea model that is structured into entity types and features\(^11\) based on Core Entities such as Person, Organisation Units and Projects and corresponding Result Entities containing Publications and Patents or some 2nd Level Entities i.e., Equipment, Prizes, Citations or Metrics. Accompanied with flexible Link Entities considered a major strength of the CERIF model being able to create relationships between mentioned above, along with Semantic and Multilingual Features deliver fully extendible metalanguage suitable for building strong base for CRIS (Current Research Information System) that the Jagiellonian University is willing to develop.

---


Finally, the CERIF model being way more complex, can still be mapped into some other popular data structures such as mentioned POL-on and PBN models that shares the idea of connecting modular entities into components that work together.

The authors will present a matching table showing development of PBN structure and its mapping with Dublin Core and CERIF dataset adding to the discussion\textsuperscript{12,13} on national and international standards of research information.

\textsuperscript{12} https://eurocris.org/blog/insights-eurocris-board-cerif-and-kdsf

\textsuperscript{13} https://www.kerndatensatz-forschung.de/version1/technisches_datenmodell/document/Einfuehrung%20Datenmodell%20KDSF%20und%20CERIF.pdf