Aleksander Nowiński
Wojtek Sylwestrzak
Wojciech Fenrich
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
Uniwersytetu Warszawskiego

Polska Bibliografia Naukowa

Aleksander Nowiński - kierownik działu rozwoju oprogramowania w Centrum Otwartej Nauki, w ICM Uniwersytetu Warszawskiego. Od lat zajmuje się budowaniem i rozwojem systemów informatycznych dla bibliograficznych baz danych oraz bibliotek wirtualnych. Był m.in. dyrektorem technicznym projektu Europejskiej Matematycznej Biblioteki Cyfrowej, a także koordynuje rozwój platformy Yadda.

Wojtek Sylwestrzak - pracuje w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego, gdzie kieruje Centrum Otwartej Nauki CeON. Zaangażowany jest w promocję otwartego modelu nauki oraz uczestniczy w szeregu inicjatyw związanych z rozwojem nowoczesnych metod komunikacji naukowej. Jest aktywnym uczestnikiem takich projektów jak OpenAIRE (Europejska infrastruktura otwartego dostępu do badań OpenAIRE), EuDML (cyfrowa biblioteka matematyczna), SYNAT (krajowa e-infrastruktura nauki i techniki), Wirtualna Biblioteka Nauki czy Polska Bibliografia Naukowa (PBN).

Wojciech Fenrich - absolwent socjologii i filozofii, ukończył studia doktoranckie w Instytucie Socjologii Uniwersytetu Warszawskiego. Od 2011 r. pracuje w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego. Analityk systemu POL-on.

***Streszczenie**: Polska Bibliografia Naukowa (PBN) jest elementem powstającego od roku 2010 systemu informacji o szkolnictwie wyższym POL-on, będącego własnością Ministerstwa Nauki i Szkolnictwa Wyższego. Zasadniczym celem PBN jest agregacja informacji o dorobku publikacyjnym polskich naukowców i jednostek naukowych zarówno dla celów ewaluacji, jak i analiz. Baza danych PBN opierać się będzie na istniejących bibliografiach instytucjonalnych oraz danych wprowadzanych bezpośrednio przez autorów. System udostępniony został we wczesnej wersji beta, która jest na bieżąco rozwijana i dostosowywana do potrzeb zgłaszanych przez użytkowników. W niedalekiej przyszłości funkcje PBN zostaną rozszerzone o repozytorium prac dyplomowych bronionych na polskich uczelniach.*
***Słowa kluczowe**: bibliografie publikacji, bazy danych, Polska Bibliografia Naukowa, ocena parametryczna, ocena dorobku naukowego*

***Abstract:** Polish Scholarly Bibliography (PBN) is a component of a country-wide system POL-on, a System for Information About Higher Education in Poland, developed for Ministry of Science and Higher*

Education in Poland. The PBN has been created to aggregate information about scientific publication of Polish reasearchers and scientific institutions. This data is necessary both for evaluation and for deep analysis of the scholarly communication in Poland. The PBN is a database, which combines data from existing institutional bibliographies with direct input of the publication metadata with web interface. The system is already available to the users in beta version, and is adopted to the user needs. It is expected that soon it will be extended as a central repostory for the thesis in Poland.
Keywords: *bibliographies of publications, databases, Polish Scholarly Bibliography, parametric evaluation, science achievements evaluation*

Prezentacja

O projekcie

Polska Bibliografia Naukowa (PBN) jest systemem realizowanym na zamówienie Ministerstwa Nauki i Szkolnictwa Wyższego. W jednej bazie danych gromadzi ona informacje o dorobku publikacyjnym polskich naukowców i jednostek naukowych. PBN jest elementem znacznie szerszego systemu informacji o szkolnictwie wyższym POL-on, tworzonego dla potrzeb zarządzania szkolnictwem wyższym i nauką w Polsce. Projekt jest finansowany z funduszy strukturalnych Unii Europejskiej w ramach programu operacyjnego „Kapitał ludzki". System POL-on realizowany jest przez konsorcjum czterech partnerów: Ministerstwo Nauki i Szkolnictwa Wyższego, Ośrodek Przetwarzania Informacji, Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego (ICM UW) oraz Index Copernicus International sp. z o.o. Moduł Polska Bibliografia Naukowa realizuje ICM UW, który jest również jego operatorem.

Cele projektu

Najistotniejszym elementem oceny dorobku, zarówno poszczególnych naukowców, jak i jednostek naukowych, są obecnie ich publikacje. Z punktu widzenia instytucji powołanych do zarządzania nauką w Polsce gromadzenie informacji o publikacjach pozwala skuteczniej realizować wyznaczone zadania i oceniać ich rezultaty. Obecnie raportowanie publikacji każdorazowo realizowane jest za pomocą dedykowanych, niezintegrowanych systemów informatycznych — informacje o publikacjach danego autora lub jednostki zbierane są w inny sposób i w innym formacie. Dotyczy to ewaluacji jednostek naukowych, ubiegania się o granty naukowe i późniejszego raportowania ich rezultatów oraz procedur związanych z nadawaniem stopni naukowych. Jednocześnie wiele instytucji gromadzi informacje o swym dorobku na zasadzie „zrywu", nierzadko realizowanego przez przypadkowe osoby, co znacząco obniża jakość danych. Istniejące w kraju instytucjonalne bazy danych pokrywają jedynie pewien wycinek dorobku naukowego polskich naukowców i jednostek, nie pozwalając na łatwe wykonanie całościowej kwerendy zarządczej. Z kolei istniejące bazy dziedzinowe fachowo opisują publikacje, zazwyczaj jednak nie przechowują pełnej informacji

o afiliacjach i nie są zintegrowane z katalogami osób i jednostek, kluczowymi w przypadku raportowania.

Celem PBN jest wypełnienie tej luki i stworzenie zintegrowanego systemu informacji o dorobku publikacyjnym polskich autorów i placówek badawczych, a co za tym idzie, umożliwienie ogólnych analiz i raportów zawierających porównania między różnymi instytucjami i różnymi obszarami wiedzy. Oznacza to ograniczenie konieczności raportowania dorobku publikacyjnego do lokalnego systemu i stopniowe zastępowanie wielorakich ankiet i raportów wymaganych przez różne instytucje. Pozwoli to na zarządzanie oparte na rzeczywistych danych i bieżące monitorowanie rezultatów wdrażanych programów i projektów.

Dodatkowym celem systemu jest dostarczenie narzędzi tym jednostkom, które są zainteresowane prowadzeniem własnej bazy dorobku, ale nie posiadają do tego odpowiednich systemów. PBN udostępnia istotne funkcje pozwalające na prowadzenie takiej bazy w skali jednostki, pozwalając na raportowanie i analizę informacji o publikacjach bez konieczności ponoszenia dodatkowych kosztów związanych z zakupem i utrzymaniem własnego serwisu.

Dane w systemie

Polska Bibliografia Naukowa gromadzi informacje o publikacjach naukowych polskich jednostek i autorów. W systemie zbierane są informacje o: artykułach naukowych, książkach (monografiach) i ich fragmentach. W przyszłości system zostanie rozszerzony o możliwość deponowania rozpraw doktorskich. Dane o osobach i jednostkach pochodzą z centralnej bazy systemu POL-on, co zapewnia ich aktualność i poprawność. Z mocy prawa każdy naukowiec i nauczyciel akademicki pracujący w polskiej instytucji naukowej rejestrowany jest w systemie POL-on, gdzie przechowywane są jego dane, m.in. numer PESEL. Podobnie wszystkie instytucje naukowe ubiegające się o dotację statutową lub prowadzące studia wyższe muszą zostać zarejestrowane w tym systemie. W efekcie PBN dysponuje stale aktualizowaną bazą osób i jednostek pozwalającą na sprawne realizowanie celów projektu.

Katalog czasopism w PBN jest kombinacją informacji z różnych źródeł i jest utrzymywany lokalnie na potrzeby systemu. Jednocześnie trwają prace, aby podstawą dla tej części bazy stał się katalog czasopism będący elementem katalogu centralnego NUKAT. W związku z rozszerzeniem systemu POL-on o informacje o grantach i projektach do końca 2013 r. w PBN dostępne będą również informacje o grantach z systemu obsługa strumieni finansowania (OSF), dzięki czemu możliwe będzie powiązanie ich z publikacjami stanowiącymi rezultat badań realizowanych ze środków przyznanych w ramach poszczególnych strumieni finansowania.

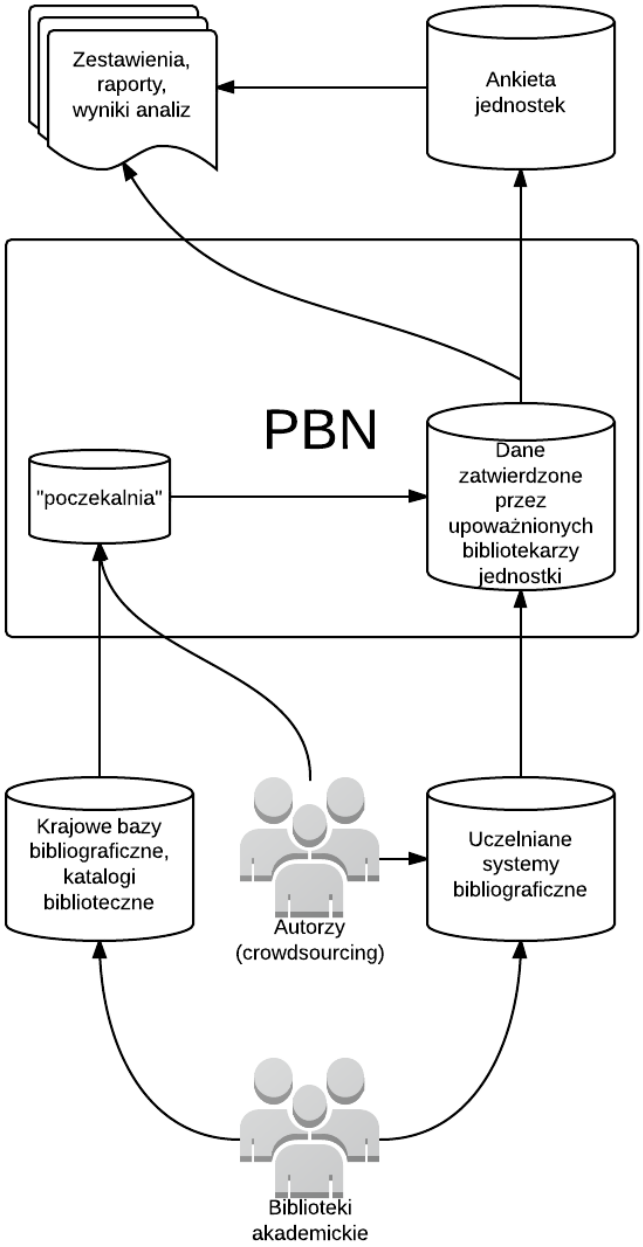
Ponieważ PBN pełni przede wszystkim rolę agregatora danych pochodzących z wielu źródeł, bardzo istotny staje się problem jakości danych w systemie. Jako że głównym celem gromadzenia danych w PBN jest identyfikacja osiągnięć osób i jednostek, kryteria dotyczące jakości tych danych są przede wszystkim definiowane tak, by umożliwić przeprowadzenie odpowiednich analiz. Jednocześnie problemem jest to, iż bibliografie instytucjonalne mają dane bardzo różnej jakości, zawierające często pola różnie zdefiniowane i zakładające różne dodatkowe ograniczenia. Dlatego zakłada się, że jakość i integralność danych w PBN będzie poprawiać się stopniowo.

Pojedyncza publikacja znajdująca się w PBN może stanowić element dorobku kilku osób i instytucji. Jest to różnica w stosunku do typowych bibliografii instytucjonalnych, w przypadku których to jednostka sama decyduje o regułach opisu. Dlatego też w PBN przyjęto zasadę opisu pracy bazującego na informacjach znajdujących się na samej publikacji. System pozwala na wprowadzenie w ten sposób wszystkich informacji, włącznie z oryginalnym zapisem afiliacji i nazwisk autorów, i osobne powiązanie ich z właściwymi jednostkami i osobami w bazie systemu POL-on. Ogólny model danych jest bardzo szeroki i pozwala na wprowadzenie bardzo wielu informacji, włącznie z wieloma wersjami językowymi abstraktu czy słów kluczowych. Jednocześnie przyjęto nietypowy sposób oznaczania typów (cech) publikacji, który wynika z konieczności dostosowania bazy do potrzeb ewaluacji i raportowania.

Możliwe jest też zdeponowanie pełnych tekstów publikacji w formacie PDF. PBN spełnia tym samym również funkcję repozytorium centralnego.

Zarządzanie danymi

Dla systemu pełniącego funkcję krajowego agregatora informacji o publikacjach problem poprawnego zarządzania danymi staje się kluczowy. Ponieważ jedną z funkcji PBN jest zbieranie informacji do celów oceny instytucji, naturalne jest, iż przede wszystkim to jednostki odpowiedzialne są za informacje, na podstawie których będzie oceniany ich dorobek. Tam, gdzie ocenie podlegają sami autorzy, również oni muszą mieć możliwość kontrolowania poprawności i kompletności informacji zawartych w ich bibliografiach. Szczególną trudność sprawia to, że poszczególne publikacje mogą wpływać na ocenę więcej niż jednej jednostki i wszystkich autorów, co powoduje konieczność dokonywania uzgodnień pomiędzy nimi w zakresie kształtu i treści rekordu.



Rys. 1. Schemat tworzenia Polskiej Bibliografii Naukowej
Źródło: Opracowanie własne.

Podstawowym źródłem danych dla Polskiej Bibliografii Naukowej są systemy instytucjonalne. W ich przypadku najprostszym sposobem zasilenia PBN jest przygotowanie eksportu danych w formacie XML, a następnie zaimportowanie ich do systemu. Dane mogą być również importowane z innych systemów — baz bibliograficznych czy katalogów bibliotecznych. Te dane są wykorzystywane przede wszystkim w charakterze pomocniczym: służą uzupełnieniu istniejących rekordów

o brakujące informacje i weryfikacji danych, mogą też wspomagać proces wprowadzania nowych rekordów przy pomocy interfejsu WWW.

Zarządzanie treścią bibliografii jednostki realizowane jest przez osobę uprawnioną, która posiada odpowiednie konto i uprawnienia nadane w systemie POL-on. Osoba ta decyduje, opisy których publikacji znajdą się w oficjalnej bibliografii jednostki, a których nie. Ma też pełne uprawnienia w zakresie modyfikacji rekordów. Jeśli opis publikacji znajduje się w bibliografiach kilku jednostek (afiliowana jest przy kilku jednostkach), zmiany wymagają akceptacji wszystkich zainteresowanych. Wyjątkiem jest wprowadzenie niewielkich zmian, które akceptowane są automatycznie, a bibliografowie jednostek są o nich jedynie informowani.

Publikacje mogą być wprowadzane nie tylko na zasadzie importu z lokalnych systemów lub przez wyznaczonych pracowników jednostki, lecz także przez samych autorów, którzy mają ponadto możliwość nanoszenia poprawek do istniejących rekordów. Należy jednak podkreślić, że bibliografowie jednostek muszą zaakceptować publikacje dodane przez autorów oraz wprowadzane przez nich zmiany. Rozwiązanie to ma kilka zalet. Po pierwsze, bibliografie jednostek wprowadzają zazwyczaj tylko wybrane pola, pomijając np. abstrakt czy słowa kluczowe. Autorzy, którzy mają w swoim dorobku kilka bądź kilkanaście publikacji i którym zależy na szerokim udostępnieniu swego dorobku, chętnie uzupełniają takie dodatkowe pola i bardzo dokładnie opiszą publikacje. Drugą zaletą jest możliwość prowadzenia bibliografii w systemie rozproszonym, tj. przy współudziale autorów, gdzie bibliograf jedynie kontroluje ich pracę. Jest to model z powodzeniem wypróbowany na kilku dużych wydziałach i dobrze sprawdzający się w jednostkach o profilu ścisłym, gdzie posługiwanie się serwisami tego typu jest dla pracowników naturalne.

W przypadku systemu o takiej skali, jak Polska Bibliografia Naukowa, pojawia się wiele wyzwań. Najtrudniejszym z nich jest wprowadzenie skutecznych algorytmów deduplikacji, które z jednej strony radzą sobie z niewielkimi błędami osób wprowadzających dane, z drugiej zaś skutecznie chronią przed tworzeniem duplikatów w bazie. Drugim poważnym wyzwaniem jest wyznaczenie ścisłych reguł, które pozwolą na elastyczną konstrukcję modelu współpracy, tak aby jednostki mogły używać PBN w sposób adekwatny do ich indywidualnych potrzeb. W szczególności oznacza to wybór optymalnych reguł, które określają zasady akceptowania zmian w rekordach oraz scalania informacji pochodzących z różnych źródeł. Ostatnim wyzwaniem jest poprawna, mechaniczna identyfikacja osób i jednostek w już istniejących rekordach, co pozwoli na zautomatyzowanie procesu pobierania danych z wielu systemów. Jest to problem nastręczający wielu trudności, nie tyle na poziomie teoretycznym, ile raczej praktycznym, z uwagi na skromną ilość danych, na podstawie których prowadzone jest automatyczne wnioskowanie.

Funkcje dostarczane przez PBN

Głównymi odbiorcami PBN są jednostki naukowe. O ile część dysponuje już własnymi systemami bibliograficznymi, to wiele mniejszych instytucji nie posiada odpowiedniej bazy danych i oprogramowania. W szczególności niewielkie instytuty badawcze nie mają własnych baz dorobku. W takim przypadku PBN pozwala na prowadzenie bazy bibliograficznej jednostki bez ponoszenia dodatkowych nakładów związanych z zakupem sprzętu i oprogramowania oraz obsługą systemów informatycznych. Pozostałe jednostki mogą samodzielnie importować dane ze swoich systemów bibliograficznych, a następnie nimi zarządzać. Obecnie intensywnie rozwijane są moduły odpowiedzialne za raportowanie i analizy zgromadzonych danych, które umożliwią pełne wykorzystanie potencjału systemu.

Dla autorów szczególnie istotna jest możliwość prowadzenia indywidualnej bibliografii w środowisku oficjalnym, która stanowi swoiste publikacyjne portfolio i pozwala na skuteczną promocję w Internecie. Możliwie jest importowanie do PBN danych w formacie BibTeX, kontrolowanie danych oraz eksport cytowań. W przyszłości istotna stanie się też możliwość wykorzystania danych dla potrzeb raportowania rezultatów grantów czy postępowania habilitacyjnego.

Ankieta czasopism

Jedną z funkcji PBN jest gromadzenie informacji o czasopismach w ramach corocznej ankiety ewaluacyjnej. Wszystkie czasopisma, które zainteresowane są przystąpieniem do ewaluacji i umieszczeniem w części B wykazu czasopism naukowych Ministerstwa Nauki i Szkolnictwa Wyższego, zobligowane są do złożenia ankiety właśnie w systemie PBN. Pierwsza ankieta tego typu została zrealizowana na przełomie lat 2011 i 2012, obecnie zaś (czerwiec 2013 r.) trwa druga edycja składanie ankiet.

Podsumowanie

Polska Bibliografia Naukowa została udostępniona we wczesnym stadium rozwoju, tak by możliwe było skorygowanie jej założeń stosownie do potrzeb użytkowników. System znajduje się obecnie w fazie testów beta, które zakończą się najprawdopodobniej pod koniec roku 2013. PBN będzie stopniowo rozbudowywana i wypełniana danymi przez kolejne lata, dzięki czemu stanie się obszerną i wyczerpującą bazą dorobku publikacyjnego polskich naukowców i polskich jednostek naukowych.

Nowiński, A., Sylwestrzak, W., Fenrich, W. Polska Bibliografia Naukowa. W: Bibliograficzne bazy danych i ich rola w rozwoju nauki. II Konferencja naukowa Konsorcjum BazTech, Poznań, 17-19 kwietnia 2013 [on-line]. Stowarzyszenie EBIB, 2013 [Dostęp: 30.08.2013]. Materiały konferencyjne EBIB, nr 24, Dostępny w World Wide Web: http://open.ebib.pl/ojs/index.php/Mat_konf/article/view/46. ISBN 978-83-63458-06-5.