



Mgr Jacek Głębocki



Mgr inż. Grzegorz Pietruszewski

Poznań – Splendor Systemy Informatyczne

PEŁNA AUTOMATYZACJA PROCESU PRZEKAZYWANIA DANYCH DO MODUŁU SPRAWOZDAWCZEGO POLSKIEJ BIBLIOGRAFII NAUKOWEJ NA PRZYKŁADZIE SYSTEMU EXPERTUS

Abstract

The process of automating the transfer of bibliographic data to the reporting module of the Polish Scholarly Bibliography is discussed in the context of the experience of several dozens of universities and research institutes that use the Expertus bibliographic software. The analysis focuses on the subsequent data deliveries and reports in the years 2015-2017.

Streszczenie

Omówiono przebieg procesu automatyzacji przekazania danych bibliograficznych do modułu sprawozdawczego Polskiej Bibliografii Naukowej w latach 2015-2017 na podstawie współpracy z kilkudziesięcioma uczelniami i instytutami badawczymi. Przedstawiono zastosowanie systemu Expertus i przeanalizowano działanie modułu sprawozdawczego Polskiej Bibliografii Naukowej w trakcie kolejnych okresów sprawozdawczych.

Wstęp

Polska Bibliografia Naukowa (PBN) została udostępniona w internecie jesienią 2012 roku [1]. Konieczność zapewnienia możliwości transportu danych do nowo powstałego serwisu, będącego w założeniu częścią Systemu Informacji o Nauce, była brana pod uwagę przez autorów systemu Expertus już w trakcie prac rozwojowych i dostosowawczych dotyczących przekazywania danych do Ankiety Jednostki w roku 2013.

Założenia i określenie celu

Podstawowym celem było zapewnienie bezpośredniego, w pełni zautomatyzowanego i kompletnego transportu danych do Polskiej Bibliografii Naukowej, a jednocześnie taka realizacja zadania, aby:

1) operacja eksportu danych z bazy i ich importu do PBN-MS mogła być obsługiwana bibliotekarzy i docelowo nie nakładała dodatkowych, a zwłaszcza „nieinformatycznych” obowiązków na uczelnianych informatyków,

2) dane nie wymagały jakiegokolwiek obróbki poza systemem Expertus, a w szczególności przygotowywania, przetwarzania, modyfikowania w arkuszach kalkulacyjnych lub innych narzędziach zewnętrznych przed importem do PBN,

3) zaimportowane dane nie wymagały poprawek edycyjnych w PBN,

4) operacja mogła być wykonywana cyklicznie i wszelkie poprawki i aktualizacje rekordu w bazie źródłowej były po imporcie widoczne w PBN,

5) nie została utracona, ani w jakikolwiek sposób zaburzona własna, przemyślana i sprawdzona na przestrzeni wielu lat struktura danych,

6) nie uległa obniżeniu jakości danych gromadzonych w bazach uczelnianych np. przez wymuszone dostosowywanie zapisu tytułów czasopism do obowiązujących w PBN,

7) nie zostały ograniczone funkcje i zakres wykorzystywania programu w sprawozdawczości wewnętrznej i innych polach eksploatacji,

8) system Expertus mógł być stosowany u wszystkich użytkowników w tej samej wersji i zapewniać transport danych polegając wyłącznie na zawartości plików konfiguracyjnych,

9) przynajmniej niektóre dane niewystępujące w bazach uczelnianych a wykorzystywane w PBN mogły zostać uzupełnione przed transportem w sposób przyspieszony, wspomagany operacjami informatycznymi.

Wstępna ocena zagrożeń

Przeprowadzona w drugiej połowie roku 2014 oraz styczniu i lutym r. 2015 analiza funkcji Polskiej Bibliografii Naukowej, dostępnej dokumentacji i wszelkich innych informacji ze źródeł zewnętrznych oraz ocena jakości danych już zgromadzonych w PBN (przede wszystkim rejestrowanych bezpośrednio przez niektóre uczelnie i pracowników nauki) dawała mocne podstawy do przyjęcia, że transport danych do Polskiej Bibliografii Naukowej może być zadaniem bardziej skomplikowanym niż przekazywanie danych do oprogramowania Ankieta Jednostki OPI. Na podstawie doświadczenia obejmującego wszystkie wersje Ankiety od r. 2002 i wobec oczywistości nowych uwarunkowań można było stwierdzić, że:

- 1) oprogramowanie Ankieta Jednostki uzyskiwało wysoki poziom stopniowo,
- 2) nawet ostatnia, bardzo dobrze opracowana wersja z roku 2013 wymagała kilkunastu poprawek i aktualizacji udostępnianych przez Ośrodek Przetwarzania Informacji już w czasie przeznaczonym na przygotowywanie zestawień,
- 3) każda nowa wersja Ankiety wymagała wprowadzania bieżących zmian w systemie Expertus; dotyczyło to zarówno aktualizacji rocznej, jak i zmian udostępnianych przez OPI w trakcie przekazywania danych,
- 4) system importowy Polskiej Bibliografii Naukowej miał być zastosowany po raz pierwszy i z oczywistych powodów nie przeszedł takiej weryfikacji, jak Ankieta Jednostki na przestrzeni lat,
- 5) platforma Polskiej Bibliografii Naukowej nadal oznaczona była jako wersja beta, a zatem „w fazie testów”,
- 6) niemal pewna była konieczność pozyskania dodatkowych danych identyfikacyjnych z systemów uczelnianych.

Przebieg prac, kalendarium

Poniżej przedstawiono działania ze strony autorów systemu Expertus (oznaczane jako **E** bezpośrednio po dacie) na tle innych zdarzeń dot. PBN (daty rozporządzeń, ogłoszeń, opisów stanu serwera PBN itp.)

Zamieszczone tutaj i poniżej w tekście cytaty pochodzą z serwera <http://pbn.nauka.gov.pl> – dostęp wg podanej daty oraz z korespondencji z użytkownikami.

2015.02.24

Ogłoszenie „Ustawy z dnia 15 stycznia 2015 r. o zmianie ustawy o zasadach finansowania nauki oraz niektórych innych ustaw”. Ustawa wchodzi w życie po upływie 3 miesięcy od dnia ogłoszenia.

Na mocy art. 9 jednostki naukowe w terminie 3 miesięcy od dnia wejścia w życie niniejszej ustawy są zobowiązane do wprowadzenia danych do Systemu Informacji o Nauce. W związku z tym terminem przekazania danych jest 25 sierpnia 2015 r.

2015.03.05 (E)

Wstępny harmonogram prac wraz ze wskazaniem możliwych zagrożeń.

2015.03.06 (E)

Opis proponowanej metody przeprowadzenia transportu danych wraz z instrukcją pobrania numerów pbn_id z wykorzystaniem skryptu udostępnianego przez PBN, tworzenia plików mapowania autorów i jednostek.

2015.04.08

Operatorem technicznym PBN zostaje Index Copernicus International.

2015.04.02-05.07 (E)

Pilotażowe wdrożenia modułu transportu danych w wybranych instalacjach.

2015.05.08 (E)

Udostępnienie modułu pozostałym użytkownikom.

2015.05.25

Platforma produkcyjna wciąż oznaczona jest jako wersja beta tj. w fazie testów, a platforma testowa została niespodziewanie wyłączona. Wg informacji udzielanej przez operatora PBN zakończenie prac *„przewidziane jest na 08.06.2015 r. Po tym okresie będzie możliwe masowe zasilanie danymi”*.

2015.07.03

Zmianie ulega udostępniany przez PBN skrypt umożliwiający pobieranie numerów pbn_id autorów. Zmiana w żaden sposób nie jest opisana w serwisie PBN, a dotyczy formatu pliku wejściowego i sposobu wywołania skryptu. Wg informacji użytkowników PBN nie wszyscy z nich, mimo dopełnienia formalności, mają dostęp do funkcji importu (zamiast przycisku „Zaimportuj” jest napis „import niemożliwy”).

2015.07.04

Wejście w życie Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 29 czerwca 2015 r. w sprawie Systemu Informacji o Nauce (Dz.U. 2015 poz. 944 ze zm.) określającego szczegółowy zakres zgłaszanych danych oraz terminy ich przekazywania

2015.07.15, 2015.07.16, 2015.07.31, 2015.07.16 (E)

Cykl internetowych sesji szkoleniowych dla użytkowników systemu Expertus poświęconych realizacji transportu danych do PBN (dwie sesje każdego dnia).

2015.08.07

Nowe wytyczne na stronach PBN (dotychczasowa wersja PBN określona zostaje jako PBN-Repozytorium; zapowiedź nowej wersji systemu określonej jako PBN-Moduł Sprawozdawczy).

2015.08.17 (E)

Udostępnienie analizy nowej sytuacji w obliczu informacji z 7 sierpnia o uruchomieniu nowego modułu nazwanego Modułem Sprawozdawczym.

Najważniejsze punkty analizy:

1) z zamieszczonych informacji wynika, że: *„Dotychczas dostępny system PBN jest systemem repozytoryjnym”* i *„jest to system opracowany pod kątem społecznej rozbudowy i wspólnej pracy nad jakością danych”* oraz wprowadzanie danych do PBN *„nie wypełnia obowiązku nałożonego ustawą”*. Ponadto operator PBN informuje,

że „Z tego powodu postanowiliśmy uruchomić Moduł Sprawozdawczy Polskiej Bibliografii Naukowej – będący jedynym oficjalnym miejscem wprowadzania danych”;

2) kluczowa zmiana polega na tym, że ogłoszona 7 sierpnia specyfikacja plików importowych do Modułu Sprawozdawczego jest różna od specyfikacji stosowanej w Polskiej Bibliografii Naukowej (nazwanej obecnie „PBN-repozytorium”). Operator PBN ogłosił, że *„dotychczasowe środowisko testowe (dla jednostek naukowych) jest nieaktualne”*, a do nowego modułu konieczne są nowe uprawnienia dla osób importujących dane;

3) w obecnej sytuacji obecna wersja modułu transportu danych do PBN-Repozytorium nie może zostać wprost zastosowana w nowym module PBN-Moduł Sprawozdawczy;

4) z uwagi na to, że *„Wprowadzenie danych do Modułu Sprawozdawczego PBN jest możliwe jedynie poprzez import masowy”* konieczne jest wprowadzenie do modułu transportu danych systemu Expertus daleko idących zmian;

5) znaczna część prac wykonanych przez bibliotekarzy nie będzie miała bezpośredniego zastosowania w nowym systemie, np. tzw. mapowanie autorów;

6) konieczne będą nowe prace poszerzające zakres rejestrowanych danych np. wprowadzenie nowych informacji np. o open access, szczegółowa rejestracja danych o monografiach obcych, zmiana wykazu cech pbn (pole nieobowiązkowe);

7) w nowych wytycznych znajduje się stwierdzenie *„Zgodnie z rozporządzeniem jednostki naukowe mają czas na wprowadzenie danych do 30 X 2015 r.”*

2015.10.16

Wersja systemu, która pojawiła się na serwerze testowym 16 października w godzinach wieczornych zawiera liczne nieukończone i nie w pełni działające zmiany; ponadto zaburza działanie funkcji, które poprzednio działały, np. przy próbie importu plików, które można było z powodzeniem importować wcześniej, wyświetla komunikat „nieoczekiwany błąd systemu”. Można zatem odnieść wrażenie, że serwer testowy służy operatorowi pbn do prac programistycznych, co prowadzi do jego czasowej nieużyteczności. Środowisko testowe powinno udostępniać cały czas wersję identyczną z tą, która jest dostępna na serwerze produkcyjnym, co jest niezbędne do przeprowadzania wiarygodnych testów importu. Termin zakończenia upływa za ok. dwa tygodnie, a wersja ostateczna nie jest gotowa.

2015.10.17 (E)

Szczegółowa instrukcja wykorzystania modułu transportu z systemu Expertus do PBN-MS.

2015.10.17

Szczegółowa instrukcja wykorzystania modułu transportu z systemu Expertus do PBN-MS. Problemy przewidywane w analizie wstępnej są obecnie widoczne

w sposób oczywisty i ich uciążliwość jest odczuwalna dla użytkowników, którzy rozpoczęli pracę z systemem „PBN-moduł sprawozdawczy”. Instrukcja dostępna na serwerze pbn.nauka.gov.pl jest niespójna i niezgodna ze stanem faktycznym (np. nieaktualny plik walidacyjny xsd). Operator pbn rozpoczął przeprowadzanie szkoleń konferencyjnych dopiero na przełomie września i października (29 września i 6 października).

2015.10.19

Nowa wersja obejmująca zatwierdzanie rekordów została zainstalowana na platformie produkcyjnej pbn i obecnie funkcja importu nie działa na żadnej z platform (ani produkcyjnej ani testowej). Pojawia się komunikat „wystąpił nieoczekiwany błąd systemu”.

2015.10.24 (E)

Instrukcja pozwalająca na łatwe wykrycie wszystkich błędów w pliku importowym (niezbędna, ponieważ system PBN-MS nadal wykazuje tylko pierwszy błąd w importowanym pliku, co dla osób wykonujących import jest bardzo uciążliwe i niezmiernie czasochłonne).

2015.10.30 (E)

Opracowanie dokumentacji licznych błędów w systemie PBN-MS na dzień 30 października 2015 (patrz także [2] i [3]).

2015.01.30

Zakończenie styczniowego transportu.

2016.04.08

Zmiana operatora technicznego PBN – operatorem technicznym PBN zostaje Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy.

2016.04.04-04.23

Wyłączenie serwerów PBN w związku ze zmianą operatora.

2016.08.04

Nowe wytyczne ze strony PBN dot. zgłaszanie autorów i wprowadzenie nowego pola tzw. liczby pozostałych redaktorów. Udostępnienie nowego, zdecydowanie usprawnionego środowiska importowego PBN.

2016.08.10 (E)

Rozbudowa modułu transportu danych o nowe pola w związku z opublikowaniem istotnych zmian w strukturze PBN i zasad zgłaszania danych (autorzy i redaktorzy wyłącznie z jednostki, nowe pole dot. liczby pozostałych redaktorów, dane o redaktorach monografii w każdym z rozdziałów), automatyczne podłączanie autorów rozdziałów

do zespołu autorskiego całej monografii, obsługę pól monografii wybitnych, wypełnianie inf. o arkuszach, możliwość importu dokumentów bez issn serii/isbn/doi oraz bez stron, uzupełnienie przebiegu diagnostycznego o kontrolę powyższych pól.

2017.01.04 (E)

Rozbudowa modułu transportu danych o inf. o tzw. obszarach wiedzy (dot. tylko autorów niezatrudnionych, afiliujących na wydziałach niejednorodnych).

2017.02.10 (E)

Udostępnienie nowej wersji modułu transportu do PBN-MS.

2017.04.28

Zakończenie parametryzacji.

2017.07.06 (E)

Udostępnianie nowej wersji modułu transportu do PBN-MS pozwalającej na przekazywanie numeru ORCID przy autorach (numery orcid są rejestrowane w bazach systemu Expertus od 2015 roku).

Napotkane problemy

Zasadniczym, fundamentalnym problemem do końca marca 2016 był niezbędnej, pełnej i precyzyjnej dokumentacji:

- 1) dokumentacja dostępna na www była dla praktycznie bezużyteczna, w wielu miejscach rozbieżna z przykładami,
- 2) modyfikacje nie były oznaczane datami, znikwały a na ich miejsce pojawiały się nowe,
- 3) opracowanie programu transportu wymagało nieustannych eksperymentów, przygotowywania rozwiązań wariantowych i doświadczalnej oceny ich przydatności,
- 4) walidacja danych na poziomie edycji była niezwykle uboga,
- 5) moduł importu raportował tylko pierwszy błąd w pliku, nierzadko z bardzo lakonicznym i zagadkowym opisem, import był niezwykle powolny, nierzadko zawieszał się,
- 6) serwer testowy było wyłączany na długie okresy,
- 7) serwer testowy i produkcyjny stosowały różne pliki walidacyjne xsd, ponadto stosowane xsd było różne od udostępnionego w dokumentacji,
- 8) we wrześniu 2015 rekordów podanych jako przykładowe nie można było zaimportować do systemu (mimo wielokrotnych prób realizowanych przez różne ośrodki),
- 9) generowanie sprawozdań było niezwykle powolne, przykładowo 35 rekordów w pliku, a komunikat przy próbie wygenerowania sprawozdania: – „Sprawozdanie zostało złożone, a dokument jest obecnie generowany. Szacowany czas jego wygenerowania wynosi 48 godzin” (22 stycznia 2016),
- 10) mimo przekazywania pbn_id autorów, w systemie nie było żadnej informacji że autor został rozpoznany,

- 11) w ocenie użytkowników odpowiedzi z helpdesku były w dużej części bezużyteczne,
- 12) niezwykle długie oczekiwanie na uaktywnienie kont, przydzielenie uprawnień: *(„... dlaczego dopiero na moją piątą prośbę o nadanie uprawnień managera i importera piszą Państwo o upoważnieniu podpisanym przez kierownika jednostki ...”),*
- 13) istotne błędy nawet w ostatnim dniu zgłaszania sprawozdania,
- 14) zawieszanie się systemu w czasie importu,
- 15) wyświetlanie 50 pierwszych prac, a przy usuwaniu rekordów (opcja dostępna tylko dla wskazanego rekordu wymagała powtarzania zapytania przed usunięciem każdego rekordu, przewijanie się wyników zawsze na początek),
- 16) opcja „pobierz wszystkie w csv” dostępna dopiero po zatwierdzeniu opisów, zawartość raportu nieprzydatna,
- 17) brak kluczowych danych w podglądzie.

Oprócz czynników wymienionych powyżej występowały także inne problemy, nieistnie z punktu widzenia importu, jak np. jakość konwersji danych pomiędzy modułami PBN (PBN-Repozytorium a PBN-MS).

Powyższe problemy udało się rozwiązać nowemu operatorowi. Zasadnicze zalety nowej aplikacji sprawozdawczej uruchomionej latem 2016:

- 1) błyskawiczne i niezawodne działanie procedury importu,
- 2) uproszczenie czynności przedimportowych (import bezpośrednio z xml, a nie z plików xml spakowanych do postaci zip),
- 3) możliwość natychmiastowego podglądu wyników importu,
- 4) bardzo dobra diagnostyka błędów importu pozwalająca na zapoznanie się od razu ze wszystkimi znalezionymi błędami wraz z precyzyjnym opisem każdego błędu,
- 5) prawidłowe rozpoznawanie identyfikatorów autorów (pbn_id).

Uzyskane rezultaty

Dane począwszy od pierwszego transportu były przekazywane terminowo.

Udało się uzyskać pełną automatyzację procesu przekazywania danych do modułu sprawozdawczego Polskiej Bibliografii Naukowej, a w szczególności:

1) jeżeli bibliotekarz posiada wszelkie niezbędne uprawnienia w systemie PBN może samodzielnie dokonywać transportów danych lub – w przeciwnym przypadku – przekazywać gotowe dane wyznaczonym importerom; pomoc specjalistów z zakresu informatyki nie jest niezbędna, aczkolwiek w dużych, wielowydziałowych uczelniach może usprawnić pracę,

2) system tworzy pliki gotowe do importu, zawierające dane kompletne, nie-wymagające wstępnego przetwarzania poza systemem Expertus ani poprawek edycyjnych w PBN,

3) dane są prawidłowo identyfikowane w PBN: autorzy na podstawie pbn_id wraz z numerem orcid, a czasopisma są przekazywane w trybie zgodności z wykazem rocznym lub czteroletnim niezależnie od zapisu w rekordzie źródłowym,

4) operacja może być wykonywana cyklicznie, rekordy posiadają identyfikatory, dzięki czemu dane aktualizują się automatycznie; selekcja rekordów może być bardzo precyzyjna z włączeniem i wyłączaniem dowolnych zbiorów rekordów,

5) struktura i funkcje systemu Expertus zostały rozbudowane, dane i zależności między różnymi rodzajami rekordów są wykorzystywane także do innych celów niż PBN (dodatkowe pola opisu bibliograficznego, powiązania rozdziałów z monografiami, dodatkowe identyfikatory autorów),

6) system Expertus wraz z modulem konwersji zachował jednolitość i może być stosowany w różnych konfiguracjach,

7) pobieranie identyfikatorów zewnętrznych zostało usprawnione.

Wnioski

1) wywiązanie się przez jednostki naukowe z obowiązku sprawozdawczego w pierwszym (2015.10.30) i drugim (2016.02.28) okresie sprawozdawczym nie dowodzi, że system „Polska Bibliografia Naukowa – moduł sprawozdawczy” spełnił pokładane w nim oczekiwania.

2) realizacja zadania odbyła się kosztem ogromnych nakładów i wysiłku wszystkich zaangażowanych: bibliotekarzy, informatyków uczelnianych i innych osób zaangażowanych w prowadzone prace

3) począwszy od transportu w lecie 2016 problemy wynikały jedynie z interpretacji przepisów, a nie z powodów podanych powyżej

4) nowym wyzwaniem jest tzw. Ustawa 2.0 wymuszająca kolejne zmiany

Bibliografia

[1] Klenczon W. Polska Bibliografia Naukowa – nowe szaty bibliografii? – Porządek rzeczy. Blog Pracowni Teorii i Organizacji Bibliografii BN. [Dostęp 2018.08.17]. <http://porzadek-rzeczy.bn.org.pl/polska-bibliografia-naukowa-%E2%80%93-nowe-szaty-bibliografii/>

[2] Rozkosz E. A., Drabek A. Moduł Sprawozdawczy Polskiej Bibliografii Naukowej. – Biuletyn EBIB [on-line], nr 3 (165)/2016. [Dostęp 2018.08.17]. <http://open.ebib.pl/ojs/index.php/ebib/article/view/438>

[3] Komperda A.: Import publikacji Politechniki Wrocławskiej do Modułu Sprawozdawczego PBN – sprawozdanie z realizacji zadania. – Biuletyn EBIB [on-line], nr 3 (165)/2016. [Dostęp 2018.08.17]. <http://open.ebib.pl/ojs/index.php/ebib/article/view/434>