

Integracja APD z Ogólnopolskim Repozytorium Prac Dyplomowych

Janina Mincer-Daszkiewicz, Łukasz Karniewski
Uniwersytet Warszawski, MUCI
jmd@mimuw.edu.pl

Warszawa, 2015-11-16

1. Uczelnie mają do wyboru trzy metody przekazywania danych:
 1. Interfejs WWW - akceptowalne dla małych uczelni,
 2. ORPD jako serwer REST - działa na zasadzie PUSH, APD wysyła dane do ORPD,
 3. APD jako serwer OAI-PMH - działa na zasadzie PULL, ORPD pobiera dane z APD.
2. Uczelniany operator archiwum wskazuje w ORPD lokalizację archiwum (PULL) lub generuje hasło dostępu do serwera REST (PUSH).
3. Uczelniane archiwum autoryzacje ORPD po numerze IP - uczelnia wie z jakiego IP łączy się ORPD i umożliwia mu pobieranie prac.
4. PULL - ORPD łączy się z zarejestrowanymi repozytoriami uczelnianymi periodycznie i pobiera nowe prace.
5. PUSH - APD wysyła prace zgodnie z własnym harmonogramem.

1. Wersja 1.0 ORPD opierała się głównie na protokole OAI-PMH:
 1. nie było możliwe przesłanie kompletu danych wymaganych Ustawą (kierunek studiów, tytuł zawodowy),
 2. protokół OAI-PMH nie wymuszał kompletności metadanych (ORPD pobierało prace bez promotora, recenzenta, daty obrony).
2. W sierpniu 2015 udostępniono wersję 2.0:
 1. nowy model danych dostosowany do obowiązującego prawa,
 2. nowy format przekazywania danych - zmiana z formatu *DSpace Intermediate Metadata* na format *ETD-MS v1.1: an Interoperability Metadata Standard for Electronic Theses and Dissertations*
<http://www.ndltd.org/standards/metadata>,
 3. nowe funkcjonalności w interfejsie: wyszukiwanie prac z błędami, usuwanie prac dyplomowych, scalanie duplikatów, wiązanie prac z wieloma autorami,
 4. lepsza walidacja danych w protokole REST.
3. W listopadzie 2015 wyjdzie wersja 2.1 (obecnie na ORPD DEMO).

1. PULL jest semantycznie proste, co APD wystawi, to ORPD powinno pobrać.
2. PULL nie zapewnia informacji zwrotnej o pobranych pracach (autor ani promotor nie wiedzą, czy praca trafiła do ORPD, administrator systemu nie wie, czy prace faktycznie są pobierane ani kiedy będą obciążone łącza).
3. PULL nie pozwala sterować procesem przesyłania prac (zbiorcza przesyłka prac z danego okresu, ponowna przesyłka prac wcześniej pominiętych lub wymagających korekty).
4. Tylko PUSH (i interfejs WWW) pozwala na przestanie prac z wieloma autorami, dla których data obrony lub dane określające kierunek studiów i uzyskany tytuł zawodowy są różne.
5. **Warto przestawić się z PULL na PUSH!**

1. Praca spełnia warunki wysłania do ORPD w trybie PUSH, jeżeli:
 - jest zarchiwizowana,
 - jest obroniona (dla co najmniej jednego autora jest wpisany numer dyplomu),
 - istnieją pliki pracy.
2. Do ORPD wysyłane są następujące dane:
 - tytuł,
 - opiekunowie,
 - recenzenci,
 - autorzy, a dla każdego autora:
 - data obrony,
 - kierunek studiów, na którym obroniono prace,
 - jednostka podstawowa związana z kierunkiem,
 - forma studiów,
 - tytuł zawodowy,
 - data rozpoczęcia studiów,
 - kod ISM.
3. Brak którejkolwiek z tych informacji spowoduje odrzucenie pracy przez ORPD.

Przykładowe komunikaty błędów:

Forma studiów STACJONARNE podana w kierunku 12402 z datą obowiązywania 2007-10-01 nie istnieje w systemie ORPD.

Instytucja Wydział Biologii nie prowadzi kierunku 23253

Instytucja o kodzie UW/40070000 nie występuje w systemie ORPD.

Kierunek 6081 z datą obowiązywania 2005-10-01 nie występuje w systemie ORPD.

Data obrony nie może być późniejsza od bieżącej

Data określająca wersję danych kierunku musi być wcześniejsza albo równa dacie obrony

Prace w APD: 121 497 (średni rozmiar pracy: 1.84 MB, zajętość dysku: 224 GB)

W testach przesyłano prace obronione po 01-10-2009.

Test z 30-09-2015:

Liczba prac wybranych do przestania: 69451

Odrzucone przez APD: 589

Odrzucone przez ORPD: 1522

Przesłane: 67340

Czas: 25 godz.

Test z 04-11-2015:

Liczba prac wybranych do przestania: 69450

Odrzucone przez APD: 589

Odrzucone przez ORPD: 1482

Przesłane: 67379

Czas: 19 godz.

Przesyłanie prac metodą PUSH jest realizowane za pomocą skryptu uruchamianego z poziomu systemu operacyjnego serwera APD.

Wymagane jest podanie identyfikatora i hasła instytucji w ORPD (do wygenerowania w interfejsie ORPD).

Domyślnie do przestania wybierane są tylko nowe prace (takie, których jeszcze nigdy nie próbowano wysłać do ORPD).

Przed wysłaniem pracy do ORPD przeprowadzana jest wstępna walidacja danych. Jeśli braki w danych są na tyle duże, że ORPD na pewno odrzuci pracę, to taka praca nie jest wysyłana.

Szczegółowy log z przebiegu przesyłania prac zapisywany jest w pliku. Wynik przestania każdej pracy zapisywany jest również w bazie danych.

Przesyłanie prac można uruchomić w trybie symulacji, aby przeprowadzić samą wstępną walidację danych bez faktycznego wysyłania ich do ORPD.

Dla każdej pracy tworzony jest wpis w bazie danych zawierający:

- opis brakujących danych, jeśli praca nie przeszła walidacji
- komunikat zwrotny z ORPD
- status przestania pracy

Te informacje pozwalają na ustalenie, ile i które prace zostały:

- wybrane do przestania
- poprawnie zdeponowane
- zdeponowane mimo braków w danych
- nie wysłane z powodu braków w danych
- wysłane, lecz odrzucone przez ORPD z powodu błędów

Dzięki temu możliwe jest uzupełnienie danych prac w bazie USOS i ponowienie ich przestania do ORPD. Zapisane statusy prac pozwalają na pominięcie prac, które zostały już poprawnie zdeponowane.

Odpowiedź z serwera REST

W przypadku powodzenia usługa zwraca kod **200** i status **OK** wraz z komunikatem o poprawnym zdeponowaniu pracy. W przypadku gdy zdeponowanie pracy nie powiodło się, usługa zwraca błąd.

Kod odpowiedzi	Znaczenie
200 OK	Sukces
400 Bad request	Np. format thesisMetadata.xml niepoprawny
401 Unauthorized	Np. błędne dane w polach User Auth.
403 Forbidden	Dana zautentykowana instytucja nie ma uprawnień do zdeponowania prac instytucji przypisanej do pracy
405 Not found	Np. nie znaleziono instytucji o podanym w adresie url kodzie

Błędy w danych z UW

PULL, ORPD wersja 1.0

111 032 – prace z autorami bez tytułu zawodowego (wszystkie przekazane),
111 032 – prace z autorami bez danych o ukończonym kierunku studiów (wszystkie przekazane),
2 171 – prace bez promotora,
1 728 – prace bez recenzenta.

PUSH, ORPD wersja 2.0

Wyszukiwarka prac z błędami

- Prace z autorami bez daty obrony
- Prace z autorami bez tytułu zawodowego
- Prace z autorami bez danych ukończonego kierunku studiów
- Prace bez promotora
- Prace bez recenzenta
- Prace z Uczelnią/Jednostką podstawową niewypełnioną lub błędnie wypełnioną

Szukaj

Czyść

Wyniki wyszukiwania

Brak prac spełniających zadane kryteria

Praca przekazana do ORPD w wersji 1.0 protokołem OAI-PMH

POMOC

Zalogowano: *jmd* [Wyloguj](#)

Centralne Repozytorium Prac Dyplomowych

[Wyszukiwarka prac](#)

[Dodaj pracę](#)

[Hasła repozytorium inst.](#)

[Serwery OAI-PMH/ORE](#)

[Generuj link tymczasowy](#)

Praca dyplomowa

Tytuł: **"Hanemann" Stefana Chwina jako wyzwanie rzucone pamięci zbiorowej**

Autor: Ewa86437 Kowalska86437

Promotor: Ewa93746 Kowalska93746, Ewa67480 Kowalska67480

Recenzent: Ewa67480 Kowalska67480, Ewa93746 Kowalska93746

Jednostka organizacyjna: Wydział Polonistyki

Uczelnia: Uniwersytet Warszawski

Data obrony: 24.06.2005

Pliki: [resource1 \(180,79 kB\)](#)

Data wprowadzenia: 10.10.2014

Osoba: harvester

IP: 127.0.0.1

[Edytuj](#)

Praca przekazana do ORPD w wersji 2.0 protokołem REST

POMOC

Zalogowano: *jmd* [Wyloguj](#)

Ogólnopolskie Repozytorium Prac Dyplomowych

[Wyszukiwarka prac](#) [Wyszukiwarka prac z błędami](#) [Dodaj pracę](#) [Hasła repozytorium inst.](#) [Serwery OAI-PMH/ORE](#) [Generuj link tymczasowy](#)

Praca dyplomowa

Tytuł: **"Hanemann" Stefana Chwina jako wyzwanie rzucone pamięci zbiorowej**

Autor: Ewa86437 Kowalska86437
Data obrony 24.06.2005
Kierunek studiów: Filologia polska
Forma studiów: Niestacjonarne
Uzyskany tytuł zawodowy: Licencjat

Promotor: Ewa93746 Kowalska93746 , Ewa67480 Kowalska67480

Recenzent: Ewa67480 Kowalska67480 , Ewa93746 Kowalska93746

Jednostka podstawowa: Wydział Polonistyki

Uczelnia: Uniwersytet Warszawski

Pliki: [T109.pdf \(4,86 MB\)](#)

Data wprowadzenia: 27.09.2015

Osoba: Uniwersytet Warszawski

IP: 193.0.109.15

[Edytuj](#) [Usuń](#)

Praca przekazana do ORPD w wersji 2.0 protokołem REST

Praca wspólna wielu autorów

Praca dyplomowa

Tytuł: [System inForm - zarządzanie zadaniami i przetwarzanie zgłoszeń](#)

Autor: Ewa132887 Kowalska132887

Data obrony 05.07.2011

Kierunek studiów: Informatyka

Forma studiów: Stacjonarne

Uzyskany tytuł zawodowy: Licencjat

Jan132861 Kowalski132861

Data obrony 15.09.2010

Kierunek studiów: Informatyka

Forma studiów: Stacjonarne

Uzyskany tytuł zawodowy: Licencjat

Jan132802 Kowalski132802

Data obrony 08.07.2010

Kierunek studiów: Informatyka

Forma studiów: Stacjonarne

Uzyskany tytuł zawodowy: Licencjat

Jan132799 Kowalski132799

Data obrony 16.09.2010

Kierunek studiów: Informatyka

Forma studiów: Stacjonarne

Uzyskany tytuł zawodowy: Licencjat

Promotor: Jan319 Kowalski319

Recenzent: Jan319 Kowalski319 , Jan593 Kowalski593

Jednostka podstawowa: Wydział Matematyki, Informatyki i Mechaniki

Uczelnia: Uniwersytet Warszawski

Pliki: [T078.pdf \(2,43 MB\)](#)

Data wprowadzenia: 26.09.2015

Osoba: Uniwersytet Warszawski

IP: 193.0.109.15

Edytuj

Usuń

Rekomendacje

1. Format danych przesyłanych do ORPD w trybie PUSH różni się od formatu stosowanego w trybie PULL.
2. Jeżeli uczelnia dotychczas przekazywała prace w trybie PULL, to zmiana na tryb PUSH spowodowałaby ponowne przestanie wszystkich danych w nowym formacie, w wyniku czego prace po stronie ORPD zostałyby zdublowane (nieprawidłowości w starych rekordach nie zostałyby skorygowane).
3. Optymalnym rozwiązaniem jest usunięcie wszystkich prac uczelni z ORPD i ponowne ich przestanie, już w trybie PUSH.
4. OPI (instytucja utrzymująca ORPD), na wniosek rektora uczelni, usunie dane z produkcyjnej bazy danych ORPD.
5. Wcześniej warto, w ramach testów, przestać prace do wyczyszczonej bazy danych ORPD DEMO.