

Helena Gałek

Biblioteka Politechniki Krakowskiej

helena.galek@pk.edu.pl

ORCID: <https://orcid.org/0000-0003-3051-6465>

Dane badawcze Politechniki Krakowskiej na platformie RODBUK

Streszczenie: W artykule omówiono problematykę funkcjonowania repozytorium danych badawczych RODBUK (wspólnej inicjatywy uczelni krakowskich) ze szczególnym uwzględnieniem rozwiązań organizacyjnych i merytorycznych Politechniki Krakowskiej. Przedstawiono szereg inicjatyw Biblioteki Politechniki Krakowskiej w zakresie funkcjonowania repozytorium, współpracy ze środowiskiem akademickim jak i wkład pracy w proces gromadzenia, zarządzania i archiwizacji danych badawczych macierzystej uczelni.

Słowa kluczowe: RODBUK, dane badawcze, repozytorium danych badawczych, data steward, zarządzanie danymi badawczymi, współpraca z pracownikami naukowymi

Wprowadzenie

Od kilku lat w środowisku bibliotekarzy szkół wyższych rośnie zainteresowanie tematyką zarządzania danymi badawczymi. Po długich latach zmagania się z problematyką dotyczącą otwierania publikacji naukowych przyszła kolej na dane wytwarzane na rodzimych uczelniach w toku badań naukowych. Początki nigdy nie są łatwe, podobnie jest i w tym przypadku. Wielu z nas musiało bardzo szybko doksztalić się w tym zakresie, aby sprostać potrzebom naukowców i móc służyć im pomocą zarówno w zaplanowaniu pracy z danymi – najczęściej poprzez korektę planu zarządzania danymi (ang. DMP Data Management Plan), organizacji danych podczas badań, jak i ich udostępnieniu w sposób otwarty dla społeczeństwa.

Warsztat bibliotekarza

Narzucony przez jednostki finansujące (NCN czy Horyzont Europa) obowiązek upublicznienia danych badawczych w zakresie niezbędnym do weryfikacji twierdzeń zawartych w publikacjach naukowych zmusił pracowników Oddziału Informacji Naukowej Biblioteki Politechniki Krakowskiej do szybkiego zapoznania się z wytycznymi grantodawców, a początkowo przede wszystkim z zasadami wypełniania planów DMP. Pracownicy naukowcy nie od razu zwracali się o pomoc do biblioteki, jednak ci, którzy do nas trafili, liczyli na wsparcie merytoryczne i korektę, co zapewniałaby im pozytywną ocenę dokumentu. Nie mogliśmy realizować tych zadań bez odpowiedniego warsztatu. Samokształcenie bibliotekarzy w tym zakresie opierało się na przeczytaniu wielu dokumentów i regulaminów, m.in. dotyczących zasad przystąpienia do grantu, zastosowania zapisów prawa autorskiego i odpowiednich licencji Creative Commons (CC) oraz odbycia szeregu szkoleń. Jedną z form kształcenia był udział w webinarium prowadzonych zarówno przez koleżanki i kolegów z polskich bibliotek, którzy już dłuższy czas współpracują z badaczami i wiedzieli, na co zwracać uwagę, a także zagranicznych, którzy mogli podzielić się swoimi doświadczeniami. To właśnie na zagranicznych przykładach mogliśmy określić, jakich narzędzi potrzebujemy w bibliotece oraz co

możemy zaproponować naukowcom w tym zakresie. Ze względu na początkowo małe zapotrzebowanie na naszą pomoc nie zdecydowaliśmy się na ukończenie certyfikowanych szkół czy kursów data stewardów.

Repozytorium Otwartych Danych Badawczych Uczelni Krakowskich – RODBUK

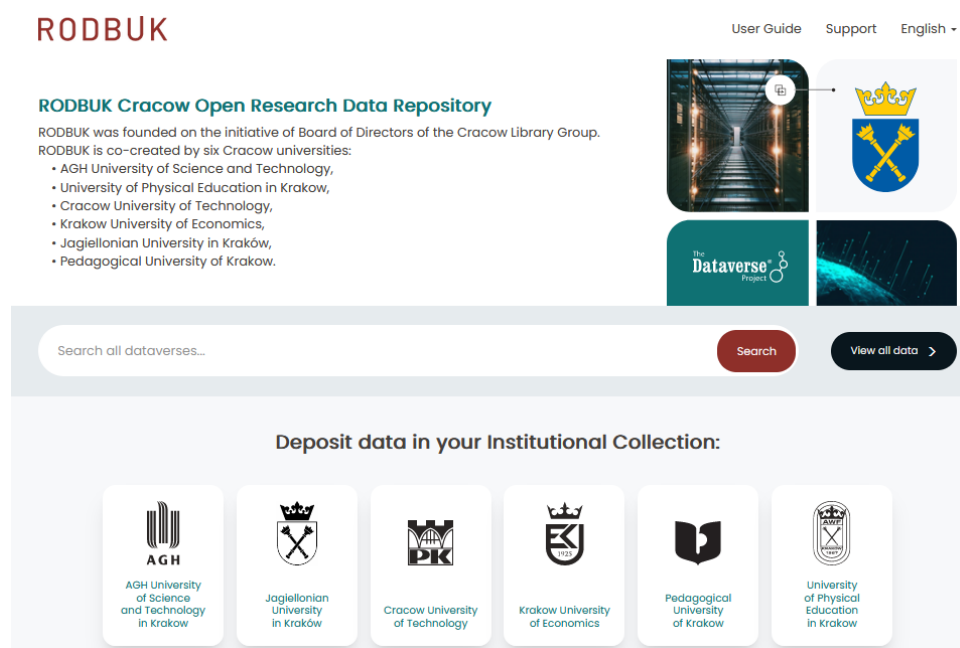
W czerwcu 2021 r. członkowie Rady Dyrektorów [Krakowskiego Zespołu Bibliotecznego](#)¹ (KZB) powołali [zespół roboczy](#), który we współpracy z Akademickim Centrum Komputerowym Cyfronet miał za zadanie opracować wytyczne dla repozytorium otwartych danych badawczych jako wspólnej inicjatywy uczelni krakowskich. Pracownicy Biblioteki Politechniki Krakowskiej aktywnie włączyli się w pracę Zespołu ds. utworzenia międzyuczelnianego repozytorium danych badawczych, który zainicjował powstanie Repozytorium Otwartych Danych Badawczych Uczelni Krakowskich (RODBUK). Spotkania zespołu odbywały się w trybie cotygodniowym. Na każdym z nich omawiany był jeden konkretny problemem czy zagadnienie. Początkowo rozpatrywano standardową budowę repozytorium, tzn. jednego „zarządcę” interfejsu i systemu oraz administratorów instytucjonalnych kolekcji jednostek. Jednakże po dołączeniu do zespołu Akademickiego Centrum Komputerowego Cyfronet AGH (ACK Cyfronet AGH), które miało nas wspierać technologicznie, wypracowaliśmy inną koncepcję współpracy i organizacji działania systemu. Oficjalne otwarcie repozytorium RODBUK odbyło się 6 marca 2023 r. w Bibliotece Jagiellońskiej, natomiast 28 czerwca 2023 r., prawie dwa lata od rozpoczęcia prac, oficjalnie uruchomiła swoją instancję Politechnika Krakowska (PK). Wcześniej, oprócz Uniwersytetu Jagiellońskiego (UJ), narzędzie wdrożono w Akademii Górniczo-Hutniczej im. S. Staszica oraz Uniwersytecie Ekonomicznym w Krakowie. Wszystkie te jednostki mają strukturę hierarchiczną, odzwierciedlającą strukturę uczelni (w zależności od decyzji administratorów instancji może być bardziej lub mniej rozbudowana). Politechnika Krakowska zdecydowała się na utworzenie 10 kolekcji: dla każdego wydziału (8) oraz dodatkowo dla Szkoły Doktorckiej oraz Biblioteki PK. Obecnie w repozytorium RODBUK jest łącznie 77 kolekcji (uczelni – 6, wydziałów – 61, jednostek organizacyjnych – 15).

Działanie systemu oparte jest na oprogramowaniu Dataverse (aktualna wersja 5.14) i zastosowanie podstawowego opisu metadanych zgodnego z obowiązującymi normami, z możliwością importu do ośmiu standardów m.in. Dublin Core, DDI (Data Documentation Initiative), JSON czy OpenAIRE. Staramy się także spełniać wszystkie możliwe wytyczne otwartości i jednostek finansujących granty, stąd też repozytorium jest indeksowane przez [re3data.org](#) oraz [OpenDOAR](#).

Obecna struktura repozytorium RODBUK składa się z niezależnych instancji (kolekcji), osobnych dla każdej instytucji (model rozproszony). Każda uczelnia zarządza swoją kolekcją i może wprowadzać indywidualne zmiany dostosowane do swoich potrzeb (w pewnym zakresie). Obecnie jest ich sześć, w tym jedna Politechniki Krakowskiej. Pracownicy naukowcy deponują dane w swoich instancjach, a ich dane podlegają korekcie uczelnianych opiekunów danych/data stewardów. Wszystkie dane zgromadzone są na serwerze administratora, tj. ACK Cyfronet AGH. Administrator odpowiada także za agregator, tj. stronę, na której pojawiają się dane wszystkich instancji, możliwe do wspólnego przeszukiwania i przeglądania. Takie rozwiązanie pozwala na swobodne zarządzanie cyklem życia danych na uczelniach. Staraliśmy się opracować i utrzymać jednolitość zasad dotyczących deponowania danych, przyjmując jeden, wspólny [Regulamin korzystania z Repozytorium Otwartych](#)

¹ Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dn. 6.11.2023 r.

[Danych Badawczych Uczelni Krakowskich](#). Co istotne, repozytorium jest otwarte także dla innych jednostek, które chciałyby deponować w nim swoje dane. Nie jest także ważne umiejscowienie topograficzne, pomimo iż nazwa sugeruje udział tylko uczelni z Krakowa. Jednostki zainteresowane dołączeniem do RODBUK-a muszą zgłosić się do koordynatora zespołu, zgodnie z informacją na stronie KZB.



Il. 1. Strona informacyjna repozytorium RODBUK

Źródło: RODBUK [online]. [Dostęp 01.10.2023]. Dostępny w: <https://rodbuk.pl/>.

Deponowanie danych badawczych składa się z trzech kroków: wypełnienia m.in. obowiązkowych metadanych, wybrania odpowiedniej licencji CC oraz zdeponowania plików. Metadane można dodatkowo podzielić na:

- opisowe – zawierające opis zasobu, tj. tytuł, twórcę, słowa kluczowe,
- strukturalne – opisujące powiązania i zależności pomiędzy zasobami/opisami, tj. link do publikacji omawiającej zamieszczone dane, do innego datasetu w ramach tego samego projektu oraz
- administracyjne – dotyczące zarządzania zasobem, tj. typ plików, licencja CC, data zdeponowania zasobu.

Do 24 października 2023 r. w RODBUK-u zdeponowano 194 datasety zawierające łącznie 3032 pliki. Znaczna większość dotyczy danych wytworzonych w roku 2023 (108 rekordów) i 2022 (46 rekordów). Na razie tylko dwie, „najmłodsze”, instancje nie mają jeszcze zdeponowanych żadnych danych badawczych (Akademii Wychowania Fizycznego w Krakowie oraz Uniwersytetu Komisji Edukacji Narodowej w Krakowie). Zdecydowana większość datasetów znajduje się w instancji Uniwersytetu Jagiellońskiego (163 datasety z prawie 2500 plikami), na drugim miejscu plasuje się Akademia Górniczo-Hutnicza (28 datasetów z prawie 530 plikami), na trzecim miejscu Politechnika Krakowska z zaledwie dwoma datasetami (zawierającymi 5 plików), zaś na czwartym Uniwersytet Ekonomiczny z jednym datasetem (zawierającym 4 pliki). W RODBUK-u zliczane są także ilości pobrań poszczególnych datasetów – do 24.10.2023 r. pliki zostały pobrane ponad 360 razy. Mając

świadomość, iż pracownicy naszych uczelni publikowali/publikują swoje dane w innych repozytoriach, polskich i światowych, w RODBUK-u możliwe jest pobranie metadanych z zewnętrznych serwisów. Takie rozwiązanie zastosowaliśmy już w przypadku repozytorium [Zenodo](#), inne znalezione przez nas źródła będą dodawane do RODBUK-a w późniejszym czasie, zgodnie z możliwościami technologicznymi.

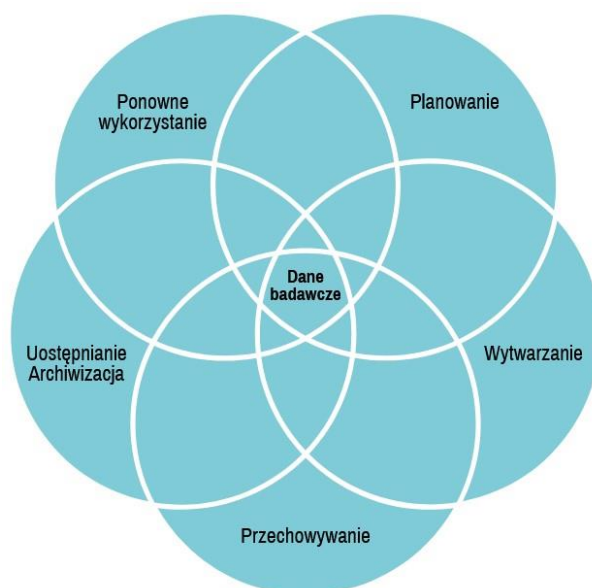
W repozytorium zastosowaliśmy dwie kategorie przedmiotowe: jedną zgodną z wykazem dziedzin i dyscyplin naukowych MEiN i drugą zgodną z wykazem OECD (Organisation for Economic Cooperation and Development). Trzy najpopularniejsze kategorie przedmiotowe, w których badacze uczelni krakowskich najczęściej publikują dane to:

- dziedzina nauk ścisłych i przyrodniczych / dyscyplina „nauki biologiczne” (46 datasetów),
- dziedzina nauk ścisłych i przyrodniczych / dyscyplina „nauki chemiczne” (31 datasetów),
- dziedzina nauk społecznych / dyscyplina „psychologia” (24 datasety).

Aby ułatwić poruszanie się po repozytorium, zastosowaliśmy możliwość przeglądania według instancji, roku czy kategorii tematycznej oraz wyszukiwanie proste i zaawansowane. To ostatnie pozwala na wyszukiwania w obrębie kolekcji (instancji/wydziału/jednostki organizacyjnej), opisu metadanych (19 kryteriów wyszukiwawczych, np. dane dotyczące projektu/grantu, słów kluczowych) oraz plików (7 kryteriów np. ID pliku, opis czy rodzaj danych). Najczęściej jednak wykorzystywane jest wyszukiwanie proste, umożliwiające przeszukiwanie jednocześnie większości pól.

Cykl życia danych badawczych oczami pracowników Politechniki Krakowskiej

Całość zagadnień związanych z poszczególnymi etapami pracy związanymi z powstawaniem, organizacją i archiwizacją danych można podzielić wg cyklu życia danych badawczych (rozumianego jako następujące po sobie etapy procesu naukowego, wynikające jeden z drugiego).



Il. 2. Cykl życia danych badawczych
Źródło: opracowanie własne.

Planowanie

Ten etap jest najczęściej łączony z powstaniem idei przeprowadzenia konkretnego badania, a co za tym idzie zaplanowaniu organizacji całego procesu badawczego. W przypadku starania się o finansowanie zewnętrzne najczęściej oznacza to utworzenie dokumentacji projektowej zawierającej plan DMP. Po zdobyciu odpowiedniej wiedzy pracownicy biblioteki potrafią doradzić i wskazać naukowcom podstawowe błędy lub braki w zapisach planu, podobnie jak ustalić, czy będą oni korzystać z danych już istniejących czy też będą „tylko” wytwarzać nowe. Zachęcamy do przeprowadzenia wyszukiwań w istniejących źródłach wiedzy i weryfikacji czy planowane dane są unikatowe, tzn. czy nikt wcześniej już takich badań nie przeprowadził lub nie zaplanował.

Wytwarzanie/przechowywanie

Podczas procesu badawczego zwracamy uwagę pracowników na kompletność dokumentacji oraz jasne i precyzyjne opisywanie każdego z jej elementów. W tym etapie jesteśmy tylko tzw. pomocą doraźną. Niestety pracownicy bardzo rzadko potrzebują wsparcia bibliotekarzy na tym etapie, z czym często wiążą się problemy w kolejnych stadiach cyklu życia danych. Pracownicy uważają, iż wytwarzanie łączy się wyłącznie z prowadzeniem badań, w merytorycznym tego słowa znaczeniu, wszystkie inne operacje są nieistotne lub mniej ważne. Dlatego też zwracamy uwagę dodatkowo na:

- stosowanie wybranej metodologii oraz procedur i opisanie ich w pliku *read-me*,
- zachowanie ustawień dot. np. kalibracji sprzętu lub zasad wyboru wytworzonych danych,
- stosowanie jednoznacznych i jasnych nazw plików (najlepiej z określeniem stanu na dzień... lub wersji),
- zabezpieczanie danych, poprzez „ohasłowanie”, tworzenie kopii zapasowych zgodnie z zasadą 3–2–1, a w przypadku danych wrażliwych szyfrowanie i anonimizację,
- zachowanie informacji potrzebnych do uzupełnienia metadanych w repozytorium.

Udostępnianie/archiwizacja

Etap ten wiąże się z deponowaniem podstawowego zestawu danych (tych powiązanych z publikacją) w określonym repozytorium, w naszym przypadku instancji Politechniki Krakowskiej w RODBUK-u. Podczas korekty planu DMP nie mogliśmy wcześniej wskazać żadnego bezpłatnego polskiego repozytorium danych badawczych, które odpowiadałoby zaleceniom grantodawców. Po przeprowadzeniu rozpoznania okazało się, że na politechnice koordynatorzy projektów nie mieli też żadnych odgórnych zaleceń czy wskazań związanych z udostępnianiem swoich danych, więc musieli sami wyszukiwać i deponować zgromadzone dane w repozytoriach odpowiadających wymogom jednostek finansujących, np. indeksowanych w odpowiednich wykazach, nadających identyfikatory obiektów cyfrowych czy też posiadających odpowiedni certyfikat jakości. Obecnie dwa pierwsze wymagania nie stanowią już dużego wyzwania zarówno dla twórców repozytoriów w ich spełnieniu, jak i dla pracowników w ich odnalezieniu. Trzeci warunek – zalecany – jest już bardzo trudny do spełnienia. Obecnie niecałe sto repozytoriów na świecie posiada ważny certyfikat jakości [CoreTrustSeal](#). Do zeszłego roku było też jedno z Polski – CLARIN-PL Language Technology Centre. Aktualnie, od 10 października 2023 r., certyfikat jakości został przyznany repozytorium Most Wiedzy Politechniki Gdańskiej. Na tym etapie pomagamy naszym pracownikom zdeponować dane w naszym systemie. Obecnie trwa wypełnianie raportów końco-

wych z grantów NCN, z którymi zgłaszają się pracownicy. Dzięki temu wiemy także, jakie dane trafią do nas, a jakie mogą trafić w inne miejsca, np. [Zenodo](#).

Na Politechnice Krakowskiej nie ma jeszcze określonych zasad korzystania z repozytorium RODBUK. Nie zostały wprowadzone rozwiązania obligatoryjne, natomiast jako biblioteka, cały czas rekomendujemy naszą inicjatywę. Dla pracowników, którzy wybrali lub wybiorą inne miejsce staramy się wprowadzać metadane do naszego systemu, poprzez funkcje API poszczególnych repozytoriów.

Ponowne wykorzystanie

Ten etap nie jest jeszcze dobrze opracowany. Wiąże się to przede wszystkim z brakiem wiedzy, jak dane już wytworzone przez pracowników PK egzystują w sieci. Część z nich została odnaleziona w innych repozytoriach czy serwisach naukowych, część pozostała ograniczona, poprzez embargo lub niechęć twórców do ich upublicznienia w otwartym dostępie. Są prowadzone akcje informacyjne dotyczące szeroko rozumianej otwartości oraz zasad otwartej nauki, która niejako nakazuje upublicznianie wyników badań naukowych zwłaszcza tych finansowanych ze środków publicznych.

Plany...

Jako Biblioteka Politechniki Krakowskiej jesteśmy nadal w początkowej fazie pracy z danymi badawczymi wytworzonymi przez naszych pracowników. Planujemy dalsze pogłębianie naszej wiedzy poprzez udział w szkoleniach organizowanych przez bibliotekarzy oraz naukowców. Jesteśmy otwarci na głosy naszej społeczności akademickiej w zakresie potrzeb i oczekiwań. Realizując prace w obszarze zarządzania platformą RODBUK i gromadzonymi danymi badawczymi, włączamy się w szeroko zakrojoną akcję promowania badań naukowych rodzimej uczelni. Czekają nas także prace nad ciągłym rozwojem repozytorium RODBUK oraz dalsza współpraca pomiędzy uczelniami, z którymi nadal chcemy dążyć do celu, jakim jest sprawne zarządzanie systemem jako pełnym źródłem wiedzy o danych badawczych uczelni krakowskich.