

Repozytoria danych otwarte dla wszystkich badaczy

Od 2020 roku naukowcy w Polsce mogą korzystać z trzech nowych repozytoriów danych badawczych. Dwa z nich mają charakter dziedzinowy: Repozytorium Danych Społecznych (RDS) służy do udostępniania jakościowych i ilościowych danych z nauk społecznych, zaś Macromolecular Xtallography Raw Data Repository (MX-RDR) do udostępniania surowych danych krystalograficznych. Trzeci serwis to Repozytorium Otwartych Danych (RepOD), czyli repozytorium ogólnego przeznaczenia, w którym deponować można dane ze wszystkich dyscyplin. Od 2015 roku działało ono w wersji pilotażowej. Teraz uruchomione zostało w nowej odsłonie w ramach projektu Dziedzinowe Repozytoria Otwartych Danych Badawczych, realizowanego przez Uniwersytet Warszawski we współpracy z Uniwersytetem im. Adama Mickiewicza w Poznaniu oraz Instytutem Filozofii i Socjologii Polskiej Akademii Nauk, partnerami zaangażowanymi m.in. w rozwój zasobów w repozytoriach dziedzinowych.

Wszystkie repozytoria wykorzystują Dataverse – oprogramowanie *open source* stworzone na Uniwersytecie Harvarda, które zostało zmodyfikowane na potrzeby projektu przez programistów z Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego (ICM UW). Oferuje ono szereg funkcji pozwalających na przechowywanie, opisanie oraz udostępnianie danych badawczych w sposób, który ułatwia spełnienie zasad FAIR oraz uwzględnia specyfikę dziedzinową, na przykład odpowiednią konfigurację schematów metadanych. Zgodnie z zasadami FAIR dane powinny być łatwe do znalezienia (*findable*), dostępne (*accessible*), interoperacyjne (*interoperable*) oraz możliwe do ponownego wykorzystania (*reusable*). Repozytoria, jako narzędzia informatyczne tworzone z myślą o gromadzeniu i udostępnianiu rezultatów badań naukowych, uwzględniają dobre praktyki i rekomendacje w zakresie komunikacji naukowej, m.in. zasady FAIR, a także potrzeby badaczy i instytucji naukowych. Udostępnianie danych w repozytoriach jest najlepszym i najbezpieczniejszym sposobem dzielenia się wynikami badań, zalecanym przez instytucje finansujące badania naukowe, które wdrażają polityki otwartości.

Wszystkie repozytoria powstałe w ramach projektu korzystają z trwałych identyfikatorów: nadają zestawom danych numery DOI (Digital Object Identifier) oraz umożliwiają wprowadzenie przypisanych badaczom numerów ORCID (Open Researcher and Contributor ID) czy przypisanych instytucjom numerów ROR (Research Organization Registry). Trwałe i unikalne identyfikatory ułatwiają wyszukanie danych, zacytowanie ich, a także powiązanie z innymi rezultatami badań – zarówno publikacjami, jak i danymi badawczymi. Służą do tego osobne pola metadanych, w których użytkownik może wskazać odpowiednie obiekty wraz z identyfikatorami, zarówno powszechnie stosowanymi w artykułach naukowych numerami DOI, jak i innego typu identyfikatorami, np. ISSN, ISBN, handle czy arXiv.

Ponadto w repozytoriach możliwe jest wskazanie źródła finansowania, co z kolei łączy rezultaty badań z odpowiednimi informacjami na temat instytucji finansującej i programu grantowego. Podanie tego typu danych wymagane jest na przykład przez Narodowe Centrum Nauki czy Komisję Europejską w programach Horyzont 2020 oraz Horyzont Europa. Repozytoria pozwalają także w jasny sposób określić warunki korzystania z danych poprzez wybranie stosownej licencji, m.in. rekomendowanych i powszechnie stosowanych licencji Creative Commons.

W repozytoriach dziedzinowych zastosowano dodatkowe rozwiązania: w Repozytorium Danych Społecznych (RDS) – schemat metadanych dla nauk społecznych oraz słowniki CESSDA, natomiast w Macromolecular Xtallography Raw Data Repository (MX-RDR) – schemat metadanych krystalograficznych oraz automatyczną analizę programem XDS, pozwalającą na wstępną weryfikację poprawności wprowadzonych metadanych. Repozytoria te stanowią ponadto miejsce udostępnienia opracowywanych w ramach projektu danych.

Naukowcy korzystają mogą ze wszystkich repozytoriów na ogólnych zasadach opisanych w regulaminach serwisów. Deponowanie danych możliwe jest po założeniu indywidualnego konta w wybranym repozytorium. Jest to proces szybki i bezpłatny. Po rejestracji użytkownicy mogą deponować zbiory danych, na które składają się metadane oraz pliki z danymi. W przypadku każdego pliku należy określić zasady jego wykorzystania, np. przypisując mu jedną z licencji Creative Commons. Dodatkowo Repozytorium Otwartych Danych (RepOD) umożliwia instytucjom naukowym założenie i prowadzenie kolekcji instytucjonalnych, czyli wydzielonych katalogów służących do gromadzenia, udostępniania i promowania danych wytwarzanych w ramach badań prowadzonych w konkretnej uczelni czy instytucji. Kolekcja dostarcza informacji o dorobku naukowym instytucji oraz ułatwia prowadzenie statystyk i zarządzanie rezultatami badań.

Instytucje naukowe mogą samodzielnie sprawować nadzór merytoryczny nad swoimi kolekcjami, współpracując przy ich prowadzeniu z ICM UW. W praktyce, szczególnie w wypadku większych instytucji, w ich obrębie często wyznaczane są osoby opiekujące się kolekcją. Do ich zadań należy np. weryfikacja wprowadzanych przez naukowców danych, sprawdzenie poprawności metadanych, a niekiedy także wspieranie badaczy w innych kwestiach, takich jak wybór formatów plików czy wybór odpowiednich licencji. Szczegółowy zakres ich obowiązków każdorazowo określa sama instytucja.

Uruchomienie serwisów otwartych dla wszystkich zainteresowanych stanowiło jeden z celów projektu Dziedzinowe Repozytoria Otwartych Danych Badawczych, który uwzględniał także działania szkoleniowe i informacyjne. Od końca 2019 roku prowadzone były szkolenia z zarządzania danymi badawczymi, które w 2020 i 2021 roku odbywały się online i cieszyły się bardzo dużym zainteresowaniem.

Czas realizacji projektu, obejmujący okres od sierpnia 2018 do października 2021, zbiegł się z istotnymi dla polskiego środowiska akademickiego inicjatywami w zakresie danych badawczych. W 2019 roku Narodowe Centrum Nauki wprowadziło wymóg sporządzania planu zarządzania danymi badawczymi, natomiast w 2020 roku przyjęło politykę otwartości, która w pewnym zakresie obejmuje także dane badawcze (podstawowe zestawy danych powiązane z publikacjami). Aktualnie otwieranie danych badawczych jest przedmiotem prac powołanego w kwietniu 2021 roku zespołu doradczego do spraw otwartych danych naukowych (zespół działa przy Ministerstwie Edukacji i Nauki) oraz jednym z elementów „Programu otwierania danych na lata 2021–2027”.

Więcej informacji na temat projektu Dziedzinowe Repozytoria Otwartych Danych Badawczych znaleźć można na stronie drodb.icm.edu.pl (<https://drodb.icm.edu.pl/>).

NATALIA GRUENPETER

Interdyscyplinarne Centrum Modelowania Matematycznego
i Komputerowego, Uniwersytet Warszawski

Strony repozytoriów:

<https://repod.icm.edu.pl/>, <https://rds.icm.edu.pl/>, <https://mxrdr.icm.edu.pl/>