

Zarządzanie danymi badawczymi i ich udostępnianie

W czasie, kiedy odpowiednie zarządzanie danymi badawczymi i ich udostępnianie przybiera na znaczeniu, konieczne jest tworzenie odpowiedniej infrastruktury i usług wspierających naukowców w procesie przygotowania, zabezpieczania i udostępniania wyników prac naukowo-badawczych.

Dane badawcze (Research Data) to wszystkie dane, które powstają w wyniku prac naukowo-badawczych, niezależnie od tego, czy mamy do czynienia z naukami ścisłymi, przyrodniczymi, społecznymi, czy humanistycznymi. Danymi mogą być m.in. wyniki analiz, pomiarów, ankiet, a także źródła historyczne, dzienniki laboratoryjne, obrazy i próbki. Mogą one przybierać formę zarówno cyfrową, jak i analogową.

Zarządzanie danymi badawczymi (Research Data Management) to temat, który jest dyskutowany i przybiera na znaczeniu od wielu lat. Podstawową przesłanką do odpowiedniego zarządzania danymi jest właściwe i efektywne wykorzystywanie publicznych funduszy. Instytucje i organizacje finansujące badania naukowe podkreślają, że fundusze przeznaczone na granty badawcze w dużej mierze się marnują. Dane, będące wynikami prac badawczych, nie są odpowiednio przechowywane ani udostępniane. Publikowane w czasopiśmie naukowych wyniki są w takim wypadku niemożliwe do zweryfikowania, a badania, które zostały zakończone wraz z projektem, niemożliwe do odtworzenia, kontynuowania i wykorzystania. Taka sytuacja powoduje np. to, że w wielu ośrodkach badawczych prowadzone są ponownie te same badania, które już zostały kiedyś przeprowadzone, nawet jeśli w przeszłości nie przyniosły oczekiwanych wyników.

Plan S – inicjatywa instytucji finansujących badania naukowe

Instytucje finansujące, w tym Komisja Europejska, kładą coraz większy nacisk na właściwe zarządzanie wynikami finansowanych przez siebie badań. Jedną z przełomowych inicjatyw w tym zakresie jest tzw. Plan S, ogłoszony w 2018 r. przez cOAlition S, czyli europejskie porozumienie instytucji finansujących badania naukowe. Zobowiązuje on grantobiorców do udostępniania wyników badań w sposób całkowicie otwarty już od stycznia 2021 r. Dotyczy to głównie publikacji naukowych, które powinny być publikowane w czasopiśmie Open Access lub udostępniane w otwartych repozytoriach natychmiast po opublikowaniu. Wytoczne zawarte w Planie S dotyczą zarówno wydawców, którzy muszą prawnie i technicznie dostosować czasopisma do ściśle określonych wymagań, jak i samych autorów, którzy podpisują zobowiązanie dotyczące otwartego publikowania w ramach umowy grantowej. Jedną z instytucji zrzeszonych w cOAlitionS jest Narodowe Centrum Nauki (NCN).

Polityka Open Access Narodowego Centrum Nauki

NCN jest członkiem Science Europe, inicjatora cOAlition S, organizacji zrzeszającej ponad 30 agencji finansujących naukę z 27 krajów europejskich, której nadrzędnym celem jest kształ-

towanie polityki finansowania nauki i wysokich standardów jej uprawiania. Do priorytetów organizacji należą działania zmierzające do otwartego dostępu do danych naukowych oraz publikacji, które są efektem realizacji badań finansowanych lub współfinansowanych ze środków publicznych. Polityka Open Access NCN została wprowadzona 31 maja 2020 r., a obowiązuje od konkursów ogłoszonych 15 czerwca 2020 r. NCN wspiera również otwarte udostępnianie danych badawczych. Od 2019 r. do wniosków grantowych, składanych w konkursach NCN należy załączyć Plan Zarządzania Danymi (Data Management Plan – DMP), który mówi m.in. o tym jakie dane badawcze będą wykorzystywane, wytwarzane i udostępniane w ramach projektu, a także jak będą one przechowywane.

Dyrektywa UE o udostępnianiu danych

4 kwietnia 2019 r. Parlament Europejski przyjął dyrektywę o otwartych danych i ponownym wykorzystaniu informacji sektora publicznego. Państwa członkowskie UE mają od tego czasu dwa lata na dostosowanie swojego prawa do nowych wymogów. Oznacza to m.in. konieczność wprowadzenia regulacji dotyczących udostępniania danych badawczych przez instytucje naukowe już w 2021 r.

Zasady FAIR Data a Open Data

Jednym z podstawowych kryteriów odpowiedniego zarządzania danymi badawczymi jest spełnianie przez nie tzw. zasad FAIR. Określają one szereg atrybutów, jakie musi posiadać opublikowany zbiór danych, aby był łatwy do odnalezienia, dostępny, interoperacyjny i nadający się do ponownego, wielokrotnego użytku. Jest to akronim słów: Findable, Accessible, Interoperable i Reusable.

Należy pamiętać, że dane powinny spełniać zasady FAIR, ale nie oznacza to jednocześnie, że wszystkie powinny być otwarte. Podstawową zasadą udostępniania danych badawczych jest zasada mówiąca, że powinny być one „na tyle otwarte, na ile to możliwe i na tyle zamknięte, na ile to konieczne”. Bowiem nie każde dane mogą, czy też powinny być udostępniane, nie wszystkie też nadają się do udostępniania. Wszystkie natomiast powinny być tworzone z uwzględnieniem zasad FAIR, co świadczy o właściwym ich generowaniu, przechowywaniu i wykorzystaniu.

Wsparcie instytucjonalne w zakresie zarządzania danymi – Centrum Kompetencji

W obliczu wszystkich tych zmian i wyzwań, przed którymi stoją polskie instytucje naukowe, konieczne jest zapewnienie odpowiedniego poziomu wiedzy i wsparcia dla pracowników

naukowych w zakresie zarządzania danymi i ich udostępniania. W dużych europejskich ośrodkach badawczych w ciągu ostatnich kilkunastu lat powstały wyspecjalizowane jednostki i tzw. centra kompetencji oraz wyodrębniła się rola tzw. data stewarda, osoby bezpośrednio odpowiedzialnej za wsparcie zespołów badawczych w procesie zarządzania danymi. W modelowej sytuacji każda dyscyplina, reprezentowana na uczelni powinna mieć wsparcie przynajmniej jednego data stewarda. Ich zadania i role powinna koordynować jednostka, zapewniająca zaplecze merytoryczne i infrastrukturę, np. biblioteka akademicka.

Centrum Kompetencji Otwartej Nauki w Bibliotece Politechniki Gdańskiej powstało w 2018 r. w ramach projektu MOST DANYCH – Multidyscyplinary Otwarty System Transferu Wiedzy – Etap II Open Research Data, finansowanego ze środków Unii Europejskiej w ramach Programu Operacyjnego Polska Cyfrowa. Centrum Kompetencji w Bibliotece Politechniki Gdańskiej jest pierwszą i jak na razie jedyną w Polsce tego typu innowacyjną jednostką, wspierającą środowisko naukowe we wszystkich aspektach Otwartej Nauki, a w szczególności w zakresie zarządzania danymi badawczymi. Pracownicy Centrum wspomagają naukowców zarówno w kwestiach otwartego publikowania i udostępniania prac badawczych w repozytorium MOST Wiedzy, jak i prowadzą szkolenia i konsultacje z zakresu m.in. tworzenia Planów Zarządzania Danymi na potrzeby wniosków grantowych.

W lutym 2020 r., po wprowadzeniu obowiązku przygotowania DMP przez wnioskodawców, NCN zorganizowało dwa szkolenia dla bibliotekarzy i pracowników administracji uczelni, wspierających naukowców w przygotowywaniu wniosków grantowych. Celem szkoleń było omówienie poszczególnych części DMP oraz idei otwartości w dostępie do danych. Ekspertami szkolącymi byli pracownicy Biblioteki PG.

W odpowiedzi na duże zapotrzebowanie na wiedzę w tym zakresie oraz brak odpowiednio wyszkolonych kadr w polskich instytucjach naukowych, Centrum Kompetencji zorganizowało serię szkoleń w Gdańsku a następnie online, zarówno dla bibliotekarzy i pracowników administracji, jak i dla naukowców, reprezentujących różne dyscypliny naukowe. Łącznie w pierwszym półroczu 2020 r. z zasad tworzenia i wsparcia w zakresie DMP przeszkolonych zostało ponad 300 osób z bibliotek i jednostek administracji wielu instytucji i uczelni z całej Polski.

Centrum Kompetencji w Bibliotece PG współpracuje z czołowymi międzynarodowymi gremiami i organizacjami, odp-

wiedzialnymi za promowanie zasad FAIR, idei otwartych danych, tworzenie standardów i opiniowanie aktów prawnych dotyczących polityki naukowej, Open Access i Open Research Data. Wśród nich są m.in. EOSC, SPARC Europe, GO-FAIR, Research Data Alliance, OCLC oraz CESAER.

Repozytorium danych badawczych

Dane badawcze przeznaczone do udostępniania deponowane są w tzw. repozytorium danych. Narzędzie to powinno spełniać szereg wymagań organizacyjnych i technicznych, aby dane spełniały zasady FAIR oraz mogły być indeksowane przez naukowe wyszukiwarki i bazy, takie jak np. Data Citation Index (Web of Science). Jest to o tyle ważne, że dane, udostępniane w ramach tzw. data setów (zbiorów danych) mogą być cytowane i mieć wpływ na ocenę dorobku naukowego autora. Inną formą udostępniania danych badawczych są tzw. data journals, w których dane są publikowane.

W ramach projektu MOST DANYCH Politechnika Gdańska, wraz z dwoma partnerami: Uniwersytetem Gdańskim i Gdańskim Uniwersytetem Medycznym, udostępnia dane badawcze z różnych dyscyplin naukowych w repozytorium MOST Wiedzy. Repozytorium to, tworzone jest w zgodzie z międzynarodowymi standardami, a jego funkcjonalność jest odpowiedzią na potrzeby naukowców. W zasobach repozytorium można znaleźć m.in. dane obrazowe, wyniki pomiarów i eksperymentów, dane medyczne, środowiskowe i biologiczne, a także z zakresu nauk społecznych, humanistycznych i nauk o sztuce. Przekrój dyscyplin, funkcjonalność, a także zastosowane rozwiązania techniczne czynią z MOST Wiedzy nie tylko w pełni funkcjonalne repozytorium danych naukowych, ale także platformę promocji dorobku, zarówno samych uczelni, jak i poszczególnych naukowców. Mimo iż platforma w części poświęconej otwartym danym została uruchomiona dopiero w marcu 2020 r., zdeponowano w niej już ponad 2000 data setów, z czego ponad połowa już została udostępniona. Tylko do końca przyszłego roku planowane jest udostępnienie ok 30 000 tysięcy data setów, w tym tysiące zdjęć tkanek i organizmów, które będą mogły być przeglądane w wysokiej rozdzielczości i jakości poprzez tworzony, również w ramach projektu, wirtualny mikroskop. Będzie to narzędzie przydatne nie tylko dla badaczy, wykorzystujących zgromadzone dane do celów naukowych, ale również dla entuzjastów nauki, hobbyistów i uczniów szkół na każdym poziomie nauczania.



Projekt „MOST DANYCH: Multidyscyplinary Otwarty System Transferu Wiedzy – etap II: Open Research Data” współfinansowany jest z Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Polska Cyfrowa na lata 2014-2020.