

PLAN ZARZĄDZANIA DANYMI BADAWCZYM

Ewa Adaszyńska

Biblioteka Uniwersytetu Zielonogórskiego

Dzielenie się danymi powstającymi podczas badań naukowych jest ważnym elementem komunikacji naukowej. To część otwartej nauki, na która wpływ mają nie tylko naukowcy i redakcje czasopism, ale także instytucje finansujące badania, które zobowiązanie otwartego dostępu do danych badawczych (*ang. Research Data*) zapisują jako element umów grantowych z beneficjentami. Otwarty dostęp do danych zapisany jest w europejskim programie Horyzont oraz programach Narodowego Centrum Nauki, m.in. w konkursach *MAESTRO*, *PRELUDIUM*, *SONATA BIS*, *OPUS*, *MINIATURA*, *GRIEG*. Przygotowanie i udostępnienie danych

badawczych jest procesem twórczym, wymaga przygotowania osobnego planu dla każdego projektu.

Plan zarządzania danymi

Plan zarządzania danymi (*ang. Data Management Plan, DMP*) to „dokument opisujący, co będzie się działo z danymi w trakcie projektu i po jego zakończeniu. Ma charakter żywego dokumentu, który może i powinien się zmieniać wraz ze zmianami pojawiającymi się w innych obszarach projektu badawczego”. Zgodnie z zasadami FAIR powinien opisywać, jakie działania zostaną podjęte, aby wytworzo-

15



FOT. MAMERT JANION

ne w ramach projektu dane były łatwe do odnalezienia, dostępne i możliwe do ponownego wykorzystania oraz powiązania z innymi danymi (zob. „[Selekcja i przygotowanie danych badawczych do udostępniania](#)”, s. 6,9).

Plan powinien powstać na etapie przygotowania projektu badawczego, musi być:

- _ opisany skrótowo (do 800 znaków);
- _ dołączony do formularza wniosku o finansowanie projektu;
- _ może się zmieniać, podobnie jak plan badawczy, z którym jest związany;
- _ nie wymaga konsultacji w trakcie zmian;
- _ skorygowany na etapie składania raportu końcowego;
- _ oceniany podwójnie; na etapie oceny wniosku i po zakończeniu projektu badawczego;
- _ oceniany według tego samego wzoru dla wszystkich konkursów.

Plany różnią się od siebie w zależności od rodzaju danych i dyscypliny naukowej, której dotyczą; niezależnie od tego powinny zawierać informacje wspólne dla wszystkich projektów konkursowych, w zakresie:

danych, ustaleń typu:

- _ jakie dane zostaną wytworzone lub zebrane podczas projektu? (typy i liczba danych, format plików);
- _ jak zostaną opisane i uporządkowane? (rodzaj metadanych, standardy, metodologia);
- _ jakie będą podstawy prawne oraz aspekty etyczne (kodeks NCN, prawa autorskie, własność intelektualna, dane niejawne);
- _ w jaki sposób dane będą udostępnione? (gdzie, jak, kiedy, komu?);
- _ jakie dane będą przechowywane długoterminowo (sposoby przechowywania danych oraz ich ochrona).
- _ **stosowanej licencji:**
- _ Creative Commons (zalecane CC0, CC-BY, CC-BY-SA);
- _ Open Data Commons (licencja stosowana wyłącznie do baz danych ujętych jako całości; nie obejmuje części składowych bazy).

Każdy z etapów planu niesie za sobą specyficzne wyzwania prawne, które należy dokładnie rozważyć (zob. „[Prawne aspekty otwierania danych badawczych - poradnik](#)”).

Do organizacji, które wymagają prowadzenia DMP, należą: Komisja Europejska (Horyzont); European Research Councils; National Science Foundation; Research Councils UK; Biotechnology and Biological Sciences Research Council Wellcome Trust; Narodowe Centrum Nauki.

Zgodnie z wytycznymi [planu zarządzania danymi w projektach Narodowego Centrum Nauki](#), podobnie jak w planach programu „Horyzont”, powinien zawierać:

- _ opis danych:
 - _ sposoby pozyskiwania i wytwarzania nowych danych oraz ponownego wykorzystania już istniejących; określenie rodzaju formatu i objętości plików danych;
- _ dokumentację i jakość danych:
 - _ określenie standardu metadanych, sposobu organizacji danych, wyboru metodologii, środków kontroli jakości;
- _ przechowywanie i tworzenie kopii zapasowych podczas badań:
 - _ bezpieczeństwo danych i metadanych; tworzenie kopii zapasowych, zabezpieczenie danych i ochrona danych wrażliwych podczas badań;
 - _ wymogi prawne, kodeks postępowania;
 - _ sposób zapewnienia zgodności z przepisami dotyczą-

**Plan zarządzania danymi
(ang. *Data Management Plan, DMP*) to „dokument opisujący, co będzie się działo z danymi w trakcie projektu i po jego zakończeniu. Ma charakter żywego dokumentu, który może i powinien się zmieniać wraz ze zmianami pojawiającymi się w innych obszarach projektu badawczego”.**

cymi danych osobowych oraz bezpieczeństwa przy ich przetwarzaniu; sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własności; aktualne przepisy;

- _ udostępnianie i długotrwałe przechowywanie danych:
 - _ sposób i termin udostępniania danych; ograniczenia w ich udostępnianiu; przyczyny embarga; sposoby selekcji danych, miejsce długoterminowego przechowywania, metody lub narzędzia programowe umożliwiające dostęp i korzystanie z danych, przypisanie identyfikatora DOI do każdego rodzaju zbiorów;
- _ zadania związane z zarządzaniem danymi:
 - _ wybór osoby zarządzającej; określenie środków przeznaczonych na zarządzanie zgodnie z zasadami FAIR.

Od 2019 r. Narodowe Centrum Nauki wprowadziło obowiązek dołączania planu zarządzania danymi badawczymi do formularza wniosku o finansowanie projektu (komunikat: „[Plany NCN w zakresie zarządzania danymi naukowymi](#)” z dnia 3.04.2019 r.)

Formularz przygotowany jest do skrótovej wersji opisu (zwiększonej teraz z 400 do 800 znaków). Plan może podlegać zmianom w trakcie realizacji projektu podobnie jak plan badawczy, z którym jest związany, bez konieczności konsultowania z Centrum. Musi on być uzupełniany na etapie składania raportu końcowego. Zgodnie z [Regulaminem dotyczącym oceny wniosków](#) podlega on ocenie po zakończeniu realizacji projektu (według tego samego wzoru dla wszystkich konkursów).

Przykładowe plany zarządzania danymi

- [DMPTool](#) - narzędzie do tworzenia planów zarządzania danymi odpowiadające wymaganiom amerykańskich instytucji finansujących projekty, udostępnia ok. 43.5 tys. planów z różnych dyscyplin naukowych z 268 instytucji.
- [DMPonline](#) elastyczne narzędzie do tworzenia spersonalizowanych planów zarządzania danymi badawczymi według wymagań instytucji finansujących projekty, dostarczane przez Digital Curation Center (DCC). Zawiera

Rekomendowany czas przechowywania danych wynosi 10 lat i wynika bezpośrednio z zasad kodeksu rzetelności naukowej oraz zapisów umowy grantowej, a także zastosowanej metodologii typowej dla określonej dziedziny wiedzy.

wskazówki, szablony i przykłady pomocne w pisaniu DMP, udostępnia 21 tys. planów z 203 instytucji z 89 krajów.

- [The Data Curation Center](#) - brytyjski serwis specjalizujący się w zarządzaniu danymi badawczymi. Zawiera gotowe plany zarządzania danymi, szablony, przewodniki, wytyczne, informacje o formatach i metadanych dla poszczególnych dyscyplin naukowych, a także listę kontrolną DMP ([Checklist for a Data Management Plan](#)), która pozwala określić, jakich informacji może brakować w przygotowywanym planie.

Przykładowy plan zarządzania danymi dla nauk społecznych

- [Doctoral Dissertation Research: An Agent-Based Model of Population Changes in a Vulnerable Coastal Environment](#)

Przykładowy plan zarządzania danymi dla projektu NCN [Plan zarządzania danymi](#) opracowany przez dr hab. Małgorzatę Jefimow z UMK w Toruniu - biologia (wersja skrócona z OSF)

Metadane

Dane badawcze powinny być odpowiednio opisane dla ich łatwego zidentyfikowania. Można to zrobić za pomocą metadanych, które służą do szczegółowego opisu danych.

Metadane to najogólniej dane o danych, można wyróżnić wśród nich trzy podstawowe typy:

- Metadane opisowe niezbędnych do identyfikacji zbioru danych; zawierają takie elementy, jak np. autor, tytuł, słowa kluczowe, streszczenie.
- Metadane strukturalne ułatwiające nawigację; opisują relacje i zależności pomiędzy poszczególnymi zbiorami oraz ich elementami.
- Metadane administracyjne pomocne w zarządzaniu zasobem; zawierają datę i sposób utworzenia danych, typ pliku, sposób dostępu, mogą dotyczyć też zarządzania prawami autorskimi oraz zawierać informacje potrzebne do archiwizacji i utrzymania zasobu.

Standardy metadanych dla różnych dziedzin wiedzy

Opis danych badawczych musi być zgodny z przyjętymi standardami dla poszczególnych dziedzin lub dyscyplin naukowych. W wyborze standardów metadanych pomogą wybrane instytucje, które informują, jakie standardy stosują poszczególne repozytoria:

- Digital Curation Centre [Standards](#) - strona z dyscyplinarnymi metadany pomocna przy poszukiwaniu standardów dla własnych danych badawczych; zawiera zestaw międzynarodowych standardów pomocnych przy planowaniu oraz zarządzaniu zasobem danych.
- [FAIRsharing](#) - wyszukiwarka standardów metadanych połączona z repozytoriami dziedzinowymi.
- [The Research Data Alliance \(RDA\) Metadata Standards](#) - katalog standardów metadanych RDA.

Dane powinny być:

- udostępniane w formacie zapewniającym powszechny dostęp i otwartość wyszukiwania;
 - udostępniane z opisem (metadany) ułatwiającymi identyfikację danych oraz całej bazy;
 - opracowane według określonej metodologii i standardów dla danej dyscypliny;
 - efektywnie zarządzane przez opiekuna lub zarządzającego, co zapewni ich trwałość i integralność;
 - bezpieczne podczas gromadzenia, przetwarzania i udostępniania danych z kopiami zapasowymi;
 - bezpiecznie przechowywane również podczas długotrwałego archiwizowania w gwarantowanych serwisach;
 - zgodne z przepisami prawa; chroniące właściciela praw autorskich;
 - ochronione w przypadku zbioru danych wrażliwych, poufnych, osobowych, chronionych prawem autorskim;
 - zdeponowane w wybranym repozytorium lub czasopiśmie zgodnie z przyjętymi w danej dyscyplinie standardami;
 - długotrwale archiwizowane i dostępne.
- Po zakończeniu projektu dane badawcze powinny być udostępniane w prestiżowych i zaufanych miejscach, tam, gdzie grupa docelowa będzie nimi zainteresowana:
- w czasopiśmie gromadzącym dane;
 - w specjalistycznych repozytoriach. (Zob. „[Jak korzystać z zasobów w repozytoriach danych](#)”)

Rekomendowany czas przechowywania danych wynosi 10 lat i wynika bezpośrednio z zasad kodeksu rzetelności naukowej oraz zapisów umowy grantowej, a także zastosowanej metodologii typowej dla określonej dziedziny wiedzy.

Aby dane cyfrowe były dostępne przez ten czas, muszą być aktywnie zarządzane. Taką gwarancję daje deponowanie zbiorów w bezpiecznych repozytoriach cyfrowych zarządzanych zgodnie z dobrymi praktykami archiwizacji cyfrowej i zasadami FAIR.

FAIR to wytyczne, które umożliwiają ponowne wykorzystanie danych, przeznaczone dla ludzi i maszyn przetwarzających dane według zasady:

- Findable - łatwe do odnalezienia;
- Accessible - dostępne dla wszystkich;
- Interoperable - interoperacyjne, umożliwiające połączenie z innymi danymi;
- Reusable - do ponownego (wielokrotnego) wykorzystania.

[Wytyczne w sprawie zarządzania danymi FAIR określa program Horyzont 2020](#)

Przed udostępnieniem należy sprawdzić, czy istnieją prawne ograniczenia dotyczące ponownego wykorzystania danych pochodzących od osób trzecich oraz wskazać licencję, na której będą udostępniane (zalecane są licencje otwarte). Wszystkie pozyskiwane i wytwarzane dane muszą mieć określonego właściciela praw autorskich i praw własności intelektualnej.

Wyniki badań mogą być składowane w repozytoriach instytucjonalnych liderów projektu lub w europejskim repozytorium ogólnego przeznaczenia Zenodo, a także w repozytoriach dziedzinowych, które można wyszukiwać w katalogu repozytoriów Open DOAR albo katalogu danych badawczych re3data.org. Naukowcy pracujący w Polsce mają możliwość przechowywania i udostępniania danych badawczych w polskim repozytorium RepOD. Należy dokładnie zapoznać się z warunkami korzystania z serwisu i sprawdzić, czy spełnia on nasze wymagania (zob. „[Jak korzystać z zasobów w repozytoriach danych](#)”).

Wyszukiwarki

- [re3data.org](#) (Registry of Research Data Repositories) to globalny rejestr repozytoriów danych badawczych ze wszystkich dyscyplin akademickich. Dostarcza informacji o repozytoriach do trwałego przechowywania i dostępu. Podaje tytuły serwisów, ich zakres tematyczny, podział według dziedzin i dyscyplin oraz informacje o ograniczeniach, dzięki czemu łatwiej znaleźć serwis do zarchiwizowania i udostępniania wyników badań. Rejestruje ponad 2 450 repozytoriów z całego świata, w tym ponad 1 200 z Europy, 8 z Polski. Rejestr uruchomiony w 2012 r. i finansowany przez German Research Foundation (DFG) jest usługą [DataCite](#) wymienianą przez wielu wydawców w ich politykach redakcyjnych jako najlepsze narzędzie do identyfikacji odpowiedniego repozytorium. Zalecane jest też w wytycznych Komisji Europejskiej w sprawie otwartego dostępu do publikacji i danych badawczych w programie „Horyzont”.
- [OpenDOAR](#) to światowy katalog repozytoriów otwartego dostępu i ich polityk, prowadzony przez University of Nottingham według usług SHERPA/RoMEO we współpracy z Lund University. Umożliwia identyfikację, przeglądanie i wyszukiwanie repozytoriów na podstawie szeregu funkcji, takich jak lokalizacja, oprogramowanie lub rodzaj przechowywanych materiałów. Wymaga otwartego dostępu do publikacji naukowych oraz kontroluje przesyłane materiały według własnego uznania. Rejestruje i umożliwia przeszukiwanie zasobów 5 313 repozytoriów z całego świata (2 113 z Europy i 118 z Polski).

Repozytoria danych badawczych

- [Zenodo](#) ogólnodostępne repozytorium opracowane w ramach europejskiego programu [OpenAIRE](#) na zlecenie Komisji Europejskiej, prowadzone i finansowane przez CERN. Umożliwia naukowcom z każdego obszaru tematycznego deponowanie zbiorów danych, oprogramowania badawczego, raportów i innych cyfrowych materiałów związanych z badaniami. Publikacje oznaczane są identyfikatorem DOI.
- [RepOD](#) Repozytorium Otwartych Danych opracowane przez ICM UW w ramach działań Platformy Otwartej Nauki archiwizujące i udostępniające dane wytworzone, zebrane i opracowane na potrzeby badań naukowych. Przeznaczone dla tzw. małych danych.

Źródła:

- [Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon](#) / Przewodnik KE dotyczący zasad otwartego dostępu do publikacji naukowych i danych badawczych w programie „Horyzont 2020”.
- [Guidelines on Data Management in Horizon 2020](#) / Przewodnik KE dotyczący zarządzania danymi badawczymi w programie „Horyzont 2020”.
- [Guidelines on FAIR DATA Management in Horizon 2020](#) / Przewodnik KE dotyczący zarządzania danymi zgodnie z zasadami FAIR w programie „Horyzont 2020”.
- [Pilotaż otwartego publikowania danych badawczych w programie Horyzont 2020](#).
- [H2020 On-line Manual Open access & Data management](#) / Podręcznik on-line - otwarty dostęp i zarządzanie danymi
- [H2020 Online-Manual Data Management](#) / Podręcznik on-line - zarządzanie danymi.
- [Dyrektor NCN o otwartości w nauce](#).
- Strona Programu [Horyzont 2020](#).
- Strona [Narodowego Centrum Nauki](#).
- [Wytyczne dla wnioskodawców do uzupełnienia Planu zarządzania danymi w projektach badawczych Narodowego Centrum Nauki](#).
- [re3data.org](#).
- [OpenDOAR](#).
- [Zenodo](#).
- [RepOD](#).
- Opracowania w ramach projektu Dziedzinowe Repozytoria Otwartych Danych Badawczych:
 - „[Jak korzystać z zasobów w repozytoriach danych](#)”, oprac. N. Gruenpeter, Warszawa, 2019.
 - „[Prawne aspekty otwierania danych badawczych - poradnik](#)”, oprac. K. Siewicz, N. Rycko, Warszawa, 2019.
 - „[Selekcja i przygotowanie danych badawczych do udostępniania](#)”, oprac. Wojciech Fenrich, Warszawa, 2019.
- Opracowania Science Europe:
 - „[Practical Guide to the International Alignment of Research Data Management](#)” Praktyczny przewodnik po międzynarodowym dostosowaniu zarządzania danymi badawczymi, opracowany przez ekspertów z Science Europe w celu dostosowania wymagań zarządzania danymi badawczymi w różnych organizacjach. Polecany przez NCN, jego zapisy zostały ujęte w nowym formularzu we wniosku grantowym projektów NCN tzw. Planie Zarządzania Danymi, wprowadzony do 33. edycji konkursów, ogłoszonych 15 marca 2019 r.
 - „[Guidance Document Presenting a Framework for Discipline-specific Research Data Management](#)” Wytyczne przedstawiające ramy zarządzania danymi badawczymi specyficznymi dla dyscypliny, opracowane przez ekspertów z Science Europe; do wykorzystania jako znormalizowane szablony przez badaczy, organizacje naukowe i podmioty finansujące.
 - „[Implementing Research Data Management Policies Across Europe. Experiences from Science Europe Member Organisations, Science Europe 2020](#)” Wdrażanie polityk zarządzania danymi badawczymi w całej Europie: doświadczenia organizacji członkowskich Science Europe. Publikacja dla organizacji, która chce opracować wymagania dotyczące planów zarządzania danymi.