

# DANE BADAWCZE

Ewa Adaszyńska

Biblioteka Uniwersytetu Zielonogórskiego

**Dane badawcze** (ang. *open research data*) oznaczają dokumenty w formie cyfrowej inne niż publikacje naukowe, które są gromadzone lub produkowane w ramach działalności badawczo-naukowej i są wykorzystywane jako dowody w procesie badawczym, bądź też powszechnie akceptowane w środowisku naukowym jako konieczne do weryfikacji poprawności ustaleń i wyników badań. ([Dyrektywa Parlamentu Europejskiego i Rady \(UE\) z 2019/1024 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego](#) (art., 2, p. 9.)

**Dane badawcze** to różnorodne materiały gromadzone w ramach prowadzonych badań naukowych, definiowane jako:

- \_ zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych;
- \_ dane zebrane, zaobserwowane lub wytworzone jako materiał do analizy w celu uzyskania oryginalnych wyników naukowych;
- \_ wszystko, co zostało wyprodukowane lub wytworzone w ramach prowadzonych badań.

Są to więc w zależności od uprawianej dyscypliny m.in. dane liczbowe, dokumenty tekstowe, kwestionariusze, wyniki ankiet, fotografie, nagrania audio i wideo, zawartość baz danych, oprogramowanie (skrypty, pliki), obserwacje laboratoryjne, wyniki eksperymentów, pomiarów, obserwacji prowadzonych w terenie, modele matematyczne, algorytmy, dane statystyczne, specyfikacje, inne obiekty cyfrowe. ([Prawne aspekty otwierania danych badawczych - poradnik](#), s. 5)

Do danych badawczych powinno się zaliczać także dokumentację badań (informacje na temat pochodzenia danych, jak, gdzie, kiedy zostały zebrane i za pomocą jakich narzędzi) oraz metadane (opis danych umożliwiający ich identyfikację i zrozumienie).

Dane badawcze można dzielić na:

- \_ dane surowe, czyli takie, które uzyskano bezpośrednio w wyniku zastosowania narzędzia badawczego, w różnych przedsięwzięciach naukowych lub też zgromadzone na potrzeby konkretnych projektów, nieprzeanalizowane;
- \_ dane, które poddane zostały obróbce.

Dane badawcze mogą być archiwizowane w modelu otwartym lub zamkniętym.

**Otwarte dane badawcze** to dane, do których każdy użytkownik ma swobodny i bezpłatny dostęp, które może

wykorzystywać i modyfikować zgodnie z licencją wskazaną przez autorów.

Otwarte dane badawcze to dane cyfrowe w otwartym formacie, udostępniane w Internecie bezpłatnie do swobodnego korzystania wraz z prawem do ponownego wykorzystania i udostępniania przez wszystkich do dowolnego celu. Dane mogą być dowolnie używane, rozpowszechniane oraz przetwarzane przez kogokolwiek i gdziekolwiek w dowolnym celu, jeżeli nie spowodują naruszenia tajemnic ani interesów związanych z komercjalizacją. (Dyrektywa PE, 16).

Niektóre dane mogą być archiwizowane w modelu **zamkniętym** ze względu na kwestie związane z prawami własności intelektualnej, ochroną danych osobowych i poufnością, komercjalizacją badań, bezpieczeństwem i uzasadnionymi interesami handlowymi, ale zgodnie z zasadą „otwarty jak to najbardziej możliwe, zamknięty jak to konieczne (Dyrektywa PE, art. 10.1).

Dane badawcze udostępniane są najczęściej w postaci tzw. datasetów, czyli zbiorów danych stanowiących pewną całość powiązaną z eksperymentem, publikacją lub projektem naukowym.

**Otwarte dane badawcze to dane cyfrowe w otwartym formacie, udostępniane w Internecie bezpłatnie do swobodnego korzystania wraz z prawem do ponownego wykorzystania i udostępniania przez wszystkich do dowolnego celu.**

**Data set** to uporządkowany zbiór danych, udostępniony w określonym repozytorium, odnoszący się do danego tematu, opatrzony metadanymi opisującymi jego zawartość. Na data set składają się:

- \_ pliki z danymi;
- \_ plik typu „readme” opisujący zawartość danych (opis metodologii badań oraz tego, czy i w jaki sposób dane zostały przetworzone).

Zbiory te muszą być zgodne z zasadami, które ułatwiają dostęp do danych i ich ponowne wykorzystanie, zarówno przez użytkowników, jak i oprogramowanie komputerowe przeszukujące bazy danych bez udziału człowieka. Są to tzw. zasady FAIR zalecane przez Komisję Europejską.

**FAIR data** to zasady przygotowywania, opracowywania i udostępniania danych badawczych.

([Selekcja i przygotowanie danych badawczych do udostępniania](#), s.11-18).

Nazwa **FAIR** pochodzi od akronimu słów Findable, Accessible, Interoperable, Reusable, określa wymogi jakie powinny spełniać udostępnione dane badawcze. FAIR oznacza dane:

- \_ **łatwe do znalezienia** (opatrzone metadanymi, które są zgodne z przyjętymi standardami dla danej dyscypliny lub rodzaju danych, zaopatrzone w trwałe identyfikatory DOI, indeksowane w ogólnodostępnych serwisach, które można przeszukiwać);
- \_ **dostępne** (bezpośredni dostęp do zbioru danych i metadanych powinien być możliwy poprzez unikalny identyfikator i otwarty protokół komunikacyjny bez dodatkowych narzędzi oraz oprogramowania);
- \_ **interoperacyjne** (udostępnione w otwartym oprogramowaniu, które umożliwi przetwarzanie, wymianę, łączenie i powiązanie danych z innymi badaniami, systemów komputerowych lub oprogramowania; dane i metadane muszą być dostarczone w formacie zapewniającym łatwy odczyt i przetwarzanie, zawierać odnośniki do powiązanych zbiorów);

## Zarządzanie danymi badawczymi (RDM) to sposób organizacji pracy z danymi badawczymi w czasie trwania projektu oraz po jego zakończeniu

\_ **możliwe do ponownego wykorzystania** (opatrzone licencją określającą warunki ponownego wykorzystania i przetwarzania danych, a także informacją o kolejnych etapach przetwarzania; metadane muszą opisywać zbiór danych według standardów przyjętych dla dyscypliny lub rodzaju danych, określać autora i miejsce powstania, ma to ułatwić użytkownikom określenie ich przydatności do własnych badań).

Dane muszą być wytworzone i zapisane, przechowywane, dostępne, łatwe do znalezienia, zrozumiałe, poufne, interoperacyjne, odtwarzalne. ([Wytyczne w sprawie zarządzania danymi FAIR określa program Horyzont 2020](#) ).

Dostęp do danych i ich udostępnianie musi się opierać na spójnych zasadach i praktykach, opisanych jako zarządzanie danymi badawczymi (ang. *Research Data Management*, RDM) oraz na planie zarządzania danymi (ang. *Data Management Plan*, DMP).

**Zarządzanie danymi badawczymi (RDM)** to sposób organizacji pracy z danymi badawczymi w czasie trwania projektu oraz po jego zakończeniu, obejmuje następujące etapy:

- \_ **wytwarzanie lub pozyskiwanie danych** - określenie pochodzenia danych, opis procedur określających warunki jakie muszą spełniać dane aby można z nich korzystać, opis metodologii oraz otrzymanych rezultatów (metadane danych);
- \_ **przechowywanie i archiwizację danych** - określenie, gdzie będą przechowywane dane, kto będzie miał do nich dostęp, kto będzie za nie odpowiedzialny, ile będzie kopii;
- \_ **analizę i wykorzystanie danych** - wskazanie, które dane zostaną poddane analizie, określenie zagadnień związanych z ponownym przetwarzaniem danych i ich interpretacją, sprawdzenie polityki grantodawcy w zakresie udostępniania danych badawczych;
- \_ **udostępnianie danych** - wybór odpowiedniego repozytorium, licencji, osób do kontaktu.

**Plan zarządzania danymi badawczymi (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”. ([Selekcja i przygotowanie danych badawczych do udostępniania](#) s.6).

DMP jest dokumentem wymaganym przy składaniu wniosków o granty w ramach programów Horyzont (obecnie Horyzont Europa) oraz w konkursach Narodowego Centrum Nauki m.in. w konkursach MAESTRO, PRELUDIUM, SONATA BIS, OPUS, MINIATURA, GRIEG. Przygotowanie oraz udostępnienie danych badawczych jest procesem twórczym i wymaga osobnego omówienia.

Ujednolicenie europejskich praktyk związanych z zarządzaniem i udostępnianiem danych badawczych dla beneficjentów europejskich agencji finansujących badania naukowe proponuje Science Europe w publikacjach:

- \_ [Practical Guide to the International Alignment of Research Data Management](#)

Praktyczny przewodnik po międzynarodowym dostosowaniu zarządzania danymi badawczymi, opracowany przez ekspertów z organizacji członkowskich 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. Został on 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 organizacji członkowskich 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.

Opracowania w ramach projektu Dziedziny Repozytoria Otwartej Danych Badawczych:

– [Prawne aspekty otwierania danych badawczych - poradnik](#), oprac. K. Siewicz, N. Rycko, Warszawa, 2019, dostęp: <https://drodb.icm.edu.pl/wp-content/uploads/2020/01/Prawne-aspekty-otwierania-danych-badawczych-%E2%80%93-poradnik.pdf>

– [Selekcja i przygotowanie danych badawczych do udostępniania](#), oprac. W. Fenrich, Warszawa, 2019, dostęp: <https://drodb.icm.edu.pl/wp-content/uploads/2019/10/Selekcja-i-przygotowanie-danych-badawczych-do-udost%C4%99pnienia>

– [Otwarta nauka w Polsce 2014. Diagnoza](#), pod red. J. Szprota, Warszawa, 2014, dostęp: [https://bg.uwb.edu.pl/DebataOA2016/pdf/Otwarta%20nauka%20w%20Polsce%202014\\_Diagnoza.pdf](https://bg.uwb.edu.pl/DebataOA2016/pdf/Otwarta%20nauka%20w%20Polsce%202014_Diagnoza.pdf)

– [Jak korzystać z zasobów w repozytoriach danych](#), oprac. N. Gruenpeter, Warszawa, 2019, dostęp: <https://drodb.icm.edu.pl/wp-content/uploads/2019/10/Jak-korzyst%C4%87-z-zasob%C3%B3w-w-repozytoriach-danych.pdf>

**Zródła:**

– [Dyrektywa Parlamentu Europejskiego i Rady \(UE\) z 2019/1024 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego](#), dostęp: <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:32019L1024&from=EN>



FOT. MAMERT JANION