Platforma Otwartej Nauki
ICM, Uniwersytet Warszawski
Wahaszalwyski
Marta Hoffmann-
Somme

Otwarte udostępnianie danych badawczych

UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
www.icm.edu.pl

OpenAIRE
Open Access Infrastructure for Research in Europe

PLATFORMA OTWARTEJ NAUKI

Co to są dane badawcze?

„…zarejestrowane materiały o charakterze faktograficznym powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.”

„Dane badawcze to dane zebrane, zaobserwowane lub wytworzone jako materiał do analizy, w celu uzyskania oryginalnych wyników naukowych.”

icm

PLATFORMA OTWARTEJ NAUKI

Co to jest otwarty dostęp?

Komisja Europejska (2013):

„Otwarty dostęp definiujemy jako praktykę udostępniania informacji naukowej *on-line* w taki sposób, by była dla użytkownika końcowego **bezpłatna** i by możliwe było jej **ponowne wykorzystanie**.”

Guidelines on *Open Access* to *Scientific Publications* and *Research Data* in *Horizon 2020*. Wersja: 16 December 2013.

Po co otwierać dane?

1. Możliwość weryfikacji naszych wyników – powtarzalność w nauce *(reproducibility)*
2. Możliwość ponownego wykorzystania – przez nas i przez innych, również komercyjnie (*re-use*)

Jakie wymagania dotyczące udostępniania danych są nakładane na naukowców?

- Wymagania nakładane przez **wydawców naukowych**: konieczność udostępniania danych.
- Wymagania nakładane **w umowach grantowych**: konieczność tworzenia Planów Zarządzania Danymi i/lub udostępniania danych.
- Wymagania **pracodawców** (uczelni): prawidłowe zarządzanie danymi, tworzenie Planów Zarządzania Danymi i udostępnianie.

Wymagania Komisji Europejskiej w programie Horyzont 2020

Pilotaż Otwartych Danych w H2020

– od stycznia 2017 rozszerzony do programu

Open Research Data by Default

„Pilotaż Otwartych Danych obejmuje dwa rodzaje danych:

1) **dane** (…) **niezbędne do weryfikacji wyników** prezentowanych w publikacjach naukowych należy udostępniać tak szybko, jak to możliwe;

2) **inne dane** (…) wymienione w planie zarządzania danymi należy udostępniać zgodnie z ustalonymi w planie terminami.

(…) Projekty objęte pilotażem są zobowiązane do deponowania opisanych powyżej danych badawczych, najlepiej w repozytoriach danych badawczych.”

„Na ile to możliwe, projekty są zobowiązane do podjęcia działań umożliwiających osobom trzecim **dostęp** do danych badawczych, ich **analizę maszynową**, **ponowne wykorzystanie**, **kopiowanie** i **rozpowszechnianie** (bez opłat ze strony użytkowników).

Prostą i skuteczną metodą osiągnięcia powyższego celu jest dołączenie do deponowanych danych licencji Creative Commons (CC-BY lub oświadczenia CC0).”

„Od finansowanych projektów […] jest wymagane korzystanie ze szczegółowego planu zarządzania danymi, odnoszącego się do poszczególnych zbiorów danych.”

⟶ *ze złożenia planu zarządzania danymi nie można się wyłączyć!*

Sytuacja w Polsce

Dokument MNiSW (październik 2015):

Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce

„…zaleca, aby krajowe podmioty finansujące badania naukowe ze środków publicznych (…) stosowały i upowszechniały zasady, zgodnie z którymi publikacje i **dane badawcze** powstające w wyniku finansowanych lub współfinansowanych przez nie badań **znajdą się w otwartym dostępie**.”

icm

PLATFORMA OTWARTEJ NAUKI

Polska Akademia Nauk — *„Kodeks etyki pracownika naukowego”*

3.1. PRAKTYKI DOTYCZĄCE POSTĘPOWANIA Z DANYMI NAUKOWYMI

„Wszystkie oryginalne **dane źródłowe** (…) powinny być skrupulatnie **udokumentowane** i bezpiecznie **zarchiwizowane** w sposób (…) zapewniający po opublikowaniu tych badań ich **dostępność** przez okres właściwy dla danej dyscypliny.”

→ dostępność w tym rozumieniu nie oznacza koniecznie otwartości

• Narodowe Centrum Nauki — „*Kodeks Narodowego Centrum Nauki dotyczący rzetelności badań naukowych i starania o fundusze na badania*” (2016):

- obowiązek posiadania planu zarządzania danymi i udostępnienia go na żądanie NCN: plan przechowywania i dostępności danych

• Narodowe Centrum Badań i Rozwoju — „*Warunki ogólne realizacji projektu*” (2016):

- obowiązek stosowania „*Kodeksu etyki pracownika naukowego*” PAN

Udostępnianie danych

– jak się za to zabrać?

FAIR data

Findable	- żeby było łatwo je znaleźć
Accessible	- żeby były dostępne dla wszystkich
Interoperable	- żeby można było je połączyć z innymi danymi
Reusable	- żeby dało się je ponownie wykorzystać

Najważniejsze kwestie:

- Selekcja danych – co jest wartościowe?
- Przygotowanie danych – pełna dokumentacja, metadane, formaty
- Archiwizacja danych
- Stan prawny danych

Plan zarządzania danymi — DMP (*data management plan)*

1. Jakie dane zostaną wytworzone lub zebrane?

   (co będą zawierać? jakie będą formaty plików? jak dużo będzie danych?)

2. Jak zostaną uporządkowane i opisane?

   (metadane, dokumentacja)

3. Kwestie etyczne i prawne

   (kwestie związane z ochroną prywatności, dane niejawne, etc.)

4. W jaki sposób dane zostaną udostępnione?

   (jak, kiedy, komu)

5. Które dane będą przechowywane długoterminowo? Gdzie, jak długo?

Slajd na podstawie materiałów DCC:
www.dcc.ac.uk/resources/data-management-plans/checklist

Archiwizacja: przechowywanie długoterminowe

1. Długofalowe bezpieczeństwo danych – repozytorium/archiwum godne zaufania
2. Widoczność – znane wśród badaczy, dobrze widoczne w wyszukiwarkach
3. Trwała lokalizacja – stały identyfikator cyfrowy (np. DOI – *digital object identifier*)

Repozytoria danych badawczych

Specjalistyczne:

- Protein Data Bank
- Genbank
- Oxford Text Archive
- Polish Social Data Archive

Ogólne:

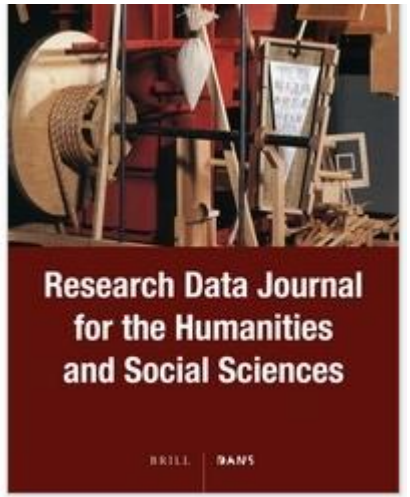
- Dryad (biologiczne)
- Zenodo (europejskie)
- RepOD (polskie)
- University of Cambridge Data Repository (instytucjonalne)

Czasopisma publikujące dane (*data journals*)

SCIENTIFIC DATA

npg nature publishing group

ELSEVIER Data in Brief



- Artykuły opisujące dane (*data descriptors*)
- Dane są deponowane w repozytoriach
- Niektóre czasopisma dopuszczają też możliwość dołączania danych w postaci Supplementary Material

→ Uzupełnienie systemu repozytoryjnego, nie alternatywa

icm

PLATFORMA OTWARTEJ NAUKI

Kwestie prawne

1. Komu przysługują prawa do konkretnego zbioru danych?

   Kto ma prawo podjąć decyzję o jego udostępnieniu?
2. Co wolno użytkownikowi danych? Kto i w jaki sposób o tym decyduje?

Ochrona prawna danych badawczych

1. Prawa autorskie – zależy od rodzaju danych – jeżeli zbiór danych ma cechy utworu.
2. Prawa ochrony baz danych – zależy od rodzaju danych – jeżeli zbiór danych ma cechy bazy danych.
3. Prawa osób trzecich – zależy od rodzaju danych – np. jeżeli zbiór zawiera dane osobowe pacjentów, mają oni prawo do ochrony tych danych.

Nie wszystkie prawa do danych przysługują autorom, w grę wchodzą jeszcze pracodawcy (instytucje naukowe) i osoby trzecie, np. poddawane badaniom.

- Jednostki naukowe regulują zasady wewnętrzne w **regulaminach nabywania i korzystania z praw własności intelektualnej**.
- Zazwyczaj głównym celem regulaminu jest ustalenie kwestii związanych z komercjalizacją; zasady otwartego udostępniania wyników badań nie są wprost omawiane.

Czasem bardzo trudno jest badaczowi ustalić, jakiej ochronie podlega konkretny zbiór danych.

→ instytucja naukowa powinna zapewnić mu odpowiednie wsparcie.

Prawny status udostępnionych danych

Dane w repozytorium możemy udostępnić:

- bez licencji: na zasadach dozwolonego użytku
- z licencją (np. na otwartej licencji Creative Commons)
- z oświadczeniem o zrzeczeniu praw (np. Creative Commons Zero)

Pilotaż Otwartych Danych w H2020

„Na ile to możliwe, projekty są zobowiązane do podjęcia działań umożliwiających osobom trzecim dostęp do danych badawczych, ich analizę maszynową, ponowne wykorzystanie, kopiowanie i rozpowszechnianie (bez opłat ze strony użytkowników).

Prostą i skuteczną metodą osiągnięcia powyższego celu jest dołączenie do deponowanych danych licencji Creative Commons (CC-BY lub CC0).”

?

FOSTER

PLATFORMA OTWARTEJ NAUKI

As open as possible, as closed as necessary

– ograniczenia prawne i etyczne oraz spójność z głównym celem prowadzonych badań (np. komercjalizacja)

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf - Guidelines on FAIR Data Management, wersja 3.0, 26 July 2016

Dane osobowe

- **Anonimizacja danych** – żeby niemożliwe było zidentyfikowanie uczestników badania
- lub: **Zgoda osób badanych** na udostępnienie – bez nadania licencji!
- Być może musimy wprowadzić ograniczenia dostępu – trzeba sprawdzić, czy wybrane przez nas repozytorium jest w stanie je zaimplementować

Czy wszystkie dane powinny być otwarte? Nie.

Dane osobowe

Bezpieczeństwo narodowe

Komercjalizacja wyników badań

Ale **informacja o istnieniu** danych zawsze powinna być publicznie dostępna:

- Inni mogą się dowiedzieć o danych i negocjować z nami dostęp
- Pozwala to uniknąć duplikacji badań

Dziękuję za uwagę

Kontakt:

msommer@icm.edu.pl

UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
www.icm.edu.pl

OpenAIRE
Open Access Infrastructure for Research in Europe

PLATFORMA OTWARTEJ NAUKI