

Plan zarządzania danymi: rekomendacje Komisji Europejskiej

Opracowanie na podstawie: Horizon Europe, Data Management Plan Template (wersja 1.0), dokument dostępny na stronie:

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx [data dostępu: 17.07.2023].

Tłumaczenie przygotowane w ramach kursu z zarządzania danymi badawczymi udostępnionego na platformie edukacyjnej Navoica (<https://navoica.pl/>).

1. Opis danych

Czy wykorzystane zostaną istniejące dane, a jeżeli tak – w jaki sposób? Należy podać powody odrzucenia ponownego wykorzystania istniejących danych, jeżeli takie wykorzystanie rozważano, ale z niego zrezygnowano.

Jakie typy i formaty danych będą wytwarzane lub ponownie wykorzystywane w ramach projektu?

Jaki jest cel wytwarzania lub ponownego wykorzystywania danych i jaki jest ich związek z celami projektu?

Jaki jest szacowany rozmiar danych, które będą w ramach projektu wytworzone lub ponownie wykorzystane?

Jakie jest źródło/pochodzenie danych, zarówno wytworzonych, jak i ponownie wykorzystanych?

Kto może wykorzystać dane („użyteczność danych”) poza projektem?

2. Zasady FAIR

2.1. Dane możliwe do znalezienia (findable), uwzględnienie metadanych

Czy dane będą identyfikowane za pomocą trwałego identyfikatora?

Czy zapewnione zostaną bogate metadane umożliwiające znalezienie danych? Jakie metadane zostaną utworzone? Jakie standardy dyscyplinowe lub ogólne będą przestrzegane? Jeśli w danej dyscyplinie nie istnieją standardy dotyczące metadanych, należy opisać, jaki typ metadanych zostanie utworzony i w jaki sposób.

Czy w metadanych zostaną podane słowa kluczowe, aby zoptymalizować możliwość znalezienia, a następnie potencjalnego ponownego wykorzystania?

Czy metadane będą przygotowane w taki sposób, aby można je było harvestować i indeksować?

2.2. Dane dostępne (accessible)

Repozytorium:

Czy dane zostaną zdeponowane w zaufanym repozytorium?

Czy rozpoznane zostały warunki i funkcje repozytorium, w którym będą przechowywane dane?

Czy w repozytorium nadaje się danym identyfikatory? Czy identyfikator w repozytorium prowadzi do obiektu cyfrowego?

Dane:

Czy wszystkie dane będą udostępnione w sposób otwarty? Jeśli określonych zbiorów danych nie można udostępnić (lub trzeba je udostępnić na warunkach ograniczonego dostępu), wyjaśnić należy dlaczego, wyraźnie oddzielając przyczyny prawne i umowne od ograniczeń o charakterze uznaniowym. Należy zauważyć, że w projektach realizowanych przez wielu beneficjentów poszczególni beneficjenci mogą również nie udostępniać danych, jeśli jest to sprzeczne z ich uzasadnionymi interesami lub innymi ograniczeniami, zgodnie z umową grantową.

Jeżeli w celu zapewnienia czasu na publikację lub zapewnienia ochrony własności intelektualnej (np. patentów) stosowane jest embargo, należy określić, dlaczego i jak długo będzie ono obowiązywało, pamiętając, że dane badawcze powinny być udostępniane tak szybko, jak to możliwe.

Czy dane będą dostępne za pośrednictwem bezpłatnego i standardowego protokołu dostępu?

Jeśli istnieją ograniczenia w korzystaniu z danych, w jaki sposób zapewniony zostanie dostęp do nich, zarówno w trakcie, jak i po zakończeniu projektu?

W jaki sposób zostanie ustalona tożsamość osoby uzyskującej dostęp do danych?

Czy istnieje potrzeba powołania komitetu ds. dostępu do danych (np. w celu oceny/zatwierdzenia wniosków o dostęp do danych osobowych/wrażliwych)?

Metadane:

Czy metadane będą udostępniane publicznie i przekazane do domeny publicznej na podstawie oświadczenia CC0, zgodnie z umową grantową? Jeśli nie, należy wyjaśnić dlaczego. Czy metadane będą zawierały informacje umożliwiające użytkownikowi uzyskanie dostępu do danych?

Jak długo dane będą dostępne i możliwe do znalezienia? Czy metadane będą dostępne, gdy dane przestaną już być dostępne?

Czy do uzyskania dostępu do danych lub ich odczytu będzie potrzebna dokumentacja bądź odniesienie do jakiegokolwiek oprogramowania? Czy możliwe będzie udostępnienie odpowiedniego oprogramowania (np. open source)?

2.3. Dane interoperacyjne (interoperable)

Jakie słowniki, standardy, formaty lub metodologie w odniesieniu do danych lub metadanych będą używane, aby dane były interoperacyjne i aby umożliwić wymianę i ponowne wykorzystanie danych w obrębie i między dyscyplinami? Czy uwzględnione zostaną najlepsze praktyki w zakresie interoperacyjności, wspierane przez społeczność naukową w ramach dyscypliny? Które?

W wypadku, gdy nie da się uniknąć korzystania z nietypowych lub specyficznych dla projektu ontologii lub słowników, czy przeprowadzone zostanie mapowanie do częściej używanych ontologii? Czy utworzone ontologie lub słowniki będą udostępnione w sposób otwarty, aby umożliwić ich ponowne wykorzystanie, udoskonalanie lub rozszerzanie?

Czy dane będą zawierały ściśle odniesienia¹ do innych danych (np. innych danych z twojego projektu lub zbiorów danych z poprzednich badań)?

2.4. Zwiększenie potencjału ponownego wykorzystania danych (reusable)

W jaki sposób zapewniona zostanie dokumentacja potrzebna do walidacji analizy danych i ułatwiająca ponowne wykorzystanie danych (np. pliki readme z informacjami o metodologii, książkach kodowych, czyszczeniu danych, analizach, definicjach zmiennych, jednostkach miary itp.)?

Czy dane zostaną udostępnione z przeniesieniem do domeny publicznej, aby umożliwić jak najszersze ponowne wykorzystanie? Czy dane będą udostępnione na standardowych licencjach pozwalających na ich ponowne wykorzystanie, zgodnie z obowiązkami określonymi w umowie grantowej?

Czy dane wytworzone w projekcie będą mogły być wykorzystane przez osoby trzecie, w szczególności po zakończeniu projektu?

Czy pochodzenie danych będzie dokładnie udokumentowane przy użyciu odpowiednich standardów?

Należy opisać wszystkie właściwe dla projektu procesy zapewniania jakości danych.

Zgodnie z zasadami FAIR plany zarządzania danymi powinny również odnosić się do wyników badań innych niż dane i wiązać się ze starannym rozważeniem kwestii związanych z alokacją zasobów, bezpieczeństwem danych i aspektami etycznymi.

3. Inne rezultaty badań

Oprócz zarządzania danymi beneficjenci powinni również rozważyć i zaplanować zarządzanie innymi rezultatami badań, które mogą zostać wytworzone lub ponownie wykorzystane w ramach ich projektów. Takie wyniki mogą być cyfrowe (np. oprogramowanie, procedury, protokoły, modele itp.) lub fizyczne (np. nowe materiały, przeciwciała, odczynniki, próbki itp.).

Należy rozważyć, które z powyższych pytań dotyczących zasad FAIR mogą odnosić się do zarządzania innymi rezultatami badań, a także starać się przedstawić wystarczająco szczegółowo, jaki będzie sposób zarządzania nimi i ich udostępniania, zgodnie z zasadami FAIR.

4. Zasoby

Jakie będą koszty opracowania danych lub innych wyników badań zgodnie z zasadami FAIR w projekcie (np. bezpośrednie i pośrednie koszty związane z przechowywaniem, archiwizacją, ponownym wykorzystaniem, bezpieczeństwem itp.)?

¹ Ściśle odniesienia to odniesienia wyjaśniające ich przeznaczenie. Na przykład: „X jest regulatorem Y” jest o wiele bardziej precyzyjnym odniesieniem niż „X jest powiązany z Y” lub „X patrz także Y”. Celem jest zatem stworzenie jak największej liczby znaczących powiązań między zasobami (meta)danych w celu wzbogacenia kontekstowej wiedzy o danych. Por. <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>.

W jaki sposób zostaną one pokryte? Należy pamiętać, że koszty związane z zarządzaniem danymi badawczymi/rezultatami są kwalifikowalne w ramach grantu Horyzont Europa (jeśli są zgodne z warunkami umowy grantowej).

Kto będzie odpowiedzialny za zarządzanie danymi w projekcie?

W jaki sposób zapewnione zostanie długoterminowe przechowywanie danych? Należy omówić zasoby niezbędne do osiągnięcia tego celu (koszty przechowywania ujęte w kontekście potencjalnej wartości danych, kto decyduje i w jaki sposób, jakie dane będą przechowywane i jak długo).

5. Bezpieczeństwo danych

Jakie zasady obowiązują lub będą obowiązywać w zakresie bezpieczeństwa danych (w tym odzyskiwania danych oraz bezpiecznego przechowywania/archiwizacji i przekazywania danych wrażliwych)?

Czy dane będą bezpiecznie przechowywane w zaufanych repozytoriach w celu długoterminowej ochrony?

6. Zagadnienia etyczne

Czy istnieją lub mogą istnieć kwestie etyczne lub prawne, mające wpływ na udostępnianie danych? Można je również omówić w kontekście oceny etycznej. W stosownych wypadkach należy uwzględnić odniesienia do dokumentów opisujących rezultaty projektu związane z aspektami etycznymi i rozdziału dotyczącego etyki w Description of the Action (DoA).

Czy świadoma zgoda na udostępnianie danych i ich długoterminowe przechowywanie będzie zawarta w kwestionariuszach dotyczących danych osobowych?

7. Inne kwestie

Czy wykorzystane będą inne procedury i zasady zarządzania danymi, np. obowiązujące na szczeblu krajowym, przyjęte przez inną instytucję finansującą badania lub wynikające ze specyfiki konkretnych sektorów? Jeśli tak, to z jakich (należy je wymienić i krótko opisać)?

Zadanie realizowane przez Narodowe Centrum Nauki na podstawie zlecenia Ministra Edukacji i Nauki dot. krajowej koordynacji partnerstwa European Open Science Cloud w latach 2022-2023.