

Ogólne zalecenia dotyczące przechowywania i wykorzystywania danych pochodzących z badań z udziałem grup wrażliwych

1. Wprowadzenie

Badania z udziałem grup wrażliwych wymagają znajomości kwestii prawnych i etycznych, a także podwyższonej wrażliwości etycznej, ponieważ uczestnicy mogą doświadczać nierównych relacji władzy, ograniczonej autonomii, zwiększonego narażenia na stygmatyzację lub ryzyk wykraczających poza sam kontekst badawczy (UKRI, 2025; Komisja Europejska, 2021). W takich warunkach etyczna praktyka badawcza opiera się na złożeniu wniosku do odpowiedniej komisji etycznej jeszcze przed rozpoczęciem pracy w terenie, przekazywaniu jasnych i przystępnych informacji uczestnikom badania, zapewnieniu dobrowolności udziału oraz na procesie uzyskiwania zgody, który w sposób wyraźny wyjaśnia granice poufności oraz przewidywane dalsze sposoby wykorzystania danych (UKRI, 2025; Komisja Europejska, 2021).

Jednocześnie zespoły badawcze powinny przestrzegać rygorystycznych standardów ochrony danych oraz wdrażać zasadę ochrony danych już na etapie projektowania. Zasada ta powinna być stosowana na wszystkich etapach cyklu życia projektu, w tym obejmować: planowanie zgodnego z prawem przetwarzania danych, ich minimalizację, zapewnienie bezpiecznego przetwarzania oraz terminowe usuwanie lub archiwizację danych (Komisja Europejska, 2021; Europejski Instytut Uniwersytecki, 2022). Ponieważ dane uczestników należących do grup wrażliwych częściej mają charakter danych wrażliwych lub mogą się nimi stać (dane szczególnych kategorii), badacze powinni stosować wzmocnione zabezpieczenia oraz zapewnić, że uzasadnienie gromadzenia takich danych jest konieczne i proporcjonalne (Komisja Europejska, 2021; ICO, 2024).

W przypadkach, gdy przetwarzanie danych może wiązać się z wysokim ryzykiem naruszenia praw i wolności osób fizycznych – co jest częste w badaniach z udziałem grup wrażliwych – projekty badawcze powinny przewidywać potrzebę przeprowadzenia ustrukturyzowanej oceny ryzyka oraz wdrożenia odpowiednich zabezpieczeń, w tym tych związanych z przetwarzaniem danych do celów badań naukowych oraz odpowiedzialnym zarządzaniem danymi (Komisja Europejska, 2021; Europejska Rada Ochrony Danych, 2021).

Wreszcie, przejrzystość dotyczy również sposobów udostępniania wyników badań i zbiorów danych. Dostęp do danych powinien być „tak otwarty, jak to możliwe”, przy jednoczesnym uzasadnieniu i proporcjonalności ograniczeń w sytuacjach, gdy są one konieczne dla ochrony jednostek i społeczności (UNESCO, 2021; ALLEA, 2023). W przypadkach, gdy udostępnianie danych jest zasadne, staranne zarządzanie danymi – takie jak bezpieczne środowiska przetwarzania, kontrolowany dostęp oraz zaawansowane metody anonimizacji lub pseudonimizacji – pozwala zachować wartość naukową danych przy jednoczesnej ochronie uczestników badań (Europejski Instytut Uniwersytecki, 2022; ENISA, 2021).

2. Kluczowe definicje

Grupy wrażliwe: Osoby, które mogą mieć ograniczoną zdolność do wyrażenia w pełni świadomej zgody lub mogą być bardziej narażone na doznanie szkody bądź wykorzystanie ze względu na wiek, niepełnosprawność, stan zdrowia, sytuację społeczno-ekonomiczną, status migracyjny, instytucjonalizację lub inne czynniki kontekstowe.

Dane osobowe: Wszelkie informacje umożliwiające bezpośrednią lub pośrednią identyfikację żyjącej osoby fizycznej.

Dane szczególnych kategorii: Dane wrażliwe wymagające dodatkowej ochrony na mocy RODO.

Pseudonimizacja: Przetwarzanie danych w taki sposób, aby nie można ich było przypisać do konkretnej osoby bez użycia dodatkowych informacji przechowywanych oddzielnie.

Anonimizacja: Nieodwracalne usunięcie informacji identyfikujących, w wyniku którego identyfikacja osób nie jest już możliwa.

3. Warunki wstępne

- Dane osobowe są gromadzone wyłącznie po uzyskaniu pełnej zgody uczestników badań na udział w badaniu, przy czym uczestnicy są świadomi, że mogą w każdej chwili wycofać się z udziału oraz że nie są w żaden sposób zmuszani do uczestnictwa;
- Uczestnikom musi zostać przekazana pełna informacja na temat celu gromadzenia danych (np. w formie materiałów informacyjnych rozdawanych uczestnikom), w języku i formie dostosowanym do możliwości potencjalnych

uczestników badań, w okresie umożliwiającym zapoznanie się z informacją i jej zrozumienie.

- Należy unikać ogólnikowych lub nieprecyzyjnych sformułowań. Zamiast tego badacze powinni w możliwie najbardziej konkretny sposób określać cel gromadzenia danych oraz zakres ich dostępności dla podmiotów trzecich;
- Zgoda etyczna oraz zgoda na gruncie RODO – pierwsza dotyczy udziału w badaniu, druga odnosi się do zgody na przetwarzanie danych osobowych. Każda z nich powinna być jasno określona i jednoznacznie zdefiniowana (Wyczik, 2025);
- Zgoda na gruncie RODO – w tym zakresie badacze powinni wyjaśnić, w jaki sposób zebrane dane będą przetwarzane, kto będzie miał do nich dostęp oraz w jakim celu zostaną wykorzystane (Wyczik, 2025).

4. Bezpieczeństwo danych, ich przechowywanie i wykorzystywanie

Anonimizacja

To proces, którego celem jest ochrona tożsamości uczestników badań poprzez uczynienie jej niedostępną i niemożliwą do ustalenia. Po skutecznej anonimizacji dane muszą nadal zachowywać wartość naukową (tj. dostarczać informacji istotnych dla danego obszaru badawczego) (Wieczorek, 2025). Jest to obowiązkowy etap po zakończeniu gromadzenia danych, a przed ich jakimkolwiek wykorzystaniem lub upublicznieniem.

W tym celu, oprócz usunięcia wszelkich imion i nazwisk, badacze powinni:

- dążyć do maksymalnego ograniczenia zakresu ujawnianych danych osobowych dotyczących uczestników badań;
- zachować szczególną czujność wobec wszelkich informacji, które mogłyby umożliwić identyfikację uczestników badań, co stanowi szczególne wyzwanie w badaniach z udziałem grup wrażliwych.

Jeżeli anonimizacja wymaga usunięcia znaczących części tekstu, prowadząc do tego, że tekst jest nieczytelny, to mamy do czynienia ze znaczącym obniżeniem jego jakości, a nawet zniszczeniem danych. Jeżeli w wyniku anonimizacji dane staną się bezużyteczne, to nie należy jej przeprowadzać. W takiej sytuacji nie udostępniamy danych publicznie, a udostępniamy jedynie metadane.

Plan zarządzania danymi (Data Management Plan)

Dokument określający szczegóły dotyczące gromadzenia, przetwarzania, zabezpieczania, przechowywania oraz wykorzystywania i ponownego

wykorzystywania danych badawczych. Dokument ten jest dynamiczny (“żywy”) i powinien być aktualizowany (wersjonowany) wraz ze zmianami wprowadzanymi w projekcie badawczym w kontekście danych badawczych. Plany zarządzania danymi uwzględniają praktyki badawcze oraz wymagania agencji finansujących badania (te mogą się różnić). W przypadku projektów badawczych realizowanych przez konsorcja występują dwa scenariusze zależne od praktyki partnerów lub wymagań grantodawcy: konsorcjum tworzy wspólny plan, uwzględniając praktyki badawcze związane z zarządzaniem danymi każdego z partnerów (na przykład Horyzont Europa) lub każdy partner posiada własny plan (niektóre konkursy NCN).

Otwarty dostęp do danych

Otwarta nauka stanowi obowiązujący standard. Dane badawcze powinny być udostępniane w otwartym dostępie, o ile nie istnieją uzasadnione ograniczenia lub inne przeszkody (np. związane z najlepiej pojętym interesem beneficjentów).

Korzystanie z zaufanych repozytoriów

To istotny element udostępniania danych zainteresowanym stronom. Zaufane repozytoria zapewniają odpowiednie warunki bezpieczeństwa, przechowania i kurowania danych. Często posiadają odpowiednie certyfikaty, np. CoreTrustSeal lub ISO.

Zestawy danych przechowywane w repozytoriach powinny:

- być opatrzone metadanymi, w tym informacją o źródłach finansowania badań,
- posiadać odpowiednią licencję otwartą, np. Creative Commons Public Domain (CC0) lub Creative Commons Attribution CC BY) lub tożsame.

Zaleca się także, aby posiadały PID (Persistent Identifier) np. DOI.

Wybór repozytorium powinien być podjęty nie tylko ze względu na wymagania techniczne i cyberbezpieczeństwo, ale także pod względem kwestii merytorycznych, a więc miejsca, gdzie dane będą cieszyć się najlepszą "widzialnością" i cytawalnością oraz możliwością ponownego wykorzystania. Jest to najbardziej optymalne rozwiązanie z punktu widzenia badacza i grantodawcy.

Do wyboru naukowcy mają repozytoria dyscyplinowe lub tematyczne; repozytoria instytucjonalne i ogólne.

Zasady FAIR (Findable, Accessible, Interoperable, Re-usable)

Dane powinny być łatwe do odnalezienia i dostępne dla zainteresowanych badaczy, interoperacyjne technicznie i interdyscyplinarnie oraz możliwe do ponownego wykorzystania. W tym celu konieczne jest również dostarczanie szczegółowych opisów danych lub zbiorów danych w formie metadanych i dokumentacji, a także nadanie im unikalnych, trwałych identyfikatorów PID (<https://www.go-fair.org/fair-principles/>).

W przypadku przechowywania danych przez instytucje zaangażowane w ich gromadzenie, bezpośrednio po ich zapisaniu należy uwzględnić następujące kwestie:

- dostęp do danych przed anonimizacją lub pseudonimizacją powinien być możliwie jak najbardziej ograniczony;
- surowe, niezanonimizowane dane powinny być przechowywane przez ograniczony czas w odizolowanej przestrzeni od danych zanonimizowanych/pseudonimizowanych (zgodnie z polityką instytucji);
- dane powinny być przechowywane, zgodnie z zasadą 3-2-1, powinna istnieć procedura zapisywania regularnego kopi zapasowych; w razie utraty/zniszczenia danych powinna istnieć procedura odzyskania danych; dane przechowywane wyłącznie w instytucjonalnych przestrzeni chmurowej (nie prywatnej), powinny być bezpieczne, w miarę możliwości, szyfrowane;
- w przypadku stosowania kopii zapasowych na serwerach, serwer powinien być odpowiednio zabezpieczony i niedostępny publicznie (np. bez dostępu przez Internet).

5. Podsumowanie

Odpowiedzialne zarządzanie danymi w badaniach z udziałem grup wrażliwych wymaga podwójnego zaangażowania: ochrony uczestników przed przewidywalnymi szkodami oraz jednoczesnego wspierania rzetelnych badań naukowych.

Zobowiązanie to rozpoczyna się od przygotowania wniosku do Komisji Etyki Badań, przez etyczne zaangażowanie uczestników, obejmujące przekazywanie jasnych i przystępnych informacji, zapewnienie dobrowolności udziału oraz wdrażanie procedur uzyskiwania zgody, które w sposób wyraźny odnoszą się do potencjalnych ryzyk, oczekiwanych korzyści oraz okoliczności, w których poufność może zostać ograniczona ze względów ochronnych lub prawnych (UKRI, 2025; Komisja Europejska, 2021).

Równie istotne jest wdrożenie rygorystycznych środków ochrony danych na wszystkich etapach cyklu życia projektu badawczego. Badacze muszą uwzględnić ochronę danych, zaczynając na etapie projektowania badań i konsekwentnie stosować ją w fazie realizacji. Obejmuje to zapewnienie, że gromadzenie danych osobowych jest konieczne i proporcjonalne, że wykorzystanie danych wrażliwych jest jasno uzasadnione, że dostęp do danych jest odpowiednio ograniczony oraz że okres przechowywania danych jest zawężony do minimum niezbędnego dla celów badawczych i wymogów rozliczalności (Komisja Europejska, 2021; Europejski Instytut Uniwersytecki, 2022).

W przypadku pracy z grupami wrażliwymi – gdzie poziom ryzyka jest z natury wyższy – kluczowe znaczenie mają ustrukturyzowane zabezpieczenia, takie jak solidne ramy zarządzania, strategie przechowywania danych oraz środki ochrony dostosowane do kontekstu, które pozwalają zachować zgodność z przepisami i ograniczyć ryzyko ujawnienia, niewłaściwego wykorzystania lub ponownej identyfikacji danych (Europejska Rada Ochrony Danych, 2021; Komisja Europejska, 2021).

Wreszcie, dobre praktyki wymagają starannego wyważenia zasad otwartej nauki z imperatywem ochrony uczestników należących do grup wrażliwych. Choć otwarta nauka i ponowne wykorzystywanie danych mogą zwiększać przejrzystość, powtarzalność i oddziaływanie badań, udostępnianie danych nigdy nie może odbywać się kosztem bezpieczeństwa, prywatności ani godności uczestników.

Wszelkie ograniczenia w zakresie udostępniania danych powinny być uzasadnione i proporcjonalne oraz wspierane przez solidne mechanizmy zarządzania danymi, w tym procedury kontrolowanego dostępu, kompleksową dokumentację oraz techniki ochrony prywatności czy bezpieczne środowiska przechowywania danych (UNESCO, 2021; ALLEA, 2023). Konsekwentne stosowanie tych zasad pozwala zespołom badawczym działać zgodnie z literą prawa oraz kodeksem etyki badań, a także utrzymywać zaufanie uczestników i ich społeczności, spełniać współczesne standardy rzetelności i odpowiedzialności badań oraz zapewnić, że postęp naukowy nie odbywa się kosztem osób najbardziej narażonych na doznanie szkody.

Literatura:

ALLEA. (2023, 23 czerwca). Europejski kodeks postępowania w zakresie rzetelności badań naukowych (wydanie zrewidowane 2023). <https://allea.org/code-of-conduct/>

Komisja Europejska, Dyrekcja Generalna ds. Badań Naukowych i Innowacji. (2021, 5 lipca). Etyka i ochrona danych (nota wytycznych programu „Horyzont Europa”).

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

Europejska Rada Ochrony Danych. (2021, 7 września). Studium prawne dotyczące odpowiednich zabezpieczeń na podstawie art. 89 ust. 1 RODO w przetwarzaniu danych osobowych do celów badań naukowych. https://www.edpb.europa.eu/our-work-tools/our-documents/legal-study-appropriate-safeguards-under-article-891-gdpr-processing_en

Agencja Unii Europejskiej ds. Cyberbezpieczeństwa (ENISA). (2021, 28 stycznia). Pseudonimizacja danych: zaawansowane techniki i przypadki użycia.

<https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases>

Europejski Instytut Uniwersytecki. (2022). Przewodnik dobrych praktyk w zakresie ochrony danych w badaniach naukowych.

<https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

Biuro Komisarza ds. Informacji (Information Commissioner’s Office). (b.d.). Oceny skutków dla ochrony danych (DPIA). Dostęp: 4 lutego 2026 r., <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/data-protection-impact-assessments/>

UK Research and Innovation, Rada ds. Badań Ekonomicznych i Społecznych (ESRC). (2025, 12 maja). Badania z udziałem osób potencjalnie wrażliwych.

<https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/research-with-potentially-vulnerable-people/>

UNESCO. (2021, 23 listopada). Rekomendacja UNESCO dotycząca otwartej nauki.

<https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>

Galica, Natalia. (2025). Wprowadzenie do zarządzania danymi badawczymi (RDM). Prezentacja ze szkolenia MANAGEMENT OF SENSITIVE RESEARCH DATA: ANONYMIZATION, STORAGE, AND CONTROLLED ACCESS, zorganizowanego w ramach projektu MIGRAEDU.

Wieczorek, Jan. (2025). Praca z danymi wrażliwymi: anonimizacja i bezpieczeństwo danych. Prezentacja ze szkolenia MANAGEMENT OF SENSITIVE RESEARCH DATA: ANONYMIZATION, STORAGE, AND CONTROLLED ACCESS, zorganizowanego w ramach projektu MIGRAEDU.

Wyczik, Jakub. (2025). Podstawy RODO dla badaczy. Prezentacja ze szkolenia MANAGEMENT OF SENSITIVE RESEARCH DATA: ANONYMIZATION, STORAGE, AND CONTROLLED ACCESS, zorganizowanego w ramach projektu MIGRAEDU.