

Wytyczne dla wnioskodawców do uzupełnienia PLANU ZARZĄDZANIA DANYMI w projekcie badawczym

W tej części formularza pod pojęciem „dane” NCN rozumie zarówno dane zebrane i dotąd nieprzetworzone, jak i dane wytworzone i poddane analizie, inne niż publikacje naukowe. Definicja ta obejmuje wszystkie możliwe formaty, zarówno cyfrowe, jak i niecyfrowe (np. próbki, wypełnione kwestionariusze, nagrania dźwiękowe, itd.).

Plan zarządzania danymi można uznać za część planu badań; jego celem jest uzupełnienie tego ostatniego o techniczny opis sposobu zarządzania danymi. NCN dopuszcza, że w ramach niektórych projektów nie będą wytwarzane, na nowo wykorzystywane, ani poddawane analizie żadne dane badawcze ani inne podobne materiały. W takich wypadkach wymagane jest jednak krótkie uzasadnienie.

W wypełnieniu poniższej części formularza pomogą Państwu pracownicy biblioteki/odpowiedniego repozytorium/działu IT Państwa instytutu lub uniwersytetu.

	Pytania	Pytania pomocnicze
1.	Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych	
	<p>1.1 W jaki sposób będą pozyskiwane lub wytwarzane nowe dane lub ponownie wykorzystywane dane już istniejące? Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none"> – Jakie standardy, metody i oprogramowanie posłużą do pozyskiwania lub wytwarzania nowych danych? – Jakie dane już istniejące (własne lub będące w posiadaniu strony trzeciej) zostaną ponownie wykorzystane? – W jaki sposób zostanie udokumentowane pochodzenie danych? – Jak wyglądać będzie organizacja plików i zarządzanie ich różnymi wersjami? – Czy konieczna będzie digitalizacja danych analogowych lub wydanych w formie papierowej (takich jak mapy, fotografie, teksty)? 	<p>Prosimy o wyjaśnienie, w jaki sposób będą pozyskiwane lub wytwarzane nowe dane lub ponownie wykorzystywane dane już istniejące. Należy również pokrótce opisać ich rodzaj, np. teksty, obrazy, zdjęcia, pomiary, statystyki, próbki fizyczne czy kody. Prosimy o określenie, w jaki sposób planują Państwo kontrolować i dokumentować ich jakość i spójność: procesy kalibracji, powtórne pobieranie próbek lub ponowne wykonywanie pomiarów, standardy zapisywania/nagrywania i utrwalania danych, wykorzystanie narzuconego słownictwa, walidacja wprowadzanych danych, specjalistyczne recenzje, itd.</p>
	<p>1.2 Jakie dane (tj. rodzaje, formaty, objętości) będą pozyskiwane lub wytwarzane w projekcie? Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none"> – Jaki rodzaj, format i objętość danych planują Państwo pozyskać, wytworzyć lub ponownie wykorzystać? 	<p>Odpowiedź powinna uwzględnić rodzaj, format i zawartość każdego planowanego zbioru danych. Należy również przedstawić ich szacunkową objętość. Prosimy o wyszczególnienie formatu, tj. sposobu, w jaki dane będą kodowane na potrzeby</p>



		przechowywania, na co często wskazuje rozszerzenie odpowiedniego pliku (np. .pdf, .xls). Pierwszeństwa należy udzielić formatom otwartym i standardowym.
2. Dokumentacja i jakość danych		
2.1 Jakie metadane i dokumentacja (np. metodologia oraz sposoby pozyskiwania i organizacji danych) będą towarzyszyć danym w projekcie? Warto rozważyć następujące kwestie: <ul style="list-style-type: none">– Jakie informacje należy uzupełnić, aby potencjalni użytkownicy (komputery lub ludzie) byli w stanie w przyszłości odczytać i zinterpretować zebrane dane?– Czy możliwe jest ich komputerowe odczytanie?– W jaki sposób powstanie odpowiednia dokumentacja?– Jakie standardy środowiskowe (jeżeli w ogóle) posłużą do objaśnienia (meta)danych?– Jakie międzynarodowe standardy lub schematy (tj. Dublin Core, DDI) posłużą do organizacji metadanych?¹	Prosimy o określenie rodzaju metadanych mających ułatwić użytkownikom odnalezienie i identyfikację poszukiwanych informacji. Mogą one obejmować informacje dotyczące tytułu plików, źródła pochodzenia danych, numeru identyfikacyjnego autora (np. ORCID), metodologii formatów służących do pozyskania danych, definicji zmiennych i jednostek pomiarowych. Prosimy o wskazanie sposobu organizacji danych w trakcie projektu poprzez podanie np. przyjętej konwencji, wersji i struktury folderów. Należy określić w jaki sposób i gdzie informacje takie zostaną utrwalone, tj. w bazie danych z łączem prowadzącym do poszczególnych elementów, plikach README, książce kodowej, itd.	
2.2 Jakie planują Państwo zastosować środki kontroli jakości? Warto rozważyć następujące kwestie: <ul style="list-style-type: none">– W jaki sposób metody pozyskiwania, analizy i przetwarzania danych mogą wpływać na ich jakość?– W jaki sposób można wyeliminować błędy pomiarowe i problem stroniczości?– W jaki sposób zminimalizować ryzyko dotyczące poprawności danych?	Prosimy o wykazanie wysokiej jakości danych w projekcie. Czy udokumentowano metody ich pozyskiwania i analizy? Należy wskazać przyjęte w instytucji mechanizmy mające na celu zabezpieczenie danych przez nieuprawnioną modyfikacją i określić w jaki sposób/na jakim etapie zostanie wdrożona ocena jakości. Należy również wyjaśnić, czy dane ilościowe wymagają oczyszczenia.	

¹ Digital Curation Centre (DCC) prowadzi listę szeroko stosowanych standardów metadanych. Można ją znaleźć pod adresem: <http://www.dcc.ac.uk/resources/metadata-standards>.



3. Przechowywanie i tworzenie kopii zapasowych podczas badań	
<p>3.1 W jaki sposób w trakcie projektu będą przechowywane dane i metadane? W jaki sposób będą tworzone ich kopie zapasowe? Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none">– Jaką dysponują Państwo pojemnością pamięci i gdzie będą przechowywane dane?– Jak wyglądają procedury tworzenia kopii zapasowych?– Czy potrzebne będą specjalne metody przenoszenia danych z urządzeń mobilnych, stanowisk terenowych lub sprzętu domowego na główny serwer w miejscu pracy?	<p>Prosimy o opisanie wymogów związanych z przechowywaniem danych i miejsca ich składowania. W przypadku przechowywania długoterminowego należy zdecydować, jakie dane zostaną zachowane, jaka będzie ich objętość i okres składowania. Należy pamiętać, że przechowywanie danych na laptopach i dyskach twardych może się okazać ryzykowne. Bezpieczniejsze będzie wykorzystanie do tego celu pomocy zespołów IT. Prosimy o opisanie procedury tworzenia kopii zapasowych (częstotliwość aktualizacji, podział obowiązków, proces automatyczny/ręcznie kontrolowany, środki bezpieczeństwa, itd.). Kto będzie odpowiedzialny za tworzenie kopii zapasowych i odzyskiwanie danych? Jeżeli w projekcie zaangażowanych jest kilkoro naukowców, prosimy o stworzenie wraz ze współpracownikami planu, który pozwoli zapewnić bezpieczeństwo przepływu danych między uczestnikami.</p>
<p>3.2 W jaki sposób zostanie zapewnione bezpieczeństwo i ochrona danych wrażliwych w okresie trwania projektu? Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none">– W jaki sposób można będzie odzyskać dane utracone w wyniku incydentu?– Kto uzyska dostęp do danych w czasie trwania projektu i jak wyglądać będzie kontrola dostępu do danych, zwłaszcza w przypadku współpracy między kilkoma partnerami?	<p>Jeżeli przechowywanie danych wymaga usługi zewnętrznej, bardzo ważne jest, by było to zgodne z polityką każdego zaangażowanego w projekt podmiotu, zwłaszcza w przypadku danych wrażliwych. Prosimy o rozważenie sposobu ochrony danych, przede wszystkim jeśli są one wrażliwe, np. zawierają dane osobowe, dane wrażliwe politycznie lub dane podwójnego zastosowania. Prosimy o określenie, jaką politykę ochrony danych wdrożono w Państwa instytucji.</p>



4.	Wymogi prawne, kodeksy postępowania 4.1 Jeżeli będzie miało miejsce przetwarzanie danych osobowych, w jaki sposób zostanie zapewniona zgodność z przepisami dotyczącymi danych osobowych oraz ich ochrony? Warto rozważyć następujące kwestie: <ul style="list-style-type: none">– Czy w procesie pozyskiwania danych niezbędna będzie ich anonimizacja?– Czy należy usunąć informacje umożliwiające identyfikację osób lub zataić tożsamość uczestników (np. za pomocą pseudonimizacji) przed udostępnieniem danych?	Prosimy pamiętać, że w przypadku prawa o ochronie danych osobowych (tj. RODO) konieczne będzie uzyskanie świadomej zgody uczestników na utrwalanie i udostępnianie ich danych osobowych. Prosimy wziąć pod uwagę możliwość anonimizacji lub pseudonimizacji, a także szyfrowania danych, co stanowi szczególny przypadek pseudonimizacji (klucz do szyfru musi być przechowywany w innym miejscu niż same dane). Prosimy o stwierdzenie, czy wprowadzono specjalną procedurę dostępu dla uprawnionych użytkowników danych osobowych.
	4.2 W jaki sposób planują Państwo zapewnić zgodność z innymi przepisami, takimi jak prawa własności intelektualnej i prawa własności? Jakie przepisy znajdują w tym przypadku zastosowanie? Warto rozważyć następujące kwestie: <ul style="list-style-type: none">– Kto będzie właścicielem danych?– Jakie zostaną zastosowane licencje?– Jakie będą ograniczenia ponownego wykorzystania danych pochodzących od osób trzecich?– Czy przed udostępnieniem danych konieczne będzie uzyskanie odpowiedniego zezwolenia w zakresie praw autorskich?	Prosimy o wskazanie właścicieli praw autorskich i praw własności intelektualnej do wszelkich pozyskiwanych i wytwarzanych danych, a także odpowiednich licencji. W przypadku konsorcjów badawczych niezbędne może się okazać podpisanie umowy dotyczącej praw własności intelektualnej. Prosimy również o wyjaśnienie, czy istnieją jakiegokolwiek ograniczenia prawne dotyczące ponownego wykorzystania danych pochodzących od osób trzecich. Czy ma to wpływ na prawa własności intelektualnej (np. dyrektywa 96/9/EC z 11 marca 1996 r. w sprawie ochrony prawnej baz danych, prawa <i>sui generis</i>)?
5.	Udostępnianie i długotrwałe przechowywanie danych 5.1 Kiedy i w jaki sposób będą udostępniane dane z projektu? Czy istnieją ewentualne ograniczenia i zakazy dotyczące ich udostępniania? Warto rozważyć następujące kwestie:	Dane należy udostępnić możliwie szybko – najpóźniej w momencie publikacji wyników badań. Prosimy również wziąć pod uwagę, w jaki sposób ponowne wykorzystanie Państwa



	<ul style="list-style-type: none">– W jaki sposób o Państwa danych dowiedzą się potencjalni użytkownicy?– Jak długo będą one przechowywane?– Czy istnieją jakieś ograniczenia i przeszkody uniemożliwiające ich pełne lub częściowe udostępnienie?– Czy wydawcy czasopism będą wymagać składowania danych na poparcie ustaleń publikacji?– Czy udostępnianie danych wymaga zgody uczestników badania?	danych zostanie ocenione i stwierdzone przez innych naukowców. Kiedy zostaną udostępnione dane? Prosimy o uzasadnienie długości ich przechowywania ² . Należy również wskazać oczekiwaną datę udostępnienia i określić, czy ulegnie ono opóźnieniu lub ograniczeniu np. w celu publikacji, ochrony własności intelektualnej, czy ubiegania się o patenty. Czy umowa o poufności będzie w stanie zapewnić odpowiednią ochronę danych poufnych?
5.2	Jak będzie wyglądać selekcja danych przeznaczonych do utrwalenia i gdzie będą one długoterminowo przechowywane (np. w repozytorium danych, archiwum)? Warto rozważyć następujące kwestie: <ul style="list-style-type: none">– Które dane trzeba zachować, a które zniszczyć z przyczyn wynikających z umów, przepisów prawnych lub regulacji?– W jaki sposób zostanie podjęta decyzja o tym, które dane zachować?– Jaka będzie procedura selekcji przeznaczonych do utrwalenia danych?– Z jakiego repozytorium będą Państwo korzystać? Czy przestrzega ono zasad FAIR Data³?– Czy Państwa instytucja tworzy regularnie kopie zapasowe?	Prosimy o określenie, jak i w jakim repozytorium ⁴ zostaną udostępnione dane. Prosimy wziąć pod uwagę koszty depozytu i pamięci. Jaka pojemność pamięci będzie niezbędna w okresie trwania projektu, a jaka po jego zakończeniu? Czy zdecydowali się Państwo na cyfrowe repozytorium prowadzone przez organizację non-profit?
5.3	Jakie metody lub oprogramowanie umożliwiają dostęp do danych i korzystanie z	Metody udostępniania danych będą zależec

² Dane surowe i przetworzone powinny być przechowywane przez okres odpowiedni dla danej dyscypliny i zastosowanej metodologii. W rozumieniu NCN uzasadniony okres przechowywania danych to minimum 10 lat.

³ Zasady FAIR Data określają szereg atrybutów, jakie musi posiadać opublikowany zbiór danych, aby był łatwy do odnalezienia, dostępny, interoperacyjny i nadający się do wielokrotnego użytku (patrz: Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, doi:10.1038/sdata.2016.18) oraz <https://www.force11.org/fairprinciples>.

⁴ Istnieje szereg międzynarodowych systemów certyfikacji określających poziom wiarygodności poszczególnych repozytoriów danych. Pośród nich najbardziej podstawowy zestaw kryteriów to Data Seal of Approval certyfikowany przez Core Trust Seal. Zaufane repozytoria cyfrowe ze znakiem jakości obejmują repozytoria z certyfikatem Data Seal of Approval, DIN-31644-, ISO-16363- lub WDS/ICSU. Przegląd takich repozytoriów można znaleźć pod niniejszym [adresem](#). Inne przydatne wykazy to: Registry of Research Data Repositories <https://www.re3data.org/>. Niektóre repozytoria, takie jak Zenodo, OpenAIRE czy CERN, pozwalają na przechowywanie zarówno danych, jak i publikacji, zapewniając również odpowiednie narzędzia dostępu. Zaleca się korzystanie w pierwszej kolejności z repozytoriów certyfikowanych lub szeroko uznanych w danej dyscyplinie, a jeżeli takie repozytoria nie istnieją, kierować się w swoim wyborze kryteriami zawartymi w dokumencie *Practical Guide to the international alignment of research data management*, https://www.scienceurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf.



	<p>danych?</p> <p>Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none">– Czy aby dane mogły być przechowywane przez dłuższy okres i zachowały długi okres ważności należy je przekształcić do formatu standardowego lub otwartego?– Czy do skanowania lub konwersji niezbędny będzie dodatkowy sprzęt lub oprogramowanie?– Jaki mechanizm posłuży do udostępniania danych (np. odpowiedzi na żądanie, repozytorium)?	<p>od kilku czynników takich, jak ich rodzaj, rozmiar danych, złożoność i wrażliwość. Prosimy o wskazanie czy potencjalni użytkownicy będą potrzebować określonych narzędzi, aby uzyskać do nich dostęp i (ponownie) je wykorzystać. Prosimy o zwrócenie uwagi na zrównoważony charakter niezbędnego oprogramowania.</p>
	<p>5.4 W jaki sposób zagwarantują Państwo stosowanie unikalnego i trwale przypisanego identyfikatora (np. takiego jak cyfrowy identyfikator dokumentu elektronicznego (DOI)) dla każdego zbioru danych?</p> <p>Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none">– Czy trwale przypisany identyfikator zostanie uzyskany?– Jakie istniejące trwale identyfikatory pozostaną w użyciu (np. cyfrowe identyfikatory dokumentu elektronicznego, numery dostępowe)?	<p>Prosimy wyjaśnić w jaki sposób dane mogą zostać ponownie wykorzystane w odmiennym kontekście. Trwale przypisany identyfikator pozwala na trafną i wydajną lokalizację danych, służy jako odniesienie oraz pozwala na łatwe śledzenie cytowań i wielokrotne użycie.</p>
6.	Zadania związane z zarządzaniem danymi oraz zasoby	
	<p>6.1 Kto będzie odpowiadał za zarządzanie danymi (tj. kto będzie ich opiekunem)?</p> <p>Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none">– Czy Państwa instytucja posiada personel odpowiedzialny za zarządzanie danymi w trakcie trwania projektu, tj. od ich utworzenia, początkowego przechowywania, aż do archiwizacji i długoterminowego przechowywania danych?– Jeśli NIE to kto jest odpowiedzialny za archiwizację i długoterminowe zarządzanie danymi w Państwa Jednostce?– Jeśli TAK to jakie zajmuje stanowisko w obrębie Jednostki?– Jaką rolę pełni Kierownik projektu w zarządzaniu danymi?	<p>Należy wyraźnie rozróżnić między właścicielem danych, a ich opiekunem. Ten ostatni to specjalista ds. jakości danych odpowiedzialny za ich ocenę (środki zaradcze), ale nie bezpośrednią korektę. Prosimy o opisanie podziału ról i obowiązków w zakresie zarządzania/opieki nad danymi i wskazanie, kto odpowiada za wdrożenie planu zarządzania, a także jego weryfikację i korektę. W przypadku projektu opartego na współpracy, prosimy o wyjaśnienie sposobu koordynacji działań w zakresie zarządzania danymi pomiędzy poszczególnymi partnerami.</p>



<p>6.2 Jakie zasoby zostaną przeznaczone na cele zarządzania danymi i zagwarantowanie przestrzegania zasad FAIR⁵?</p> <p>Warto rozważyć następujące kwestie:</p> <ul style="list-style-type: none">– Czy potrzebne są dodatkowe zasoby do zarządzania danymi, takie jak osoby, czas, sprzęt lub oprogramowanie?– Jakie koszty związane będą z zapewnieniem standardów FAIR w projekcie?– W jaki sposób zostaną opłacone?	<p>Prosimy o wyjaśnienie, w jaki sposób koszty niezbędne do przygotowania danych do ich udostępniania i przechowywania będą oszacowane? Czy do przygotowania danych do przechowywania i archiwizowania niezbędne będą dodatkowe zasoby? Jeżeli tak, jak wysokie będą to koszty i w jaki sposób zostaną opłacone? Rozważ konieczność wydatkowania środków na długoterminowe przechowywanie danych?</p>
---	---

Dodatkowe informacje:

Jeżeli składają Państwo wniosek w konkursie międzynarodowym organizowanym przez NCN w oparciu o procedurę agencji wiodącej, w którym rolę agencji wiodącej pełni zagraniczna instytucja partnerska, to bez względu na to, czy ta agencja wymaga przedłożenia planu zarządzania danymi czy nie, nadal mają Państwo obowiązek uzupełnienia go w formularzu wniosku krajowego składanym w systemie ZSUN/OSF. Treść planu dotyczy tylko danych badawczych, które powstaną lub będą ponownie wykorzystywane w trakcie realizacji projektu przez polski zespół badawczy.

Jeżeli składają Państwo wniosek w konkursie międzynarodowym organizowanym przez NCN we współpracy wielostronnej, to bez względu na to, czy na poziomie międzynarodowym wymagane jest przedłożenie planu zarządzania danymi czy nie, nadal mają Państwo obowiązek uzupełnienia go w formularzu wniosku krajowego składanym w systemie ZSUN/OSF. Treść planu dotyczy tylko danych badawczych, które powstaną lub będą ponownie wykorzystywane w trakcie realizacji projektu przez polski zespół badawczy.

Plany zarządzania danymi mają charakter bardzo indywidualny. Istnieje wiele ich rodzajów, a ich organizacja może się różnić. Różnorodność tę ilustrują [przykłady przedstawione na stronie Digital Curation Centre](#) (UK).

⁵ Istnieje szereg narzędzi pozwalających samodzielnie ocenić stopień, w jakim dane spełniają kryteria FAIR, np. <https://www.ands-nectar-rds.org.au/fair-tool>.