



eR POZYTORIUM
I POZYCJONOWANIE
WYDAWNICTW KUL

**PORADNIK – Plan zarządzania danymi
badawczymi**

Spis treści

- 1. Wprowadzenie**
- 2. Plan Zarządzania Danymi**
- 3. Narzędzia pomocne przy tworzeniu planu**
- 4. Co ważne przy konstruowaniu Planu Zarządzania Danymi**
- 5. Plan Zarządzania Danymi krok po krok w pytaniach**
- 6. Przykładowy plan zarządzania danymi badawczymi.**

WPROWADZENIE

Dane badawcze (Research Data) – zarejestrowane materiały o charakterze faktograficznym (w postaci liczbowej, tekstowej, graficznej czy dźwiękowej), powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.

Dane badawcze dzielić można 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.

Otwarte dane badawcze (Open Research Data) – dane, do których każdy ma dostęp. Można je ponownie wykorzystywać, modyfikować, redystrybuować i udostępniać z poszanowaniem prawa.

Deponować można dane ze wszystkich obszarów wiedzy, które zostały wytworzone, zebrane lub opisane na potrzeby działalności badawczej. Zakres danych zebranych w jednym zbiorze ustala autor.

Plan zarządzania danymi (Data Management Plan – DMP) – określa, w jaki sposób dane badawcze mają być zarządzane zarówno w czasie trwania projektu badawczego, jak i po jego zakończeniu. Plan zarządzania danymi uznawany jest za część planu badań.

Instytucje i programy finansujące badania naukowe, czasopisma coraz częściej wymagają od naukowców przedstawienia DMP na etapie składania i oceny wniosków grantowych czy przed publikacją artykułu (*np. Horyzont 2020, NCN*).

Dane badawcze powinny być łatwe do znalezienia. W planie zarządzania danymi powinny znaleźć się informacje na temat sposobów pozyskiwania danych, ich walidacji, pochodzenia, możliwości ponownego wykorzystania oraz ich rodzaj.

Plan zarządzania danymi powinien zawierać informacje nt.:

Danych

- ✓ jakie dane zostaną wytworzone lub zebrane (format i typ plików, liczba danych)
- ✓ jak zostaną uporządkowane i opisane (metodologia, standardy, metadane)
- ✓ kwestie etyczne i prawne (własność intelektualna, prawa autorskie, dane niejawne)
- ✓ w jaki sposób dane zostaną udostępnione (jak, kiedy, komu)
- ✓ które dane będą przechowywane długoterminowo (kwestia sposobu przechowywania i ochrony danych)

Licencji (aspekty prawne)

- ✓ Creative Commons – CC0, CC-BY, CC-BY-SA
- ✓ Open Data Commons – licencje wyłącznie do baz danych – posiadają skomplikowane rozwiązania prawne; obejmują całą bazę danych, ale nie obejmują jej części składowych

Narzędzia pomocne podczas tworzenia DMP

DMPtool (US) – narzędzie przygotowujące szablony DMP dostosowane do wymagań amerykańskich grantodawców

DMPonline (UK) – narzędzie bardzo podobne do DMPtool zawierające jednak bazę instytucji finansującej naukę z Wielkiej Brytanii

DCC Data Management Plan Content Checklist – lista kontrolna zawartości planu zarządzania danymi, szablon projektu

(http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf) – pozwala szybko określić, jakich informacji może brakować w przygotowywanym DMP

Serwis brytyjskiej instytucji specjalizującej się w zarządzaniu danymi badawczymi – [The Data Curation Center](#) – znajdują się tam gotowe plany zarządzania danymi, szablony, przewodniki, wytyczne, informacje na temat formatów zamieszczanych danych badawczych, metadanych w zależności od dyscypliny oraz lista kontrolna ([Checklist for a Data Management Plan](#)) dotycząca zawartości DMP, która pozwala szybko określić, jakich informacji może brakować w przygotowywanym DMP.

Jak napisać dobry Plan Zarządzania Danymi? Pamiętajmy!

- ✓ Powinien być prosty i w miarę krótki, ale konkretny
- ✓ Powinien być realistyczny, oparty na dostępnych informacjach
- ✓ Może się zmieniać

Co ważne przy konstruowaniu Planu Zarządzania Danymi

1. Opis danych w projekcie badawczym
2. Dokumentacja dotycząca danych i ich jakości
3. Przechowywanie danych i ich bezpieczeństwo
4. Wymogi prawne
5. Udostępnianie i długotrwałe przechowywanie danych

Rodzaje zbieranych danych

- ✓ Dokumenty tekstowe, notatki
- ✓ Dane liczbowe
- ✓ Kwestionariusze
- ✓ Ankiety, wyniki badań ankietowych
- ✓ Nagrania audio i video, zdjęcia
- ✓ Próbkki, artefakty, obiekty
- ✓ Zawartość baz danych (video, audio, teksty, obrazy)
- ✓ Modele matematyczne, algorytmy
- ✓ Oprogramowanie (skrypty, pliki wejściowe...)
- ✓ Wyniki symulacji komputerowych
- ✓ Protokoły laboratoryjne,
- ✓ Opisy metodologiczne

Formaty plików

Formaty mogą być dowolne, jednak zasadniczo powinien być wybierany format zapewniający powszechny dostęp i otwartość. Ważne jest by w jak największym stopniu wykorzystywać otwarte oprogramowanie do odczytu danych. Należy również dobrze przemyśleć nazewnictwo plików.

Dokumentacja i metadane

Dokumentacja powinna opisywać metodologię prowadzonych badań oraz ich kontekst i źródło. Informuje o sposobie organizacji danych w trakcie projektu, np. przyjętej konwencji, wersji, strukturze folderów.

Metadane umożliwiają scharakteryzowanie danych badawczych w taki sposób, żeby potencjalny użytkownik wiedział, jakiego rodzaju są to dane. Dane badawcze zawsze muszą być udostępniane z metadanymi. Do tych metadanych zalicza się przykładowo: autor, tytuł, data powstania, licencja, dyscyplina naukowa.

Przechowywanie i bezpieczeństwo danych

Należy dokładnie przeanalizować tę kwestię dla całego procesu gromadzenia i ewentualnego przetwarzania danych badawczych. Należy również przeanalizować kwestię dostępu do danych, zwłaszcza danych poufnych, ochrony danych, a w związku z tym tworzenia kopii zapasowych.

Ochrona danych (dane wrażliwe, dane chronione prawem autorskim)

Dane wrażliwe to dane ujawniające pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz dane genetyczne, dane biometryczne jednoznacznie identyfikujące osoby fizyczne lub dane dotyczące zdrowia, seksualności lub orientacji seksualnej tej osoby.

Jeżeli badania zakładają gromadzenie i/lub przetwarzanie danych wrażliwych konieczne jest zapewnienie właściwej ochrony.

Dostęp i ponowne użycie danych badawczych

Zapewnienie dostępu do danych badawczych polega na ich udostępnieniu oraz opisaniu. Należy określić, kiedy dane zostaną udostępnione, czy będzie to dostęp pełny, czy ograniczony oraz czy będzie możliwe ponowne użycie.

Czas przechowywania danych

Długoterminowa archiwizacja danych badawczych to przechowywanie w dłuższym okresie czasu. Dane surowe i przetworzone powinny być przechowywane przez okres odpowiedni dla danej dyscypliny i

zastosowanej metodologii. W rozumieniu NCN **uzasadniony okres przechowywania to minimum 10 lat.**

Miejsca przechowywania danych

Aby dane cyfrowe pozostały dostępne i mogły być wykorzystywane także po długim czasie, muszą być aktywnie zarządzane. Deponowanie danych w bezpiecznych repozytoriach cyfrowych daje nam pewność, że będą zarządzane w sposób zgodny z dobrymi praktykami archiwizacji cyfrowej.

Zasady FAIR Data służą jako wytyczne dla umożliwienia ponownego wykorzystania danych naukowych w wyraźnie opisanych warunkach, zarówno przez ludzi, jak i przez maszyny.

Zasady FAIR Data:

Findable – łatwo znajdowane i wyszukiwane.

Accessible – dostępne dla wszystkich.

Interoperable – interoperacyjne, tak aby można było je połączyć z innymi danymi.

Reusable – wielokrotnego użytku.

Cyfrowe repozytoria danych

- ✓ specjalistyczne
- ✓ instytucjonalne
- ✓ szeroko zakrojone tematycznie
- ✓ ogólne

Wyszukiwarka re3data.org (Registry of Research Data Repositories) to rejestr repozytoriów danych badawczych ze wszystkich dyscyplin akademickich. Jest to darmowe narzędzie, które oferuje naukowcom, organizacjom finansującym badania, bibliotekom i wydawcom przegląd repozytoriów danych badawczych. Umożliwia wyszukiwanie repozytoriów według dziedziny wiedzy, kraju oraz typu danych badawczych.

Istnieje wiele międzynarodowych systemów certyfikacji określających poziom wiarygodności poszczególnych repozytoriów danych. Podstawowy zestaw kryteriów to Data Seal of Approval. 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.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf).

Polecane repozytoria

Zenodo

Repozytorium zostało opracowane dzięki inicjatywie OpenAIRE i CERN. Przeznaczone dla tzw. małych danych. Umożliwia proste archiwizowanie i dzielenie się opracowanymi przez siebie danymi przez naukowców prowadzących badania we wszystkich dziedzinach wiedzy.

RepOD

Repozytorium otwartych danych badawczych opracowane przez ICM UW. Przeznaczone dla tzw. małych danych. Umożliwia archiwizowanie i udostępnianie różnych rodzajów danych wytworzonych przez naukowców z różnych dziedzin wiedzy.

ReKUL

Repozytorium Instytucjonalne KUL umożliwia archiwizację i udostępnianie różnych rodzajów danych wytworzonych przez pracowników naukowych KUL. Przeznaczone do tzw. małych danych.

Prawa autorskie i licencje

Należy wskazać właścicieli praw autorskich i praw własności intelektualnej do wszelkich pozyskiwanych i wytwarzanych danych. Trzeba określić, czy istnieją jakiegokolwiek ograniczenia prawne dotyczące ponownego wykorzystania danych pochodzących od osób trzecich. Koniecznym jest wskazanie licencji dla udostępnianych danych badawczych. Rekomenduje się korzystanie z otwartych licencji.

PLAN ZARZĄDZANIA DANYMI KROK PO KROKU W PYTANIACH

1. Identyfikacja projektu badań

- ✓ Jaki jest tytuł projektu?
- ✓ Kto jest autorem projektu?
- ✓ Jaka jest afiliacja?
- ✓ Jaka jest data utworzenia/modyfikacji planu?
- ✓ Streszczenie projektu
- ✓ Kto ma prawa autorskie do planu zarządzania danymi?

2. Opis danych w projekcie badawczym

- ✓ W jaki sposób dane będą pozyskiwane, wytwarzane i czy będą wykorzystywane dane już istniejące?
- ✓ Jakie dane zostaną zebrane lub wytworzone (formaty, rodzaje, woluminy, wielkości)?
- ✓ W jaki sposób planowane jest kontrolowanie i dokumentowanie jakości i spójności danych?
- ✓ Jakimi będą standardy zapisywania/nagrywania i utrwalania danych?
- ✓ Jak będzie wyglądać organizacja plików?

3. Dokumentacja dotycząca danych i ich jakości

- ✓ Jakie dodatkowe informacje mogą być potrzebne do korzystania z tworzonych danych (dokumentacja)?
- ✓ Jakie metadane będą wykorzystane, w jakim formacie, dzięki którym będzie możliwe wyszukanie i identyfikacja poszukiwanych informacji?
- ✓ Jaka będzie struktura folderów?
- ✓ W jaki sposób i gdzie te informacje zostaną utrwalone?
- ✓ Jak będzie kontrolowana jakość i poprawność danych?

4. Przechowywanie danych i ich bezpieczeństwo

- ✓ Gdzie przechowujemy nasze dane na bieżąco?
- ✓ W jaki sposób je zabezpieczamy?
- ✓ Jakie są procedury przenoszenia danych?
- ✓ Jak wyglądają procedury tworzenia kopii zapasowych (częstotliwość aktualizacji, środki bezpieczeństwa)?
- ✓ Kto jest odpowiedzialny za tworzenie kopii zapasowych i odzyskiwanie danych?
- ✓ Kto uzyska dostęp do danych w czasie trwania projektu?
- ✓ Kto będzie miał do nich dostęp i na jakich zasadach?
- ✓ Jaka jest polityka ochrony danych w instytucji, przy której realizowany jest projekt?

5. Wymogi prawne

- ✓ Jak chronione są dane osobowe w przypadku konieczności ich przetwarzania (RODO)?
- ✓ Czy konieczne będzie uzyskiwanie świadomej zgody na przetwarzanie i udostępnianie danych osobowych?
- ✓ Czy konieczna będzie anonimizacja danych?
- ✓ Kto będzie właścicielem danych?
- ✓ Jakie zostaną zastosowane licencje?
- ✓ Jak będą zabezpieczane prawa autorskie i prawa własności intelektualnej do pozyskiwanych i wytwarzanych danych?
- ✓ Czy będą istniały ograniczenia prawne dotyczące ponownego wykorzystania danych przez osoby trzecie?

6. Udostępnianie i długoterwałe przechowywanie danych

- ✓ Jakie dane chcemy przechowywać po zakończeniu projektu?
- ✓ Gdzie zdeponujemy dane do przechowywania długoterminowego?
- ✓ Jak długo będziemy je przechowywać?
- ✓ W jaki sposób będzie można uzyskać informację o danych (np. kiedy zostaną udostępnione)?
- ✓ Czy istnieją ograniczenia w pełnym udostępnieniu danych, jeśli tak to jakie?
- ✓ Czy wydawcy czasopism będą wymagać składowania danych na poparcie ustaleń publikacji?
- ✓ Jaka będzie procedura selekcji danych przeznaczonych do utrwalenia?
- ✓ W jakim repozytorium dane zostaną zarchiwizowane?
- ✓ Czy dane będą miały nadany identyfikator dokumentu elektronicznego (np. DOI, Handle)?

7. Zadania związane z zarządzaniem danymi oraz zasoby

- ✓ Kto będzie odpowiadał za zarządzanie danymi?
- ✓ Czy do przygotowania danych do przechowywania i archiwizowania niezbędne będą dodatkowe zasoby?
- ✓ Jakie będą koszty i w jaki sposób zostaną opłacone?

PRZYKŁADOWY PLAN ZARZĄDZANIA DANYMI BADAWCZYMI

Pole	Opis	Elementy
Informacje ogólne	Zawiera informacje ogólne dotyczące projektu i gromadzonych danych	<ol style="list-style-type: none"> 1. Wnioskodawca. 2. Afiliacja. 3. ORCID. 4. Projekt i opis. 5. Finansowanie.
Opis zbioru danych lub wykorzystanie istniejących danych	A\ Zawiera opis, jak będą wyglądać nowe dane zebrane lub wyprodukowane i czy/lub zostaną ponownie wykorzystane	<ol style="list-style-type: none"> 1. Jakie metody lub oprogramowanie zostanie użyte do zebrania lub wytworzenia danych? 2. Czy przewidywane są ograniczenia dotyczące ponownego wykorzystania danych. 3. Dokumentacja dotycząca pochodzenia danych.
	B\ Zawiera opis, jakie dane (np. rodzaj, format) zostaną zebrane lub wyprodukowane	<ol style="list-style-type: none"> 1. Szczegóły dotyczące rodzaju danych. 2. Szczegóły dotyczące formatu danych. 3. Uzasadnienie dla użycia konkretnych formatów. 4. Szczegóły dotyczące użycia otwartych formatów. 5. Szczegóły dotyczące wielkości woluminów.
Dokumentacja i jakość danych	A\ Zawiera opis, jakie metadane i dokumentacja potrzebne są przy wybranych danych	<ol style="list-style-type: none"> 1. Jakie metadane zostaną udostępnione? 2. Określić standardy metadanych (najlepiej użyć tych najbardziej powszechnych). 3. Określić sposób, w jaki dane zostaną zorganizowane i uporządkowane podczas projektu. 4. Wskazać opcjonalnie inną dokumentację potrzebną do ponownego użycia.
	B\ Zawiera opis, jakie środki kontroli jakości danych zostaną zastosowane	<ol style="list-style-type: none"> 1. Wyjaśnienie, jak spójność i jakość zebranych danych będzie

		kontrolowana i udokumentowana.
	C\ Zawiera opis, w jaki sposób, gdzie dane będą udostępniane i w jakim stopniu otwarte	1. Restrykcje dotyczące ponownego wykorzystania danych, ich udostępniania i przekształcania.
Przechowywanie i kopia zapasowa podczas badań	A\ Zawiera opis, gdzie dane będą archiwizowane i na jak długo	1. Metody zastosowane do archiwizowania i zapewnienia długotrwałego dostępu. 2. Ochrona danych i ich trwałość.
	B\ Zawiera opis, jakie działania należy wykonać, żeby zapewnić bezpieczeństwo danych	1. Procedury odzyskiwania danych. 2. Dostęp do danych, zwłaszcza kontrolowany danych wrażliwych. 3. Ochrona danych instytucjonalnych.
Wymogi prawne i etyczne, kodeks postępowania	A\ Zawiera opis, jakie dane osobowe są przetwarzane (zgodność działań z ustawą o ochronie danych osobowych)	1. Dane osobowe, zgody na ich przetwarzanie, zachowanie i udostępnianie. 2. Zastosowanie anonimizacji, pseudoanonimizacji.
	B\ Zawiera opis, jak jest rozwiązany problem z własnością intelektualną	

Przykładowy plan wygenerowany z DMPTool

Green city development plan for sebeta city

A Data Management Plan created using DMPTool

Creator: manaye teshome

Affiliation: Non Partner Institution

Template: Digital Curation Centre (DCC)

Project abstract:

In its effort to change the backward economic, social and Environmental status of the city, the Sebeta city administration is undertaking various construction and natural preservation project to provide public services corresponding to the needs of the users of the city administration as part of "Good Governance"& sustainable economic development. It is a fact that town/city planning is experiencing a crisis; at least the concept of town planning that appeared with the Modern Movement and its myriad of planners. The city phenomenon is continuously increasing and extending the boundaries of the city area. Some indicators and experts' opinions indicate a lower index of city planning than expected by the UN. However, the reality in the cities presents a number of problems that are worth mentioning. The invasion of pericity spaces causes the loss of land, natural resources and landscape. All these problems will be solved by 'Green City' concept.

Last modified: 07-17-2019

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Data Collection

1. Topographic data of sebeta city and surroundings from sebeta city administration.
2. Climatic data (annual rainfall, temperature, humidity, wind speed and direction) of the past 50 years for sebeta city
3. DEM (digital elevation model) open access databases
4. Demographic data of sebeta city from sebeta city administration
5. Master plan of sebeta city both hard and soft copy from sebeta city administration
6. Aerial photographs of the city with high resolution
7. **Collect the relevant socio-economic data:** Collect demographic data by using data collection formats, Collect Social

information by using data collection formats Collect the key economic data by using data collection formats .

8. **Collect the relevant physical/spatial and environmental data:** Prepare topographic maps in 1:2000 or 1:2000 scale, Collect the existing detailed land use data and indicate on the prepared map.

9. **Collect the necessary institutional information**

- **Compile collected data systematically:** Encode the empirical information, Tally the qualitative information, Interpret the data organized, and attach the necessary explanatory notes and Site sources of data properly
- survey using field tools, such GPS device.

Documentation and Metadata

- date created, size, coverage, type
- Arch-GIS version 10 or above is needed
- Auto Desk Revit 20 is needed to develop and read drawing and maps.

Ethics and Legal Compliance

- data will be secured and stored (anonymously)
- The Data will be owned by the project team
- CC BY-NC-ND- type of creative common license will be used.

Storage and Backup

- 2 TB separate hard disk 3 personal computers will be used for data storage
- open access sites such as ZENODO will be used for backup
- proper data security and storage is needed to avoid missing data in case of PC failure. we will use online open access data storage for backup.

Selection and Preservation

- Master plan of the city is used in the project period and should hand over to the city administration office
- re3data.org and zenodo.org will be used to store data.

Responsibilities and Resources

- core team of the project will be responsible to implement the DMP
- separate team will be established to handle data management plan issues
- software version controls should be maintained.

