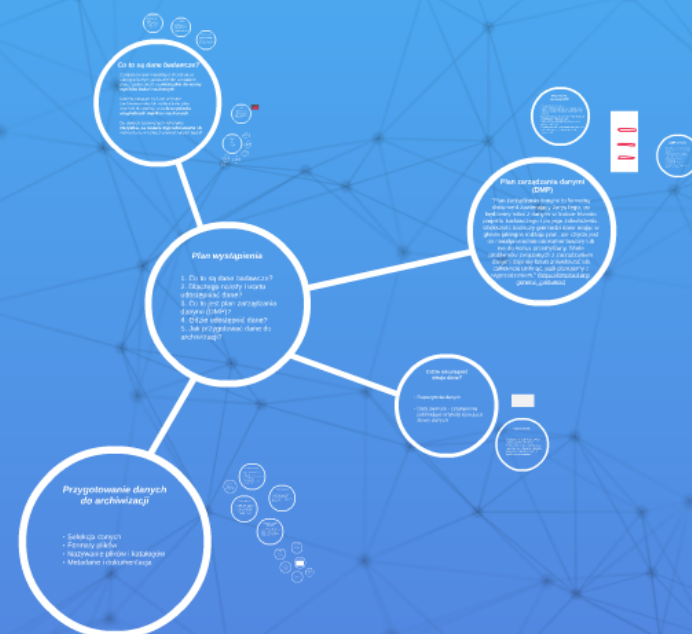


Warsztaty z zarządzania danymi badawczymi

Wojciech Fenrich - wf@icm.edu.pl

Link do tej prezentacji: <https://bit.ly/rdmlodz>

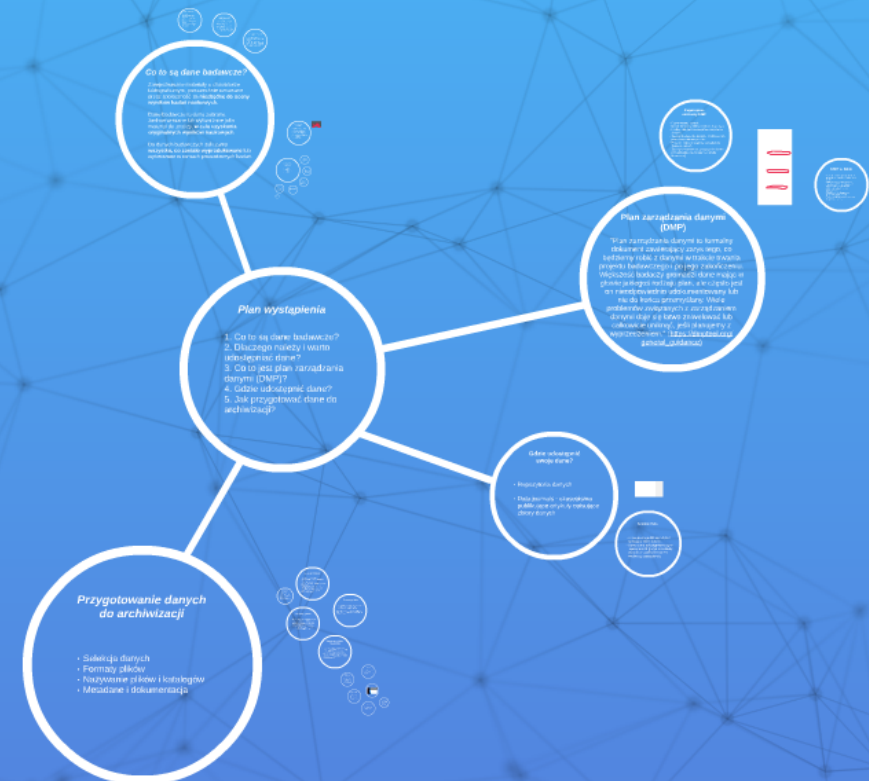


Łódź, 25 października 2019 r.

Warsztaty z zarządzania danymi badawczymi

Wojciech Fenrich - wf@icm.edu.pl

Link do tej prezentacji: <https://bit.ly/rdmlodz>



Plan wystąpienia

1. Co to są dane badawcze?
2. Dlaczego należy i warto udostępniać dane?
3. Co to jest plan zarządzania danymi (DMP)?
4. Gdzie udostępnić dane?
5. Jak przygotować dane do archiwizacji?

Co to są dane badawcze?

Zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane przez społeczność 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**.

Do danych badawczych zaliczamy **wszystko, co zostało wyprodukowane** lub wytworzone w ramach prowadzonych badań.

Horyzont 2020 -
Wymogi

- Domyślnie dane muszą zostać udostępnione najszybciej co do zakresu koniecznej weryfikacji ustalonej w publikacjach.
- Możliwość "opt-out"

Horyzont 2020 -
FAIR Data

- Findable
- Accessible
- Interoperable
- Reusable

Przebieg

W ramach projektu...

W ramach projektu...

W ramach projektu...

W ramach projektu...

Dane badawcze to:

- Dane liczbowe.
- Dokumenty tekstowe, notatki.
- Kwestionariusze, wyniki badań ankietowych.
- Nagrania audio i wideo, zdjęcia.
- Zawartość baz danych.
- Modele matematyczne, algorytmy.
- Oprogramowanie (skrypty, pliki wejściowe).
- Wyniki symulacji komputerowych.
- Protokoły laboratoryjne, opisy metodologiczne.
- ...

Dlaczego należy i warto udostępniać dane?

- Bo to ułatwia prowadzenie badań.
- Bo nie trzeba tych samych danych wytwarzać kilkukrotnie.
- Bo niektórych danych nie sposób wytworzyć ponownie.
- Bo mogą stać za tym argumenty etyczne.
- Bo to dla udostępniającego szansa na zdobycie cytowań i zareklamowanie siebie.

Dlaczego to nas dotyczy?

- Bo czasopisma naukowe coraz częściej wymagają zdeponowania danych.
- Bo instytucje finansujące coraz częściej wymagają przygotowania planu zarządzania danymi (DMP) i/lub udostępnienia danych.

Horyzont 2020 - Wymogi

- Domyślnie dane powinny zostać udostępnione jak najszybciej co najmniej w zakresie koniecznym do weryfikacji ustaleń opisanych w publikacjach.
- Możliwość "opt out".

Opting out – partially or entirely

By extending the pilot, open access data generated in Horizon 2020.

However, not all data can be made open (either before or after signing the obligations associated with the pilot).

- participation is incompatible with the obligations reasonably be expected to be generated by the project
- participation is incompatible with security issues
- participation is incompatible with the obligations of the project
- participation would mean that the project will not generate open access data
- there are other legitimate reasons for not making the data open (proposal stage).

The Commission's approach can be closed as necessary".

Opting out – partially or entirely

By extending the pilot, open access becomes the default setting for research data generated in Horizon 2020.

However, not all data can be open. Projects can therefore opt out at any stage (either before or after signing the grant) and so free themselves retroactively from the obligations associated with the conditions – if:

- participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited
- participation is incompatible with the need for confidentiality in connection with security issues
- participation is incompatible with rules on protecting personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate / collect any research data or
- there are other legitimate reasons (*you can enter these in a free-text box at the proposal stage*).

The Commission's approach can therefore be described as "*as open as possible, as closed as necessary*".

Horyzont 2020 - FAIR Data

- Findable
- Accessible
- Interoperable
- Reusable

- Czy dane
metadane
- Czy będą
przyjętymi
- Czy będą
- Czy konie
wersjonov

Accessi

- Jakie oprogramowanie,
bo można było wykorzy
- Jeśli korzystanie z danych
ograniczone, w jaki spo
zostanie zapewniony?

Findable

- Czy dane zostaną opatrzone metadanymi?
- Czy będą opisane zgodnie z przyjętymi standardami?
- Czy będą posiadać DOI?
- Czy konieczne będzie ich wersjonowanie?

Accessible

- Które dane zostaną udostępnione w sposób otwarty?
- Jeśli część danych nie może zostać udostępniona - dlaczego?
Co nas tu ogranicza: prawo, umowa, nasza własna uzasadniona wstrzemięźliwość?
- W jaki sposób dane zostaną udostępnione (np. w repozytorium danych)?

Accessible

- Jakie oprogramowanie jest potrzebne, bo można było wykorzystać te dane?
- Jeśli korzystanie z danych musi zostać ograniczone, w jaki sposób dostęp zostanie zapewniony?
- Czy powinien zostać powołany *data access committee*?
- Czy warunki dostępu są jasno określone (np. w postaci licencji możliwej do maszynowego odczytu)?

Interoperable

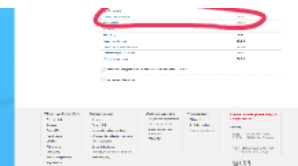
- Czy będzie możliwa wymiana i ponowne wykorzystanie danych przez inne osoby pochodzące z innych instytucji oraz państw?
- Czy możliwe będzie ich połączenie z innymi zbiorami pochodzącymi z innych źródeł?
- Czy ich przetwarzanie będzie możliwe za pomocą otwartego oprogramowania?
- Czy wykorzystane zostaną standardowe słowniki metadanych, standardowe metodologie?

Reusable

- Czy dane zostaną opatrzone licencją, która pozwoli na ich ponowne wykorzystanie w stopniu tak szerokim, jak to możliwe?
- Kiedy możliwe będzie ponowne wykorzystanie danych?
- Czy dane objęte zostaną karencją (tzw. embargo), by umożliwić publikację lub uzyskanie patentu? Jeśli tak - jak długo?

Unusable data

- "Otwarty zbiór danych".
- "Dane publicznie dostępne".
- "Dane ogólnodostępne".
- "Brak ograniczeń".
- "Freely available".
- "Dane dostępne na licencji Creative Commons"



1. Tytuł	2. Data
3. Autor	4. Wersja
5. Data	6. Wersja
7. Data	8. Wersja
9. Data	10. Wersja
11. Data	12. Wersja
13. Data	14. Wersja
15. Data	16. Wersja
17. Data	18. Wersja
19. Data	20. Wersja
21. Data	22. Wersja
23. Data	24. Wersja
25. Data	26. Wersja
27. Data	28. Wersja
29. Data	30. Wersja
31. Data	32. Wersja
33. Data	34. Wersja
35. Data	36. Wersja
37. Data	38. Wersja
39. Data	40. Wersja
41. Data	42. Wersja
43. Data	44. Wersja
45. Data	46. Wersja
47. Data	48. Wersja
49. Data	50. Wersja
51. Data	52. Wersja
53. Data	54. Wersja
55. Data	56. Wersja
57. Data	58. Wersja
59. Data	60. Wersja
61. Data	62. Wersja
63. Data	64. Wersja
65. Data	66. Wersja
67. Data	68. Wersja
69. Data	70. Wersja
71. Data	72. Wersja
73. Data	74. Wersja
75. Data	76. Wersja
77. Data	78. Wersja
79. Data	80. Wersja
81. Data	82. Wersja
83. Data	84. Wersja
85. Data	86. Wersja
87. Data	88. Wersja
89. Data	90. Wersja
91. Data	92. Wersja
93. Data	94. Wersja
95. Data	96. Wersja
97. Data	98. Wersja
99. Data	100. Wersja

Plan zarządzania danymi (DMP)

"Plan zarządzania danymi to formalny dokument zawierający zarys tego, co będziemy robić z danymi w trakcie trwania projektu badawczego i po jego zakończeniu. Większość badaczy gromadzi dane mając w głowie jakiegoś rodzaju plan, ale często jest on nieodpowiednio udokumentowany lub nie do końca przemyślany. Wiele problemów związanych z zarządzaniem danymi daje się łatwo zniwelować lub całkowicie uniknąć, jeśli planujemy z wyprzedzeniem." (https://dmptool.org/general_guidance)

Zwyczajowe elementy DMP

- Pozyskiwanie danych.
- Dobór formatów plików i ich nazewnictwo.
- Krótko- i długoterminowe przechowywanie danych.
- Zasady dostępu do danych. Możliwość ich ponownego wykorzystania.
- Prawne i etyczne aspekty zarządzania zbiorem danych.
- Zasoby potrzebne do zarządzania danymi (infrastruktura, kompetencje, środki finansowe).

Na skróty
[Baza projektów NCN](#)
[Konkursy](#)
[Dla wnioskodawców](#)
[Szkolenia dla wnioskodawców](#)
[Dla realizujących projekty](#)
[Dla ekspertów](#)
[Najczęściej zadawane pytania](#)
[Statystyki konkursów](#)
[Oferty pracy w projektach](#)

Instrukcje przygotowania wniosków

— Instrukcje przygotowania wniosków w konkursach na projekty badawcze, staże i stypendia

W związku z wdrożoną od konkursów ogłoszonych 15 marca 2019 r. standaryzacją formularzy wniosków w systemie ZSUN/OSF  informujemy, że sposób wypełniania poniższych sekcji formularza wniosku jest taki sam dla konkursów NCN wymienionych w tabeli.

Nazwa sekcji	Nazwa konkursu				
Logowanie, rejestracja, wybór roli w systemie	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Zakładanie nowego wniosku	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Wniosek nowy/powtórzony	OPUS	PRELUDIUM	-	-	-
Informacje podstawowe	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Wnioskodawca/Podmiot realizujący	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Pomoc publiczna	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Współpraca międzynarodowa	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Streszczenie	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Streszczenie popularnonaukowe	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Plan badań	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Kwestie etyczne	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Zbliżone zadania badawcze	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Opis skrócony	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Opis szczegółowy	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Wynagrodzenia i stypendia	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Aparatura	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Inne koszty	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Obniżenie pensum dydaktycznego	-	-	SONATA BIS	-	-
Koszty - zestawienia - staże	-	-	-	-	UWERTURA
Plan zarządzania danymi	-	-	SONATA BIS	MAESTRO	-
Wynagrodzenia administracyjne	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA

Streszczenie popularnonaukowe	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Plan badań	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Kwestie etyczne	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Zbliżone zadania badawcze	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Opis skrócony	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Opis szczegółowy	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Wynagrodzenia i stypendia	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Aparatura	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Inne koszty	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Obniżenie pensum dydaktycznego	-	-	SONATA BIS	-	-
Koszty - zestawienia	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-
Koszty - zestawienia - staże	-	-	-	-	UWERTURA
Plan zarządzania danymi	-	-	SONATA BIS	MAESTRO	-
Wydatki administracyjne	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Ochrona danych osobowych	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Wysyłka elektroniczna	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Informacje o stażu	-	-	-	-	UWERTURA
Opis	-	-	-	-	UWERTURA

Instrukcje wypełniania sekcji "Zespół badawczy" i "Wynagrodzenia i stypendia" ze względu na specyfikę poszczególnych konkursów jest przygotowana oddzielnie dla każdego konkursu.

Nazwa sekcji	Nazwa konkursu				
Zespół badawczy	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	UWERTURA
Wynagrodzenia i stypendia	OPUS	PRELUDIUM	SONATA BIS	MAESTRO	-

Informacje podstawowe	MOZART	ALPHORN	BIODIVCLIM
Wnioskodawca/Podmiot realizujący	MOZART	ALPHORN	BIODIVCLIM
Pomoc publiczna	MOZART	ALPHORN	BIODIVCLIM
Współpraca międzynarodowa	MOZART	ALPHORN	BIODIVCLIM
Streszczenie popularnonaukowe	MOZART	ALPHORN	BIODIVCLIM
Plan badań	MOZART	ALPHORN	BIODIVCLIM
Zespół badawczy	MOZART	ALPHORN	BIODIVCLIM
Zobowiązania badawcze	MOZART	ALPHORN	BIODIVCLIM
Kwestie etyczne	MOZART	ALPHORN	BIODIVCLIM
Plan zarządzania danymi	MOZART	ALPHORN	BIODIVCLIM
Wynagrodzenia i stypendia	MOZART	ALPHORN	BIODIVCLIM
Aparatura	MOZART	ALPHORN	BIODIVCLIM
Inne koszty	MOZART	ALPHORN	BIODIVCLIM
Koszty – zestawienia	MOZART	ALPHORN	BIODIVCLIM
Oświadczenia administracyjne	MOZART	ALPHORN	BIODIVCLIM
Ochrona danych osobowych	MOZART	ALPHORN	BIODIVCLIM
Wysyłka elektroniczna	MOZART	ALPHORN	BIODIVCLIM

Instrukcja przygotowania wniosków w konkursie GRIEG

Nazwa sekcji	Nazwa konkursu
Logowanie, rejestracja, wybór roli w systemie	GRIEG
Nowy wniosek	GRIEG
Informacje podstawowe	GRIEG

Wnioskodawca/Podmiot realizujący	GRIEG
Pomoc publiczna	GRIEG
Współpraca międzynarodowa	GRIEG
Streszczenie popularnonaukowe	GRIEG
Harmonogram	GRIEG
Zespół badawczy	GRIEG
Zbliżone zadania badawcze	GRIEG
Opis	GRIEG
Kwestie etyczne	GRIEG
Plan zarządzania danymi	GRIEG
Wynagrodzenia	GRIEG
Aparatura	GRIEG
Inne koszty	GRIEG
Koszty – zestawienia	GRIEG
Oświadczenia administracyjne	GRIEG
Ochrona danych osobowych	GRIEG
Wysyłka elektroniczna	GRIEG



Instrukcja przygotowania wniosków w konkursie **MINIATURA 3**



Archiwalne instrukcje

DMP w NCN

- Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych.
- Dokumentacja i jakość danych.
- Przechowywanie i tworzenie kopii zapasowych podczas badań.
- Wymogi prawne, kodeksy postępowania.
- Udostępnienie i długotrwałe przechowywanie danych.
- Zadania związane z zarządzaniem danymi oraz zasoby.

Gdzie udostępnić swoje dane?

- Repozytoria danych
- Data journals - czasopisma publikujące artykuły opisujące zbiory danych

- D
- Czasopi
opisując
 - Same da
repozyto
w postac
wydawc

[Search](#)

[Browse](#) ▾

[Suggest](#)

[Resources](#) ▾

[Contact](#)



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

 Search



2,000 Data Repositories

Three new DOI Fabrica

One step closer towards

Data journals

- Czasopismo publikuje artykuł opisujący zbiór danych.
- Same dane udostępniane są w repozytoriach (jedynie niekiedy w postaci suplementu przez wydawcę czasopisma).

Przygotowanie danych do archiwizacji

- Selekcja danych
- Formaty plików
- Nazywanie plików i katalogów
- Metadane i dokumentacja

Selekcja danych

Forma
Na co dzień n
na takich for
jakich nam w
jest, by prze
przekonw
odpowie

Selekcja danych

- Do czego jesteśmy **zobowiązani**?
- Jaka jest **wartość naukowa lub historyczna** danych?
- Jak **unikalne** są nasze dane?
- Czy istnieje **możliwość ich ponownego zebrania/** wytworzenia?
- Czy dane są **dobrej jakości** i dają się łatwo wykorzystać przez innych?
- Jakie **koszty** wiązałyby się z zarządzaniem i przechowywaniem danych?
- Czy dysponujemy kompletną, poprawną **dokumentacją** danych?

Selekcja danych

- Jeśli czegoś nie archiwizujemy, **udokumentujmy**, co zostało usunięte ze zbioru i dlaczego.
- Nasze **dane nigdy nie będą idealne** - udostępniajmy to, co jest wystarczająco dobre. Nie bójmy się opisywać wad i braków w naszych danych.
- Selekcja danych zależy od tego, co **chcemy** zrobić, co **wolno** nam zrobić i **na co nas stać**.

Formaty plików

- Niewymagające komercyjnego oprogramowania
- Otwarte, z dostępną dokumentacją
- Niewymuszające stratnej kompresji.

Przykłady dobrych formatów plików: JPEG2000, PNG, TIFF, TXT, HTML, PDF/A, WAVE, AIFF, TAR, GZIP, ZIP, XML, CSV.

Formaty plików

Na co dzień możemy pracować na takich formatach plików, w jakich nam wygodnie. Istotne jest, by przed archiwizacją przekonwertować je na odpowiedni format.

Nazywanie plików i katalogów

- Katalog najwyższego rzędu powinien zawierać tytuł projektu, identyfikator i datę.
- Na niższych poziomach stosowana powinna być konsekwentnie przyjęta konwencja (np. jeden plik/katalog - jedna osoba, jeden przebieg eksperymentu albo jedna wersja danych).

Co poza danymi?

- **Metadane:** podstawowe informacje stanowiące opis całego zbioru danych (autor, tytuł zbioru, data utworzenia, licencja etc.)
- **Dokumentacja:** informacje o charakterze metodologicznym, kontekst powstania, skrypty (dodatkowe pliki konieczne, by można było skorzystać z danych), słowniki, klucze kodowe etc.

Dokumentacja - poziom projektu

- Opis celu i kontekstu badań.
- Opis metodologiczny sposobu pozyskania danych.
- Struktura plików z danymi, relacje między plikami.
- Linki do publikacji powiązanych z danymi.
- Formularze zgody na udział w badaniu.
- Zastosowanie metody kontroli danych.
- Informacja o wersjach zbiorów danych.
- Informacja o zbiorach danych, których nie można udostępnić.

Dokumentacja projektu - poziom danych

- Nazwy i opisy zmiennych/rekordów.
- Opis schematów klasyfikacyjnych.
- Informacje o urządzeniach pomiarowych, ich ustawieniach i metodach kalibracji.
- Oznaczenia brakujących wartości oraz wyjaśnienie, dlaczego ich brakuje.
- W przypadku badań jakościowych - lista przypadków lub obiektów objętych badaniem.
- Wyniki testów jakości danych.

Dokumentacja

- Może znajdować się w samych plikach z danymi.
- Może znajdować się w osobnym pliku.

http://data.research.cornell.edu/sites/default/files/SciMD_ReadMe_Guidelines_v4_1_0.pdf

Guidelines for writing “readme” style metadata

by Wendy Kozlowski (wak57@cornell.edu). Last updated 05.30.2014

Adapted from Dryad's recommendations for authors (<http://datadryad.org/depositing>) now available at <https://web.archive.org/web/20120413115438/http://www.datadryad.org/depositing> and from The Knowledge Network for Biocomplexity's Introduction to Ecological Metadata Language (EML) (http://knb.ecoinformatics.org/eml_metadata_guide.html), now available at https://web.archive.org/web/20120424124714/http://knb.ecoinformatics.org/eml_metadata_guide.html

Best practices:

Create one readme file for each data file, whenever possible. It is also appropriate to describe a “dataset” that has multiple, related, identically formatted files, or files that are logically grouped together for use (e.g. a collection of Matlab scripts). When appropriate, also describe the file structure that holds the related data files. See Example 2 for describing grouped or multiple files.

Name the readme so that it is easily associated with the data file(s) it describes.

Write your readme document as a plain text file, avoiding proprietary formats such as MS Word whenever possible.¹ Format the readme document so it is easy to understand (e.g. separate important pieces of information with blank lines, rather than having all the information in one long paragraph).

Format multiple readme files identically. Present the information in the same order, using the same terminology.

Use standardized date formats. Suggested format: [W3C/ISO 8601 date standard](#), which specifies the international standard notation of YYYYMMDD or YYYYMMDDThhmmss.

Follow the scientific conventions of your discipline for taxonomic, geospatial and geologic names and keywords. Whenever possible, use terms from standardized taxonomies and vocabularies, a few of which are listed below.

Guidelines for writing “readme” style metadata

by Wendy Kozlowski (wak57@cornell.edu). Last updated 05.30.2014

Adapted from Dryad’s recommendations for authors (<http://datadryad.org/depositing>) now available at <https://web.archive.org/web/20120413115438/http://www.datadryad.org/depositing> and from The Knowledge Network for Biocomplexity’s Introduction to Ecological Metadata Language (EML) (http://knb.ecoinformatics.org/eml_metadata_guide.html, now available at https://web.archive.org/web/20120424124714/http://knb.ecoinformatics.org/eml_metadata_guide.html)

Best practices:

Create one readme file for each data file, whenever possible. It is also appropriate to describe a “dataset” that has multiple, related, identically formatted files, or files that are logically grouped together for use (e.g. a collection of Matlab scripts). When appropriate, also describe the file structure that holds the related data files. See Example 2 for describing grouped or multiple files.

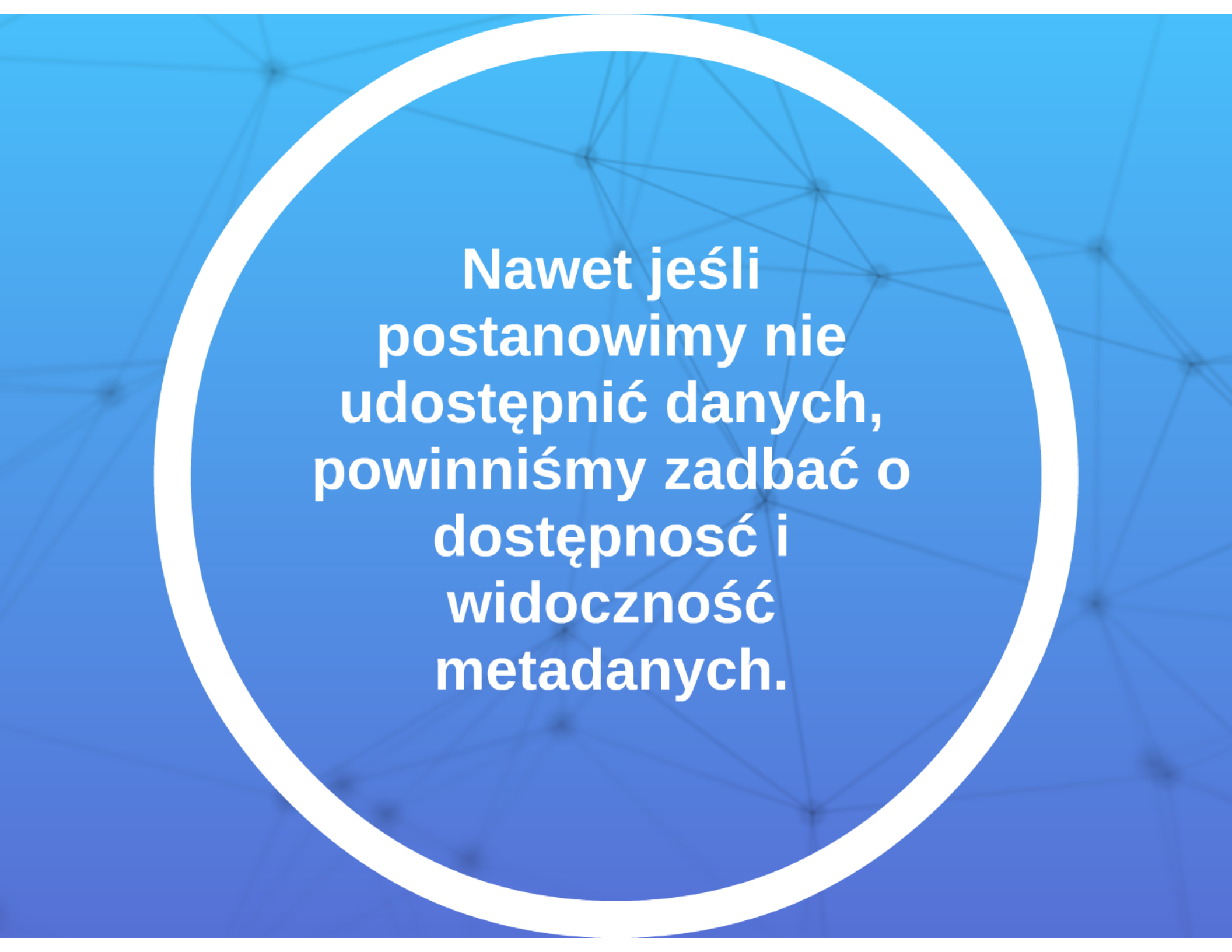
Name the readme so that it is easily associated with the data file(s) it describes.

Write your readme document as a plain text file, avoiding proprietary formats such as MS Word whenever possible.¹ Format the readme document so it is easy to understand (e.g. separate important pieces of information with blank lines, rather than having all the information in one long paragraph).

Format multiple readme files identically. Present the information in the same order, using the same terminology.

Use standardized date formats. Suggested format: [W3C/ISO 8601 date standard](http://www.w3.org/TR/1995/ISO-8601-199507-1/), which specifies the international standard notation of YYYYMMDD or YYYYMMDDThhmmss.

Follow the scientific conventions of your discipline for taxonomic, geospatial and geologic names and keywords. Whenever possible, use terms from standardized taxonomies and vocabularies, a few of which are listed below.



**Nawet jeśli
postanowimy nie
udostępnić danych,
powinniśmy zadbać o
dostępność i
widoczność
metadanych.**

Warsztaty z zarządzania danymi badawczymi

Wojciech Fenrich - wf@icm.edu.pl

Link do tej prezentacji: <https://bit.ly/rdmlodz>

