

Zarządzanie danymi badawczymi w naukach o Ziemi i środowisku

Dariusz Ignatiuk*, Aleksandra Osika

Uniwersytet Śląski w Katowicach

**Polskie Konsorcjum Polarne*



NARODOWE CENTRUM NAUKI

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.



Ministerstwo
Edukacji i Nauki



Dlaczego zarządzanie danymi badawczymi jest ważne w naukach o Ziemi i środowisku

Dzielić się danymi,
czy się nie dzielić...
Oto jest pytanie!



<https://doi.org/10.1371/journal.pbio.1001779.g001>

License: CC BY 4.0





Dlaczego zarządzanie danymi badawczymi jest ważne w naukach o Ziemi i środowisku

Zaleca się wprowadzanie przez poszczególne jednostki naukowe szczegółowych procedur i regulacji. Pełna odpowiedzialność za informowanie pracowników naukowych o przyjętych zasadach postępowania spoczywa na jednostkach naukowych. Naukowcy oraz jednostki naukowe powinni znać regulacje prawne, które odnoszą się m.in. do przetwarzania danych osobowych, ochrony własności intelektualnej oraz obowiązujących zasad etycznych.

Ministry of High Education and Science, Danish Code of Conduct for Research Integrity, listopad 2014
Kodeks Narodowego Centrum Nauki dotyczący rzetelności badań naukowych i starania o fundusze na badania



<https://doi.org/10.1371/journal.pbio.1001779.g001>

License: CC BY 4.0





Dlaczego zarządzanie danymi badawczymi jest ważne w naukach o Ziemi i środowisku

Udostępnianie danych wymaga wysiłku, zasobów i wiary w innych. Konieczna jest zmiana kulturowa

Dlaczego to robić?

- Wspiera przejrzystość i otwartość badań
- Zoptymalizuje finansowanie badań
- Promuje nieprzewidziane badania
- Kariera i korzyści naukowe
- Wiele więcej!

Kto na tym korzysta?

- Opinia publiczna
- Darczyńca
- Społeczność naukowa
- Naukowcy

Dlaczego zarządzanie danymi badawczymi jest ważne w naukach o Ziemi i środowisku

80% danych jest niedostępnych po 20 latach od publikacji.
Gibney and Van Noorden (2013), Nature



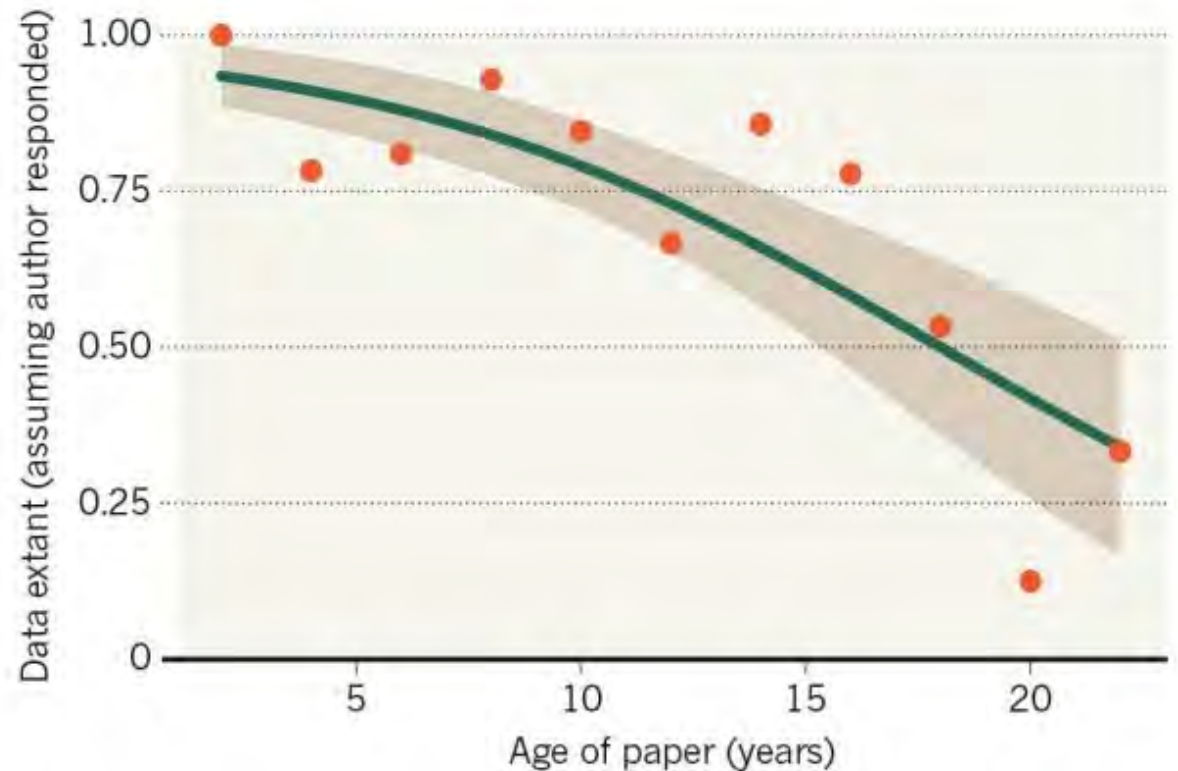
DANE
są niedostępne



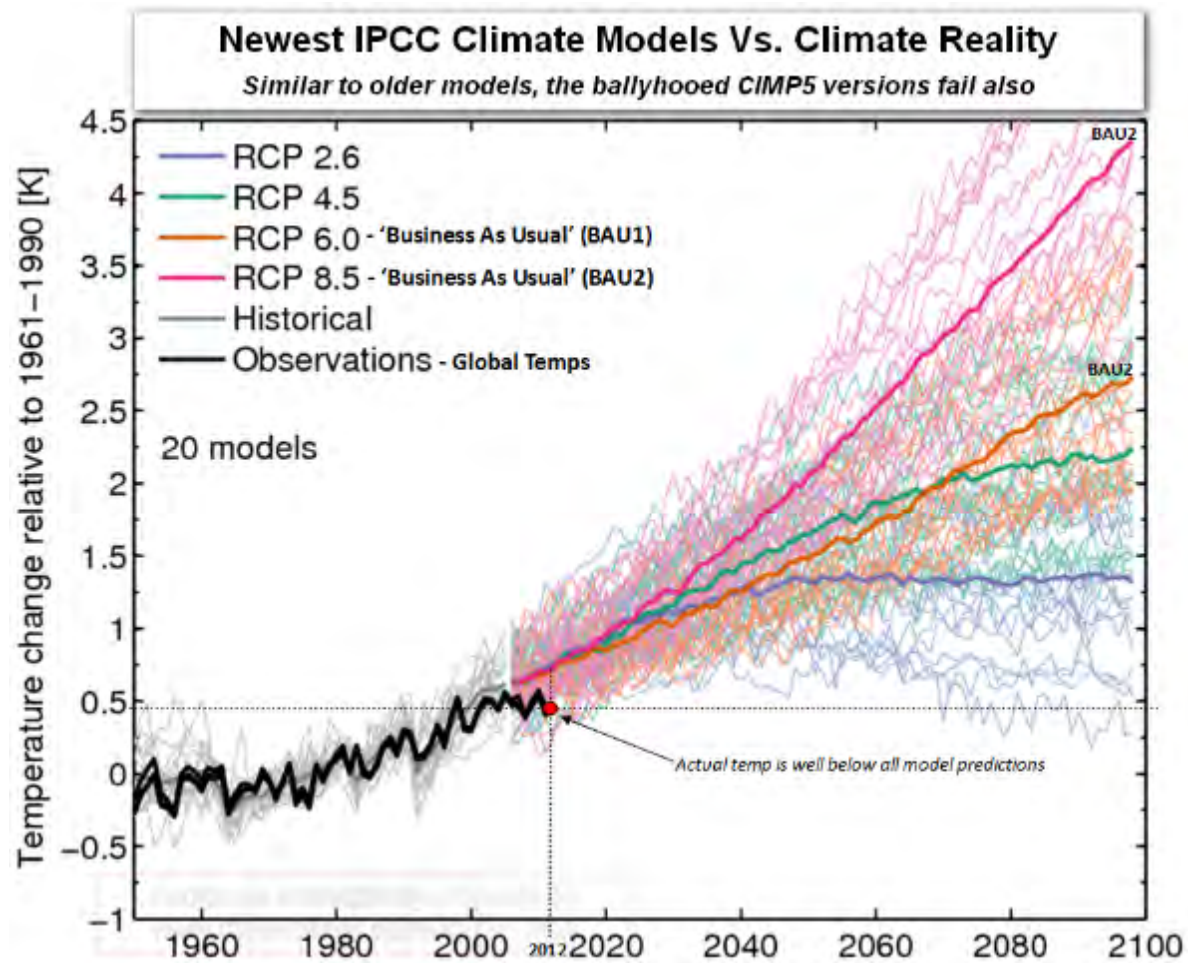
LUDZIE
są niedostępni

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Dlaczego zarządzanie danymi badawczymi jest ważne w naukach o Ziemi i środowisku



Źródło: <http://claesjohnson.blogspot.com/>





Charakterystyka danych w naukach o Ziemi i środowisku

MEiN vs Organizacja Współpracy Gospodarczej i Rozwoju (OECD)

Geonauki multidyscyplinarne

Mineralogia

Paleontologia

Geochemia i geofizyka

Geografia fizyczna

Geologia

Wulkanologia

Nauki o środowisku (aspekty społeczne należą do Geografia społeczna i gospodarcza)

Meteorologia i nauki o atmosferze

Badania klimatyczne

Oceanografia, hydrologia, zasoby wodne





Charakterystyka danych w naukach o Ziemi i środowisku

Przykład różnego poziomu danych i ich processingu (na przykładzie NASA EOSDIS)

Poziom 0	Dane poziomu 0 to nieprzetworzone surowe dane instrumentu w rozdzielczości przestrzennej i czasowej instrumentu. Wszystkie dostępne informacje uzupełniające potrzebne do dalszego przetwarzania są zawarte w plikach danych, dołączone do nich lub dostarczane z nimi.
Poziom 1	Dane poziomu 1 pochodzą z danych poziomu 0, które zostały przetworzone na wartości jednostek czujnika, np. temperatury jasności, radiancje lub współczynniki rozproszenia wstecznego. Dane poziomu 1 są dostępne w pełnej rozdzielczości czasowej/przestrzennej instrumentu. Wszystkie dodatkowe informacje potrzebne do dalszego przetwarzania są zawarte w plikach danych, dołączone do nich lub dostarczane z nimi. Dane poziomu 1 mogą już być geolokalizowane.
Poziom 2	Dane poziomu 2 pochodzą z danych poziomu 1, które zostały przetworzone do wielkości geofizycznych będących przedmiotem zainteresowania, np. wysokość fali, frakcja chmur, głębokość optyczna aerozolu lub promieniowanie na szczycie atmosfery. Oszacowania niepewności wyszukanych parametrów są włączane do plików danych, dołączane do nich lub dostarczane wraz z nimi.



Charakterystyka danych w naukach o Ziemi i środowisku

Przykład różnego poziomu danych i ich processingu (na przykładzie NASA EOSDIS)

Poziom 3	Dane poziomu 3 pochodzą z danych poziomu 2, które zostały ponownie próbkowane przestrzennie i/lub czasowo, np. średnie miesięczne na globalnej siatce szerokości geograficznej $1^{\circ} \times 1^{\circ}$. Oszacowania niepewności ponownie próbkowanych danych są zawarte w plikach danych.
Poziom 4	Dalsze produkty pochodne, np. wskaźniki klimatyczne, uzyskane z analizy wielu produktów niższego rzędu. Oszacowania niepewności, jeśli są możliwe do obliczenia, mogą być zawarte w plikach danych.





Zasady FAIR

Protokół pomiarów
Opis metadanych
Trwałe identyfikatory



Odnajdywalne
(Findable)

Dostępne
(Accessible)



Dokumentacja danych
M2M
Standardy formatu danych
Uwierzytelnianie i autoryzacja
Repozytoria danych

Interoperacyjne
(Interoperable)



Licencje i prawa
Pochodzenie

Do ponownego
wykorzystania
(Reusable)



life is
not
FAIR,
our
DATA
is





Zasady otwierania danych badawczych

Dane powinny być tak otwarte, jak to możliwe i na tyle zamknięte, na ile to jest konieczne.

[jęz. ang.: as open as possible, as closed as necessary]

Potencjalne powody ograniczenia dostępu do danych:

- **Polityka obronna kraju (Oceanografia, Hydrologia);**
- **Ochrona strategicznych zasobów naturalnych (Geologia);**
- **Polityka wewnątrz firmy (Geologia);**
- **Dane wrażliwe (geografia społeczna i gospodarcza).**



Plan Zarządzania Danymi

Ogólne Informacje



Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku



Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Dlaczego warto zarządzać danymi: perspektywa badacza

Zarządzaj swoimi danymi samodzielnie:

- Utrzymuj porządek – znajdź swoje pliki (dane wejściowe, skrypty analityczne, wyjścia na różnych etapach procesu analitycznego itp.).
- Śledź swoje procesy naukowe pod kątem odtwarzalności – bądź w stanie dopasować swoje wyniki do dokładnych danych wejściowych i transformacji, które je wygenerowały.
- Lepsza kontrola wersji danych – łatwo identyfikuj wersje, które można okresowo czyścić.
- Bardziej wydajnie kontroluj jakość swoich danych.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Dlaczego zarządzanie danymi: perspektywa badacza

- Aby uniknąć utraty danych (np. tworzenie kopii zapasowych).
- Sformatuj swoje dane do ponownego wykorzystania (przez Ciebie lub inne osoby).
- Bądź przygotowany: Udokumentuj swoje dane do własnego przypomnienia, odpowiedzialności i ponownego wykorzystania (przez siebie lub innych).
- Zyskaj wiarygodność i uznanie dla swoich wysiłków naukowych dzięki udostępnianiu danych!

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Dlaczego zarządzanie danymi: podstawa do postępu naukowego

- Dane są cennym zasobem – ich gromadzenie jest kosztowne i czasochłonne.
- Danymi należy zarządzać, aby:
 - zmaksymalizować efektywne wykorzystanie i wartość zasobów danych i informacji;
 - ciągle doskonalenie jakości, w tym: dokładności danych, integralności, integracji, terminowości pozyskiwania i prezentacji danych, przydatności i użyteczności;
 - zapewnić właściwe wykorzystanie danych i informacji;
 - ułatwić udostępnianie danych;
 - zapewnić trwałość i długoterminową dostępność do ponownego wykorzystania w nauce.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Dobrze zarządzane dane mogą skutkować ponownym wykorzystaniem, integracją i nową nauką!

eBird



Land Cover



Meteorology



MODIS – Remote sensing data



$$F(X,s,t) = \frac{1}{n(s,t)} \sum_{i=1}^m f_i(X,s,t) I(s,t \in \theta_i)$$

Spatio-Temporal Exploratory Models predict the probability of occurrence of bird species across the United States at a 35 km x 35 km grid.

Model results

Occurrence of Indigo Bunting (2008)



Potential Uses-

- Examine patterns of migration
- Infer impacts of climate change
- Measure patterns of habitat usage
- Measure population trends

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

Sposób pozyskiwania i opracowywania nowych danych i/lub ponownego wykorzystania dostępnych danych

The glaciological data will be obtained from <http://wgms.ch/> and the [university] archive.

Meteorological data from will be taken from [link to database in PANGAEA].

We will acquire, process and use new [types of data], as well as the archival ones owned by (...) and from external reliable databases and publications.

Professional and licenced software will be used: [list of software]

Documentation will be well described and include metadata.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

Pozyskiwane lub opracowywane dane (np. rodzaj, format, ilość)

Data will be collected and analyzed in standard formats.

Text files and publications will be provided in .txt, .rtf, .doc, .docx, .odt and .pdf formats, data basis and statistics in .xls and .csv, pictures in .jpeg and .tif.

The data will be saved in accordance with the global FAIR standards, i.e. they will be available on scientific platforms; their recording will make them easy to read - the data will be categorized; their format will be universal so that they can be opened in different systems, and their reuse will be possible.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

2. Dokumentacja i jakość danych

Metadane i dokumenty (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych) towarzyszące danym

Documentation will enclose:

- (1) description of the purpose and context of the research**
- (2) methodology, including methods of data collection and analysis, equipment**
- (3) data organization (folders and version structure), relationships between files**
- (4) links to data-related publications**
- (5) data control methods**
- (6) information about data that cannot be shared**
- (7) glossary**

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

2. Dokumentacja i jakość danych

Metadane i dokumenty (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych) towarzyszące danym

It will contain an additional ReadMe.txt file or URL address with an explanation of how to use the data.

Metadata will characterise the research data collection (a dataset), including the following information: a title of the project, a title of data set, an author with ORCID number, a year of creation, keywords, a short description, discipline, a financing/co-financing institution, licence.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

2. Dokumentacja i jakość danych

Stosowane środki kontroli jakości danych

Reliable data sources: peer-reviewed academic papers and publications, databases of the World Glacier Monitoring Service and IGF PAS and others.

Relevant data cleaning process and data quality check in line with completeness, timeliness, correctness, consistency. Only after data quality check and cleansing the data will be presented and stored.

The trustworthiness of the research will be enhanced by the following principles: credibility, transferability, dependability /reliability and confirmability (avoiding the bias of the researchers).

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

The data will be kept on the computers of researchers.

The backup copy will be available on the cloud and the external portable disk.

The access to computers, the external disk and the cloud will be login and password protected.

The data will be organised following FAIR principles to allow for easy transfer to open data repository, including documentation, metadata, folders and txt.read.me files.

The computers are protected by AntiVirus software. The long-term data storage is envisaged in a repository.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

The computer access and the cloud access to data will be log-in and password protected.

PI and the support system will be responsible for data recovery.

Krótki opis procedur w przypadku utraty danych z komputera lub chmury.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

4. Wymogi prawne, kodeks postępowania

Sposób zapewnienia zgodności z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych w przypadku przetwarzania danych osobowych

Sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własnością. Obowiązujące przepisy

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

5. Udostępnianie i długotrwałe przechowywanie danych

Sposób i termin udostępnienia danych

The data will be accessible in open access journals.

It will be also stored for ten years in the Zenodo repository, which is an international, general-purpose, open-data repository developed under the European OpenAIRE program. It provides FAIR data and DOI to datasets.

The information about open access to the dataset of the research will be provided on the website under the research project list of the University.

All data will be available in the Open Data Repository (RepOD) after the publication.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

5. Udostępnianie i długotrwałe przechowywanie danych

Sposób wyboru danych przeznaczonych do przechowania oraz miejsce długotrwałego przechowywania danych (np. repozytorium lub archiwum danych)

The Zenodo repository has been chosen based on the following criteria: the world-wide range, the application of FAIR principles of data management, datasets with prescribed DOI number, free of charge access to data, easy access, usage of Creative Commons Licence. These criteria effectively make the data open.

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

5. Udostępnianie i długotrwałe przechowywanie danych

Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie z danych

Datasets in the Zenodo repository will meet requirements of FAIR data principles. They will be available in Open Access under CC Licences. Most of the data will be provided in the standard formats, which do not require conversion to open formats. However, data collected in the unique formats will undergo conversion to open formats. The re-use of data will neither need purchasing additional software licenses nor additional devices. The data will be accessible through the repository and on-demand.

The stored data can be used for:

(1) knowledge transfer, (2) re-analysis (for validity and trustworthiness of the research), (3) new interpretations by another researcher/s, (4) social utility (for policymakers and public institutions, including governmental and local, regional, metropolitan authorities; NGOs, informal groups, individuals), (5) impact on the research portfolio – for the CV of researchers and citation statistics.

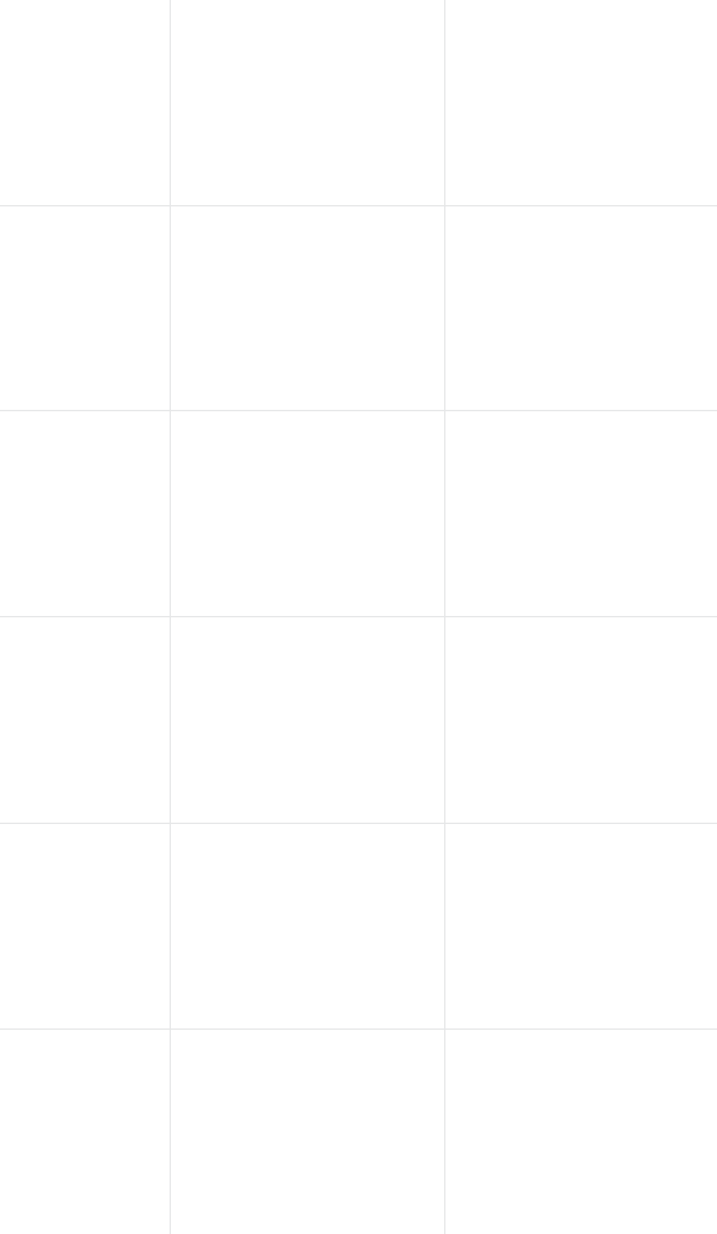
Dlaczego Plan Zarządzania Danymi jest ważny w naukach o Ziemi i środowisku

Przykład z planu zarządzania danymi

5. Udostępnianie i długotrwałe przechowywanie danych

Sposób zapewniający stosowanie unikalnego i trwałego identyfikatora (np. DOI) dla każdego zestawu danych

A unique and permanent identifier – digital ID-number (DOI) will be assigned by the Zenodo repository when the data is deposited.



Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

Opis danych

Przebieg pracy związanej ze zbieraniem danych:

- Zbieranie danych
- Kontrola jakości
- Dokumentacja danych

Opis danych

Sposób pozyskiwania lub wytwarzania nowych danych lub ponownego wykorzystywania danych istniejących.

Dane pierwotne (nowe):

- Pobór prób
- Pomiary terenowe
- Automatyczne stacje pomiarowe
- Nowoczesne metody np. geofizyczne
- Analizy laboratoryjne
- Dane teledetekcyjne

Opis danych

Przykład: Automatyczne Stacje Meteorologiczne (AWS)



Opis danych

Przykład: Automatische Stacje Meteorologiczne (AWS)

- Charakterystyka infrastruktury (WIGOS)
- Certyfikaty i kalibracja (WMO)
- Protokół pomiarowy (WMO)
- Protokół danych (WMO, NetCDF, FAIR data)
- Kontrola jakości (WMO)
- Weryfikacja danych
- Archiwizacja danych
- Udostępnianie danych (dane wejściowe i dane pierwotne)
- Serwis infrastruktury



Opis danych

Pobór prób do datowania radiowęglowego:

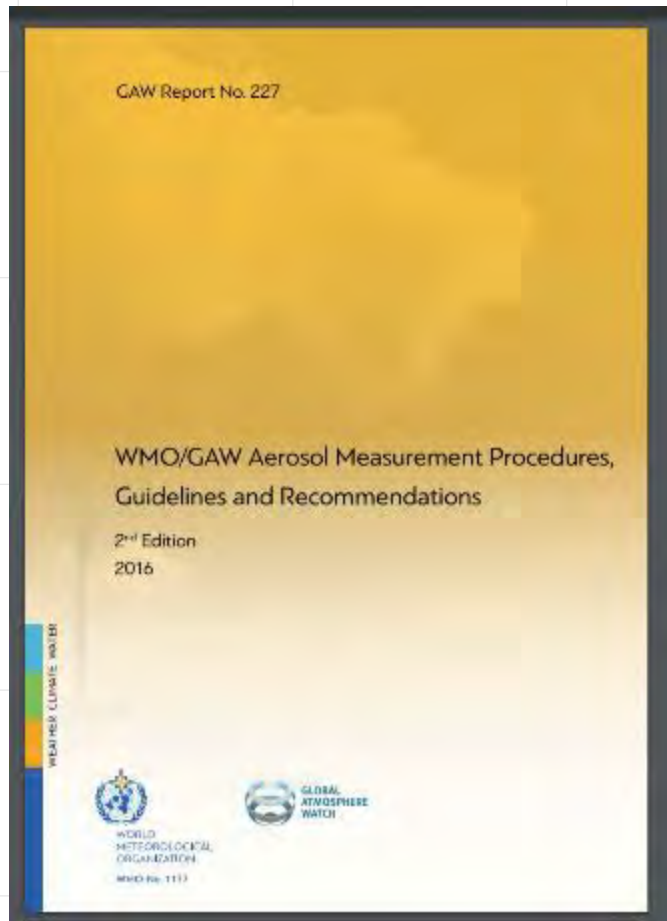
- Lokalizacja
- Opis stanowiska
- Sposób poboru materiału
- Rodzaj materiału, gatunek
- Dokumentacja: opis odsłonięcia, profil stratygraficzny, inne

Kartowanie geomorfologiczne

- Notatki
- Szkice
- Pomiary
- Zestawienie danych w formie szkicu / mapy geomorfologicznej (np. warstwy wektorowe w programach GIS)

Opis danych

Protokoły pozyskiwania danych



ACTRIS Data Centre - an atmospheric data portal

Search | Search Results | GAW Reports | GAW Activities | ACTRIS | GAW (GAW) News | GAW (GAW) News | GAW (GAW) News

Standard Operating Procedures and Measurement Guidelines for ACTRIS In situ aerosol particle variables

Below is a list of WMO/GAW publications and standard operating procedures (SOPs) used for aerosol in situ measurements with ACTRIS. The SOPs provide guidance for good measurement practice. The methodologies, guidelines and SOPs are also recommended within CMDFP and GAW. SOPs and recommendations within CMDFP for other variables are included in ACTRIS's website from the ACTRIS-CCMIP and the <https://data.ccmip.org/>

Component	Category	Description	WMO/GAW Recommendation and Report
AQ, in situ	General guideline	Guidelines on site design for in situ aerosol measurements and station siting	WMO/GAW Recommendation for aerosol in situ siting
AQ, in situ	General guideline	The station audit checklist highlights the points most critical for station operation concerning data quality assurance	Station audit checklist for atmospheric aerosol in situ operations
AQ, in situ	General guideline	Guidelines on conditioning and drying the aerosol sample depending on ambient temperature	WMO/GAW Recommendation for aerosol drying
AQ, in situ	General guideline	Reference for model aerosol variables, normal WMO/GAW Aerosol Measurement procedures, guidelines and recommendations published in 2010	WMO/GAW Report 1227
Cloud condensation nuclei number concentration	SOP	Standard operating procedure for operating ACTRIS cloud condensation nuclei counter	ACTRIS protocol and measurement guidelines on cloud condensation nuclei counter
Wavelength absorption coefficient	Measurement guideline	Reference publication for data processing and quality assurance for the Thermo 2112 Multi-Angle Absorption Photometer (MAAP), including instructions on how to convert the equivalent absorption coefficient reported by the instrument to the property of interest	WMO/GAW Recommendation for aerosol in situ measurements

ICOS Ecosystem Theme Centre

Home | Applications | Publications | Training | News | Events

INSTRUCTION ON CO₂ TURBULENT FLUXES

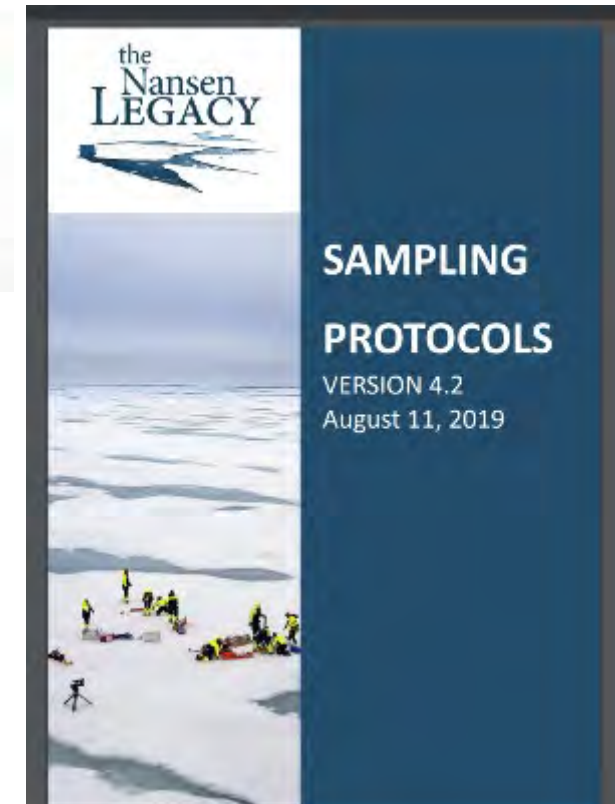
Version 4.2 (2019) - The Global Carbon Project (GCP) has developed this instruction to provide a common framework for CO₂ turbulent flux measurements. It is based on the following documents:

- [WMO \(2011\) The Global Carbon Project \(GCP\) - A common framework for CO₂ turbulent flux measurements](#)
- [Edwards et al. \(2014\) The Global Carbon Project \(GCP\) - A common framework for CO₂ turbulent flux measurements](#)
- [Edwards et al. \(2016\) The Global Carbon Project \(GCP\) - A common framework for CO₂ turbulent flux measurements](#)

and also other relevant documents and reports. The instruction is available in English and Spanish. The instruction document is available at: <https://www.icos-europe.org/~/media/icos-europe/2019/08/01/instruction-on-co2-turbulent-fluxes-v42.pdf>

Instruction document history and download

Version	Change respect to the previous
01 (2011)	Initial version of the instruction, developed by the GCP. It is the first version of the instruction, developed by the GCP. It is the first version of the instruction, developed by the GCP.
02 (2014)	First update of the instruction, based on the GCP. It is the second version of the instruction, developed by the GCP. It is the second version of the instruction, developed by the GCP.
03 (2016)	Second update of the instruction, based on the GCP. It is the third version of the instruction, developed by the GCP. It is the third version of the instruction, developed by the GCP.
04 (2019)	Third update of the instruction, based on the GCP. It is the fourth version of the instruction, developed by the GCP. It is the fourth version of the instruction, developed by the GCP.



Opis danych

pułapki: 2/13, 3/15, 4/107201

Przykład błędnego wprowadzania danych

1	Site	Date	Plot	Species	Weight	Adult	Rodent Trapping 3/15/2010							
2	DeepWell	2/13/2010		1 DIPO	12.1	j	Site	Plot	Adult	RodentSp	Weight			
3	Deep Well	Feb-10		2 Pero	13.22	j	DW		1 y	Pero		12		
4	rioSalado	2/13/2010	1a	pero	16	N	RS		2 j	PERO	escaped	<15		
5	riuSladu	"	1*	CleGap	18.92	gut away	RS		3 n	Clegap		91		
6				Mean1	15.06									
7														
8														
9														
10														
11														
12	Rodent Trapping MJK & ALN 10-Apr-10													
13	Site	Plot	Adult	Species	grams	Comments								
14	deep well		1 y	woodrat										
15	riosalado		2 y	PERO	24.5									
16	riosalado		3 y	Clegap	91									
17														
18														
19														
20														

Niespójność między opisem gromadzonych danych

- Lokalizacja i informacji o dacie
- Niespójny format daty
- Nazwy kolumn
- Kolejność kolumn

Opis danych

Dobre praktyki

- Kolumny danych są spójne: tylko liczby, daty lub tekst
- Spójne nazwy, kody, formaty (data) używane w każdej kolumnie
- Wszystkie dane znajdują się w jednej tabeli, co jest znacznie łatwiejsze w obsłudze programu statystycznego niż wiele małych tabel, z których każda wymaga interwencji człowieka

The image displays two Excel spreadsheets. The top spreadsheet, titled 'data.xls', shows a table with columns A through N. The data includes Site, Date, Plot, Species, Weight, Acult, and Rodent Trapping. The bottom spreadsheet, titled 'SEV_SmallMammalData_v.5.25.2010.xls', shows a table with columns A through H. The data includes Date, Site, Plot, Species, Weight, Adult, and Comments.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Site	Date	Plot	Species	Weight	Acult		Rodent Trapping	3/15/2010					
2	DeepWell	2/13/2010		1 DIPO	12.1	j		Site	Plot	Adult	RodentSp	Weight		
3	Deep Well	Feb-10		2 Pero	13.22	j		DW		1 y	Pero	12		
4	rioSalado	2/13/2010	1a	pero	16	N		RS		2 j	PERO	escaped <15		
5	riuSladu	"	1*	CleGap	18.92	gut away		RS		3 ri	Clegap	91		
6				Mean1	15.06									
7														
8														
9														
10														
11														
12	Rodent Trapping													
13	Site	Plot												
14	deep well													
15	riosalado													
16	riosalado													
17														
18														
19														
20														

	A	B	C	D	E	F	G	H
1	Date	Site	Plot	Species	Weight	Adult	Comments	
2	2/5/2010	Deep Well		1 DIPO	13.2	y		
3	2/4/2010	Deep Well		1 CLEGAP	11.6	j		
4	2/5/2010	Rio Salado		1 DIPO	14.2	y		
5	2/5/2010	Rio Salado		2 PERO	10.1	y		
6	3/15/2010	Deep Well		1 DIPO	15.2	y	plot burned	
7	3/15/2010	Deep Well		2 DIPO	21.7	y	pregnant	
8	3/15/2010	Rio Salado		1 CLEGAP	16.2	j		
9								
10								
11								
12								
13								

Opis danych

Dobre praktyki

Brakujące dane

- Zidentyfikuj brakujące wartości za pomocą dedykowanego kodu.
- Wskaż przyczynę braku wartości w kodzie.
- W polach numerycznych użyj odrębnej wartości, takiej jak 9999, aby wskazać brakującą wartość.
- W polach tekstowych użyj NA („Nie dotyczy” lub „Niedostępne”). (“Not Applicable” or “Not Available”).
- Oddziel te terminy, przydatne podczas interpretacji danych.
- Dodaj oświadczenia dotyczące jakości.
- Ogólne dla zestawu danych.
- Indywidualne na wartość danych np. niepewność.

Opis danych

Dobre praktyki

- Organizuj dane jako kolumny wzdłuż wymiarów w arkuszach kalkulacyjnych
- Wprowadź informacje nagłówka jako pary klucz-wartość
- Dołącz informacje o ważnych rekordach danych w pliku

Użyj opisowych nazw kolumn bez spacji i znaków specjalnych

- Gleba T30 powinna mieć wartość:
Gleba_Temp_30cm
- Kod gatunku powinien mieć postać: Kod_gatunku
- Unikaj używania -,+,*,^ w nazwach kolumn.
- Niektóre programy mogą interpretować te symbole jako operator

Użyj opisowej nazwy pliku

SEV_SmallMammalData_v_20100510.csv

- Projekt (SEV),
- teamt (SmallMammalData)
- Data utworzenia (v_20100510).

VS
mydata.xls

Dodatkowy komentarz:

używaj formatu daty zgodnej z ISO8601 (np. 20100105)

Czego unikać?

Lab ID	Locality	Lat. N	Long. E	Elevation (m a.s.l.)	Collection year	Dated material	Age ¹⁴ C (BP)	Calibrated ages BP (Marine20, ΔR = -56 ± 33)			References
								Median	68.3% Confidence interval	95.4% Confidence interval	
U-2753	Treskelen	N/A	N/A	5–10	1974	<i>Mya truncata</i>	9690 ± 160	10524	10725–10245 cal. BP (68.3%)	11061–10116 cal. BP (95.4%)	Birkenmajer, Olsson (1997)
U-2757	Treskelen	N/A	N/A	5–10	1974	<i>Chlamys islandica</i>	8770 ± 120	9314	9475–9175 cal. BP (65.1%) 9170–9151 cal. BP (3.1%)	9618–8986 cal. BP (95.4%)	Birkenmajer, Olsson (1997)
U-2755	Treskelen	N/A	N/A	5–10	1974	<i>Hiatella arctica</i>	8400 ± 140	8849	9052–9046 cal. BP (0.7%) 9026–8620 cal. BP (67.5%)	9261–8457 cal. BP (95.4%)	Birkenmajer, Olsson (1997)
U-2833	Treskelen	N/A	N/A	5–10	1974	<i>Astarte borealis</i>	4280 ± 170	4278	4519–4018 cal. BP (68.3%)	4779–3822 cal. BP (95.4%)	Birkenmajer, Olsson (1997)
U-2763	Selodden	N/A	N/A	5	1974	<i>Mya truncata</i>	9770 ± 150	10637	10871–10382 cal. BP (68.3%)	11101–10220 cal. BP (95.4%)	Birkenmajer, Olsson (1997)
U-2969	Treskelen	N/A	N/A	5–20	1974	<i>Chlamys islandica</i>	9750 ± 160	10609	10851–10338 cal. BP (68.3%)	11107–11107 cal. BP (0.2%) 11089–10117 cal. BP (95.3%)	Birkenmajer, Olsson (1997)
U-2971	Treskelen	N/A	N/A	5–20	1974	<i>Mya truncata</i>	9610 ± 140	10412	10619–10612 cal. BP (0.8%) 10595–10201 cal. BP (67.5%)	10877–10001 cal. BP (95.2%) 9985–9974 cal. BP (0.2%)	Birkenmajer, Olsson (1997)
U-2985	Treskelen	N/A	N/A	5–20	1974	<i>Astarte borealis</i>	1850 ± 150	1309	1499–1492 cal. BP (1.0%) 1483–1127 cal. BP (67.3%)	1669–1658 cal. BP (0.4%) 1651–956 cal. BP (95.0%)	Birkenmajer, Olsson (1997)
Gd-3178	Treskelen	N/A	N/A	90–100	N/A	<i>Chlamys islandica</i>	8680 ± 60	9219	9336–9102 cal. BP (66.3%)	9424–9005 cal. BP (95.4%)	Marks, Pękala (1986)
GIN-7083	Treskelen	N/A	N/A	5–40	1991	<i>Chlamys islandica</i>	8380 ± 80	8820	8974–8688 cal. BP (68.3%)	9095–8541 cal. BP (95.4%)	Tarasov et al. (1992)
GIN-7084	Selodden	N/A	N/A	N/A	1991	<i>Chlamys islandica</i>	7090 ± 200	7449	7649–7245 cal. BP (68.3%)	7872–7012 cal. BP (95.4%)	Tarasov et al. (1992)
Poz-125715	S coast of Hornbukta	77°01'13"	16°36'30"	1.8	2019	<i>Chlamys islandica</i>	9180 ± 50	9823	9950–9680 cal. BP (68.3%)	10093–9576 cal. BP (95.4%)	This study
Poz-125672	S coast of Hornbukta	77°01'13"	16°36'30"	1.8	2019	<i>Hiatella arctica</i>	9840 ± 50	10727	10851–10842 cal. BP (1.7%) 10834–10579 cal. BP (66.5%)	11040–11035 cal. BP (0.2%) 11016–10497 cal. BP (95.2%)	This study
Poz-125671	S coast of Hornbukta	77°01'56"	16°38'44"	1.8	2019	shell debris	9870 ± 50	10776	10927–10909 cal. BP (2.9%) 10900–10889 cal. BP (1.9%) 10883–10630 cal. BP (63.4%)	11055–10550 cal. BP (95.3%) 10542–10539 cal. BP (0.2%)	This study

<https://doi.org/10.25384/SAGE.19658257.v1>

Opis danych

Sposób pozyskiwania lub wytwarzania nowych danych lub ponownego wykorzystywania danych istniejących.

Dane wtórne (istniejące wcześniej):

- Digitalizacja danych analogowych;
- dane historyczne;
- dane udostępnione.

Opis danych

Normy używania danych

Podczas korzystania z danych:

- Uznaj autorów danych
- Bądź odpowiedzialny za dane
- Podziel się tym, czego się nauczyłeś
- Przestrzegaj licencji danych
- Zrozum i przestrzegaj wszelkich ograniczeń lub przepisów



Photo Attribution: <http://soctheory.iheartsociology.com/wp-content/uploads/2011/10/IMG02369-20111028-1602.jpg>

Opis danych

Etyczne używanie danych



Photo Attribution:
<https://www.flickr.com/photos/michaelgallagher/14592386702/>

Zanim udostępnisz swoje dane:

- Zapoznaj się z polityką fundatora/institucji dotyczącą udostępniania danych
- Usuń dane osobowe

Korzystanie z danych innej osoby:

- Dać kredyt
- Szanuj licencję
- Odpowiednio chroń dane

Opis danych

Licencje

LICENCJE CREATIVE COMMONS		Prawo do kopiowania, rozpowszechniania, przedstawiania i wykonywania	Obowiązek oznaczenia autora oryginału	Użycie w celach komercyjnych	Prawo do tworzenia utworów zależnych*	Prawo do zmiany licencji dla utworu zależnego*
 DOMENA PUBLICZNA		✓	✗	✓	✓	✓
 CC BY		✓	✓	✓	✓	✓
 CC BY-SA		✓	✓	✓	✓	✗
 CC BY-ND		✓	✓	✓	✗	✓
 CC BY-NC		✓	✓	✗	✓	✓
 CC BY-NC-SA		✓	✓	✗	✓	✗
 CC BY-NC-ND		✓	✓	✗	✗	✓

 Możesz kopiować, publikować, przedstawiać, wykonywać itp.

 Musisz podać autora oryginału

 Możesz zarabiać na utworze

 Możesz utwór modyfikować, tj. tworzyć utwory zależne

 Możesz wybrać dowolną licencję dla utworu zależnego, jaki stworzysz

* utwór zależny = np. tłumaczenie, przeróbka, adaptacja itp.

Autor: by Foter, modyfikacja by: ASK



<https://creativecommons.org/licenses/>

<https://archiwum.nauka.umw.edu.pl/otwarta-nauka-licencje-cc>

Opis danych

Prawa autorskie a licencja

- Prawa autorskie: „Zbiór praw wyłącznych przyznanych przez prawo właścicielom praw autorskich w celu ochrony ich dzieła” (Urząd ds. Praw Autorskich Stanów Zjednoczonych).
- Fakty i dane* nie mogą być chronione prawem autorskim.
- Metadane i układ danych mogą być chronione (czasami).
- Licencja: określa, co można zrobić z danymi i jak te dane mogą być redystrybuowane (CC)

*Granica między danymi „czysto opartymi na faktach” a danymi opisowymi jest niewyraźna i czasami może być subiektywna. W niektórych przypadkach, takich jak zbiór danych o różnorodności biologicznej, wiele danych można uznać za czysto faktyczne (np. nazwa miejsca, nazwa gatunku), podczas gdy inne dane w ramach tego samego zbioru danych mogą zawierać opisy miejsca, które, chociaż które mają charakter faktyczny, są w rzeczywistości twórczą treścią osoby(osób) dokonującej(ych) obserwacji.

Opis danych – format danych

Większość projektów (słusznie) koncentruje się na zawartości swoich plików danych:

musisz również wziąć pod uwagę format.

Ponieważ TY utworzyłeś/przechujesz dane w SWOICH WŁASNYCH plikach, **Ty wiesz:**

- w jaki sposób dane są zorganizowane
- jak je czytać
- jak z nich korzystać
- znasz cechy danych, które mogą ograniczać ich użycie

Celem dobrego formatu danych jest **ułatwienie innym** odczytywania danych.



Opis danych – format danych

Zawsze wybieraj **otwarte formaty plików**

- Pamiętaj, że dane mają być przetwarzane w perspektywie 50-100 lat.
- Używaj samoopisujących się formatów.
- Nie bądźiesz w pobliżu, aby odpowiadać na pytania w nieskończoność.

Zawsze używaj **ISO8601** dla formatu daty

- Zawsze preferuj podejście rok, miesiąc, data, czas, szczególnie w nazwach plików.

Wybierz formaty, które:

- zezwalają na przesyłanie strumieniowe danych,
- pozwalają na użycie ogólnego modelu danych.



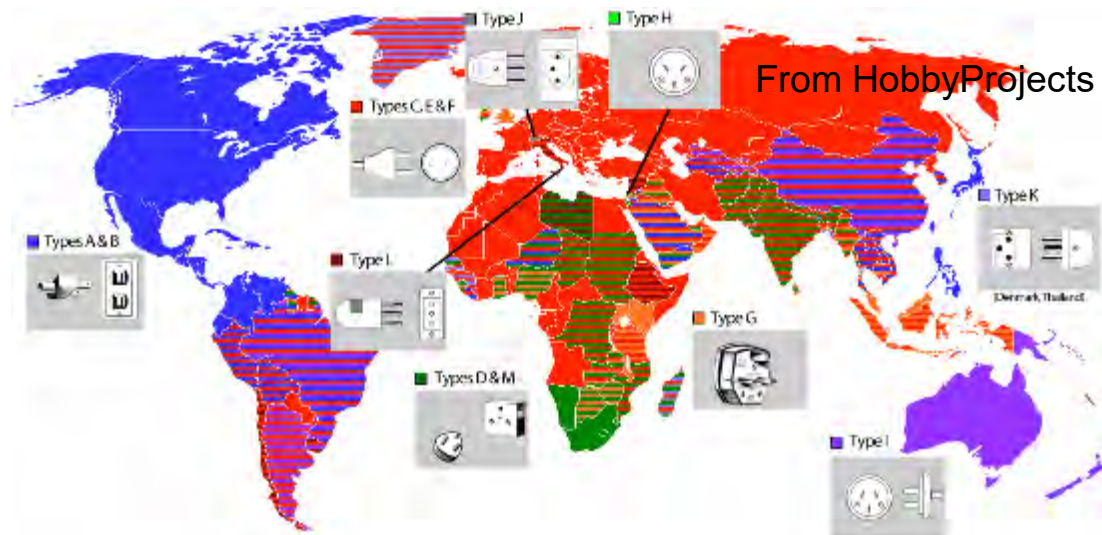
Typy formatu danych

Opis danych – format danych

Ułatwia życie!

Promuje wspólne rozumienie treści

- Poprawia wydajność
- Zmniejszyć ryzyko
- Promuje zrównoważony rozwój
- Zachęcaj do innowacji
- Redukcja kosztów
- Poprawa jakości



HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

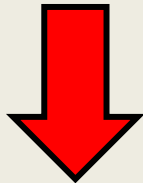


- Format danych

NetCDF

Samo opisujące się dane

Może być używany do strumieniowania danych (np. **OPeNDAP**)



Pobieranie danych nie jest potrzebne

Standardowe tworzenie:

- Dane siatkowe
- Serie czasowe
- Trajektorie (& serie czasowe)
- Profile (& serie czasowe)

Może być **zintegrowany z**

- R
- Matlab
- Ferret
- Python
- Excel
(ex. NETCDF4Excel on GitHub)

Może zawierać każde dane!

Dokumentacja i jakość danych



Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

Co to są metadane?

Metadane to: Dane „raportujące”

KTO stworzył dane?

JAKA jest treść danych?

KIEDY dane zostały utworzone?

GDZIE to jest geograficznie?

W JAKI SPOSÓB opracowano dane?

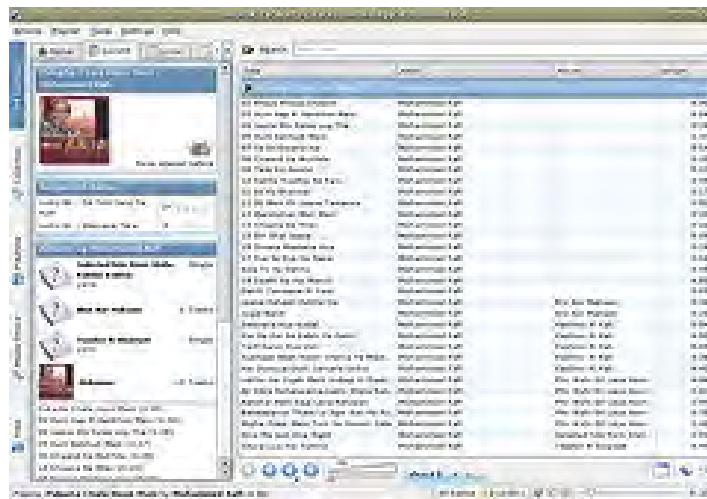
DLACZEGO dane zostały opracowane?



Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

Metadane są wokół nas!



Author(s) Boulosa, Carmen.
Title(s) They're cows, we're pigs /
by Carmen Boulosa
Place New York : Grove Press, 1997.
Physical Descr viii, 180 p ; 22 cm.
Subject(s) Pirates Caribbean Area Fiction.
Format Fiction

Nutrition Facts	
Serving Size 4 OZ. SERVING (112g) Servings Per Container VARIED	
Amount Per Serving	
Calories 170	Calories from Fat 70
% Daily Value*	
Total Fat 8g	12%
Saturated Fat 3g	15%
Cholesterol 65mg	22%
Sodium 70mg	3%
Total Carbohydrate 0g	0%
Dietary Fiber 0g	0%
Sugars 0g	
Protein 23g	
Vitamin A 0%	• Vitamin C 0%
Calcium 0%	• Iron 15%
*Percent Daily Values are based on a 2,000 calorie diet.	

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

Materiał wyjściowy i dane pierwotne muszą być opisane w sposób pozwalający na zidentyfikowanie osoby prowadzącej badania i osób lub instytucji odpowiedzialnych za zebranie materiału wyjściowego oraz danych, a także za analizę wyników końcowych.

„Brak takich danych, w przypadkach postępowania w sprawie zarzutu nierzetelności naukowej, powinien być traktowany jako okoliczność obciążająca”¹⁾

¹⁾ Ministry of High Education and Science, Danish Code of Conduct for Research Integrity, listopad 2014

Kodeks Narodowego Centrum Nauki dotyczący rzetelności badań naukowych i starania o fundusze na badania

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

Typ	Cel	Opis	Przykład
Discovery metadata	Służy do wyszukiwania odpowiednich danych	Metadane Discovery są również nazywane metadanymi indeksowymi i stanowią cyfrową wersję karty indeksowej biblioteki. Opisuje, kto co zrobił, gdzie i kiedy, jak uzyskać dostęp do danych i potencjalne ograniczenia dotyczące danych. Zawiera również łącze do dalszych informacji na temat danych, takich jak metadane witryny. Metadane wykrywania są zatem metadanymi WIS/.	ISO19115 GCMD DIF ACDD
Use metadata	Służy do zrozumienia znalezionych danych	Użyj metadanych, które opisują rzeczywistą zawartość zestawu danych i sposób jego kodowania. Celem jest umożliwienie użytkownikowi zrozumienia danych bez dalszej komunikacji. Opisuje zawartość zmiennych za pomocą ustandaryzowanych słowników, jednostek zmiennej, kodowania braków danych, odwzorowań map itp.	Climate and Forecast Convention BUFR GRIB DwCA

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

Typ	Cel	Opis	Przykład
Configuration metadata	Służy do dostrajania usług portalu pod kątem zestawów danych dla użytkowników.	Metadane konfiguracyjne służą do ulepszania usług oferowanych społeczności użytkowników za pośrednictwem portalu. Może to być np. jak najlepiej zwizualizować produkt	
Site metadata	Służy do zrozumienia znalezionych danych	Metadane witryny służą do opisywania kontekstu danych obserwacyjnych. Opisuje lokalizację obserwacji, oprzyrządowanie, procedury itp. Do pewnego stopnia pokrywa się z metadanymi wykrywania, ale w większym stopniu naprawdę rozszerza metadane wykrywania. Metadane witryny mogą być wykorzystywane do projektowania sieci obserwacyjnych.	WIGOS OGC O&M

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

**Co powoduje, że metadane są poprawne i dobre?
Używanie standardów!**

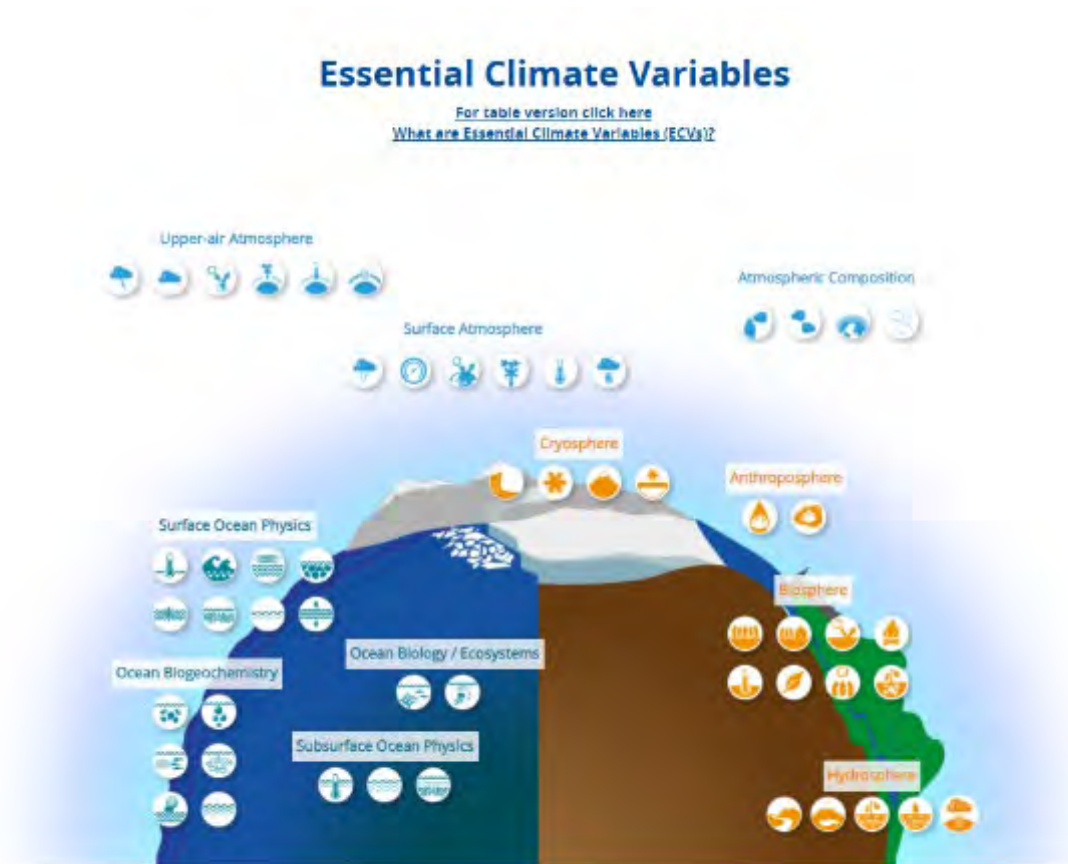
- Global Change Master Directory
- Geographic Names Information System
- Getty Thesaurus of Geographic Names
- ISO 19115 Topic Category Thesaurus



Photo by mxgirl2014 on flickr

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku
GCOS Essential Climate Variable (ECV)



<https://gcos.wmo.int/en/essential-climate-variables/>

Dokumentacja i jakość

ECV Products and Requirements

These products and requirements reflect the Implementation Plan 2022 ([GCOS-244](#)).

The requirements are found in the complete 2022 ECVs Requirements document as well: [ECV](#)

Products	(*)	Unit	Air Temperature (near surface)
	G		10
Horizontal Resolution	B	km	100
	T		500
	G		-
Vertical Resolution	B		-
	T		-
	G		< 1
Temporal Resolution	B	h	1
	T		3
	G		6
Timeliness	B	h	24
	T		720
	G		0.1
Required Measurement Uncertainty (2-sigma)	B	K	0.5
	T		1
	G		0.01
Stability	B	K/decade	0.05
	T		0.1

ECV Products and Requirements

These products and requirements reflect the Implementation Plan 2022 ([GCOS-244](#)).

The requirements are found in the complete 2022 ECVs Requirements document as well: [ECV Lakes](#).

Products	(*)	Unit	Values	Unit	Values	Unit	Values
Horizontal Resolution	G		-		10		0.1
	B	m	-	m	30	km	1
	T		100		1000		2
Vertical Resolution	G		-		-		-
	B		-		-		-
	T		-		-		-
Temporal Resolution	G		1		5		3
	B	d	30	d		h	24
	T		365		30		240
Timeliness	G		1		5		1
	B	d	30	d		d	30
	T		365		365		365
Required Measurement Uncertainty (2-sigma)	G		5		5		0.1
	B	cm		%		°C	0.3
	T		10				0.6
Stability	G		1		5		0.1
	B	cm/decade		%/decade		°C/decade	
	T		10				0.25

Dokumentacja i jakość



1.9.4 Entities should rely on WMO regulatory and other material for ensuring data quality, such as *Technical Regulations* (WMO-No. 49), Volume III – Hydrology; *Guide to Hydrological Practices* (WMO-No. 168), Volumes I and II; *Manual on Stream Gauging* (WMO-No. 1044), Volumes I and II; and *Guidelines for Hydrological Data Rescue* (WMO-No. 1146), the latter being especially useful for safeguarding long-term series for climate analysis.

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

ISO Topic Category

farming	rearing of animals or cultivation of plants. For example, resources describing irrigation, aquaculture, herding, and pests and diseases affecting crops and livestock.
biota	naturally occurring flora and fauna. For example, resources describing wildlife, biological sciences, ecology, wilderness, sea life, wetlands, and habitats.
boundaries	legal land descriptions.
climatologyMeteorologyAtmosphere	atmospheric processes and phenomena. For example, resources describing cloud cover, weather, atmospheric conditions, climate change, and precipitation.
economy	economic activities or employment. For example, resources describing labor, revenue, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, and exploration and exploitation of resources such as minerals, oil, and gas.
elevation	height above or below sea level. For example, resources describing altitude, bathymetry, digital elevation models, slope, and products derived from this information.
environment	environmental resources, protection, and conservation. For example, resources describing pollution, waste storage and treatment, environmental impact assessment, environmental risk, and nature reserves.
geoscientificinformation	earth sciences. For example, resources describing geophysical features and processes, minerals, the composition, structure and origin of the earth's rocks, earthquakes, volcanic activity, landslides, gravity information, soils, permafrost, hydrogeology, and erosion.
imageryBaseMapsEarthCover	base maps. For example, resources describing land cover, topographic maps, and classified and unclassified images.
intelligenceMilitary	military bases, structures, and activities. For example, resources describing barracks, training grounds, military transportation, and information collection.
inlandWaters	inland water features, drainage systems, and their characteristics. For example, resources describing rivers and glaciers, salt lakes, water use plans, dams, currents, floods, water quality, and hydrographic charts.
location	positional information and services. For example, resources describing addresses, geodetic networks, postal zones and services, control points, and place names.
oceans	features and characteristics of salt water bodies excluding inland waters. For example, resources describing tides, tidal waves, coastal information, and reefs.

https://gcmd.nasa.gov/add/serfguide/iso_topic_category.html

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o Ziemi i środowisku

Global Change Master Directory (GCMD) Keywords

Słowa kluczowe GCMD są zorganizowane przy użyciu wielopoziomowej struktury hierarchicznej. Ta hierarchiczna struktura zapewnia ramy, według których pojęcia mogą być klasyfikowane i powiązane. Zestawy słów kluczowych dotyczących nauk o Ziemi i usług naukowych mają następującą strukturę hierarchiczną:

Kategoria > Temat > Termin > Zmienna > Zmienna szczegółowa
(*Category > Topic > Term > Variable > Detailed Variable*)

<https://earthdata.nasa.gov/earth-observation-data/find-data/idn/gcmd-keywords>
<https://gcmd.earthdata.nasa.gov/static/kms/>

Dokumentacja i jakość

Global Change Master Directory (GCMD) Keywords

Earth Science Keywords Structure
Data Center/Service Provider Keywords Structure
Instrument/Sensor Keyword Structure
Platform/Source Keyword Structure
Location Keyword Structure
Horizontal Data Resolution Keyword Structure
Vertical Data Resolution Keyword Structure
Temporal Data Resolution Keyword Structure
URL Content Type Keyword Structure
Granule Data Formats Keyword Structure
Measurement Names Keyword Structure
Chronostratigraphic Units Types Keyword Structure

Dokumentacja i jakość

Earth Science Keywords Structure

Keyword Level	Example
Category	Earth Science
Topic	Atmosphere
Term	Weather Events
Variable Level 1	Subtropical Cyclones
Variable Level 2	Subtropical Depression
Variable Level 3	Subtropical Depression Track
Detailed Variable	(Uncontrolled Keyword)

<https://earthdata.nasa.gov/earth-observation-data/find-data/idn/gcmd-keywords>

<https://gcmd.earthdata.nasa.gov/static/kms/>

Dokumentacja i jakość

Global Change Master Directory (GCMD) Keywords

Category	Topic	Term	Variable_Level_1	Variable_Level_2	Variable_Level_3	Detailed_Variable	UUID	Case native
Hits: 3552	page_num: 1	page_size: 2000	Keyword Version: 16.0	Revision: 2023-05-11 12:45:05	Timestamp: 2023-05-23 03:20:26	Terms Of Use: https://cdn.earthdata.nasa.gov/conduit/upload/5182/KeywordsCommunityGuide_Baseline_v1_SIGNED_FINAL.pdf	The most up to date XML representations can be found here: https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/?format=xml	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC CHEMISTRY	CARBON AND HYDROCARBON COMPOUNDS	ATMOSPHERIC CARBON DIOXIDE			c3b81888-8a39-4b3f-8033-4c077797bcba	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC CHEMISTRY	CARBON AND HYDROCARBON COMPOUNDS	ATMOSPHERIC CARBON MONOXIDE			88a1b416-1589-45a4-9923-452975ec35c7	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC TEMPERATURE	ATMOSPHERIC TEMPERATURE INDICES	COOLING DEGREE DAYS			2590519a-c2bb-448a-b2f3-d10aaa7e057c	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC TEMPERATURE	SURFACE TEMPERATURE	AIR TEMPERATURE			f634ab55-de40-4d0b-93bc-691bf5408ccb	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC TEMPERATURE	SURFACE TEMPERATURE	BOUNDARY LAYER TEMPERATURE			e9c3b6ca-a534-4f3e-82de-b8b921e8f312	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC TEMPERATURE	SURFACE TEMPERATURE	DEICED TEMPERATURE			6e923275-f9e3-4faf-8a7f-2c96f3d5a280	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC TEMPERATURE	SURFACE TEMPERATURE	DEW POINT TEMPERATURE			0c28d9e4-c848-4628-9c00-45a540707b59	
EARTH SCIENCE	ATMOSPHERE	ATMOSPHERIC TEMPERATURE	SURFACE TEMPERATURE	MAXIMUM/MINIMUM TEMPERATURE			5164162a-60eb-4c94-a0f0-2caaa3bb1754	

Norma PN ISO19115 definiuje metadane w zakresie informacji geograficznej przeznaczone dla celów ogólnych.

Norma PN ISO19115 przeznaczona jest do:

- katalogowania zbiorów danych, wykorzystania w hurtowniach danych oraz pełnego opisu zbiorów danych;
- opisu zbiorów danych geograficznych, serii zbiorów i pojedynczych obiektów geograficznych oraz ich właściwości

Norma definiuje:

- obligatoryjne i warunkowe sekcje metadanych;
- podstawowy zbiór metadanych wymagany do zapewnienia pełnego zakresu zastosowań metadanych (wyszukiwanie danych, określanie przydatności danych, dostęp do danych, transfer danych i wykorzystanie danych cyfrowych);
- fakultatywne elementy metadanych, za pomocą których można szczegółowo opisać dane geograficzne w sposób znormalizowany (normatywny), o ile opis taki jest wymagany;
- metodę rozbudowy metadanych (definiowania nowych metadanych) w celu zaspokojenia specyficznych potrzeb.

Dokumentacja i jakość

Środki kontroli jakości danych:

1. Pozyskanie danych zgodnie z obowiązującymi normami:

Np. Pobór prób wraz z odpowiednim opisem lokalizacji (współrzędne GPS, rodzaj pokrycia terenu, zachowanie reprezentatywności poszczególnych próbek) wykonany zostanie zgodnie z zaleceniami przedstawionymi w Rozporządzeniu Ministra Środowiska z dn. 5.09. 2016 r. poz. 1395 w sprawie sposobu prowadzenia oceny zanieczyszczenia powierzchni ziemi (przy zachowaniu zasad określonych w normach: PN-ISO 10381-1, PN-ISO 10381-2, PN-ISO 10381-3, PN-ISO 10381-4, PN-ISO 10381-5).

Dokumentacja i jakość

Środki kontroli jakości danych:

2. Praca na aparaturze spełniającej odpowiednie standardy jakości:

Np. Badania wykonane zostaną w laboratorium XYZ metodą ICP-MS zgodnie z wytycznymi i normami technicznymi: ISO/IEC 17025 (Testing and calibration laboratories).

Dokumentacja i jakość

Środki kontroli jakości danych:

- 3. Kalibracja terenowych urządzeń pomiarowych (metrologia)**
- 4. Kontrola jakości danych – czynnik ludzki**
- 5. Weryfikacja jakości danych**

Przechowywanie i tworzenie kopii zapasowych podczas badań

Przechowywanie i kopie zapasowe

- Klęski żywiołowe
- Awaria infrastruktury obiektów
- Awaria pamięci masowej
- Awaria sprzętu/oprogramowania serwera
- Awaria oprogramowania aplikacji
- Ludzki błąd
- Złośliwy atak (ludzki lub automatyczny)
- Utrata kompetencji personelu
- Utrata zaangażowania instytucjonalnego
- Utrata stabilności finansowej
- Zmiany oczekiwań i wymagań użytkowników



Przechowywanie i kopie zapasowe

Przechowywanie danych i metadanych – kto jest właścicielem danych?

- Zależy to od umowy między agencją finansującą a instytucją powiązaną z naukowcem
- W przypadku NCN i EU **własność należy do instytucji**, nie do indywidualnego naukowca.



"No, it's MY data!"

Amo

Przechowywanie i kopie zapasowe

Przechowywanie danych i metadanych – kto jest właścicielem danych?

- Naukowcy ponoszą odpowiedzialność za przechowywanie materiałów wyjściowych oraz danych pierwotnych
- Jednostka naukowa powinna posiadać regulacje dotyczące przechowywania materiału wyjściowego i danych pierwotnych.

Prepublication Data Sharing. Opinion, „Nature” nr 461(7261), 10 września 2009, s. 168–170

Przechowywanie i kopie zapasowe

Przechowywanie danych i metadanych

Zasada 3-2-1:

- przechowywać przynajmniej trzy kopie swoich danych (3 x backup)
- przechowywać dwie kopie backupu na różnych urządzeniach (2 x backup)
- z których jedno (1 x backup) znajduje się w innej lokalizacji

Przechowywanie i kopie zapasowe

Przechowywanie danych i metadanych

O czym musimy pamiętać przy przechowywaniu danych:

- bezpieczeństwo danych
- odzyskiwanie danych w przypadku utraty/uszkodzenia
- częstotliwość wykonywania kopii
- przenoszenia danych z urządzeń mobilnych, stanowisk terenowych lub sprzętu domowego na główny serwer w miejscu pracy]

Przechowywanie i kopie zapasowe

W jaki sposób zostanie zapewnione bezpieczeństwo i ochrona danych wrażliwych w okresie trwania projektu?

- Zgodnie z wytycznymi RODO
- Szyfrowane pliki (np. .pdf).
- Kontrola dostępu do danych wrażliwych pomiędzy partnerami – odpowiednie oświadczenia RODO

Wymogi prawne, kodeksy postępowania



Wymogi prawne, kodeksy postępowania

Zapewnienie zgodności z przepisami dotyczącymi danych osobowych oraz ich ochrony (np. geografia społeczna).

Podczas przeprowadzania badań naukowych, **zgoda uczestnika jest kluczowym elementem**, szczególnie gdy badanie wymaga gromadzenia danych osobowych. Takie badania powinny zawsze zapewniać uczestnikom pełną informację o celu badania, jakie dane będą zbierane, jak będą one używane i przechowywane, oraz jakie mają prawa w odniesieniu do swoich danych.

Jeżeli badanie ma na celu generowanie otwartych danych badawczych, to należy to w szczególności uwzględnić w procesie uzyskiwania zgody. Zgoda powinna obejmować informacje, że **dane będą dostępne dla innych naukowców**, jaki jest cel tej otwartości, oraz jakie środki bezpieczeństwa zostaną podjęte, aby zanonimizować dane i chronić prywatność uczestników.

Pamiętaj, że **otwarte dane badawcze powinny być anonimowe**, co oznacza, że nie powinno być możliwe zidentyfikowanie poszczególnych uczestników na podstawie tych danych. W praktyce oznacza to, że osobiste identyfikatory, takie jak imiona, adresy, numery telefonów itp., powinny być usunięte lub zmienione tak, aby nie można było zidentyfikować konkretnej osoby.

Również warto podkreślić, że **zgodę można w dowolnym momencie cofnąć**. Oznacza to, że jeżeli uczestnik zdecyduje, że nie chce już, aby jego dane były używane, te dane muszą być usunięte z badań i nie mogą być dalej wykorzystywane.

W Unii Europejskiej najważniejszym aktem prawnym dotyczącym ochrony danych osobowych jest Ogólne Rozporządzenie o Ochronie Danych (GDPR), które weszło w życie w 2018 roku. GDPR ma na celu ujednoczenie przepisów dotyczących ochrony danych w całej UE i wprowadza szereg wymogów dotyczących przetwarzania danych osobowych.

Wymogi prawne, kodeksy postępowania

Zapewnienie zgodności z przepisami dotyczącymi danych osobowych oraz ich ochrony (np. geografia społeczna).

Pseudonimizacja danych, podobnie jak **anonimizacja**, jest procesem, który ma na celu ochronę tożsamości osób, których dane są używane w badaniu. W przeciwieństwie do anonimizacji, pseudonimizacja pozwala na ponowne powiązanie danych z konkretnymi osobami za pomocą unikalnego identyfikatora, ale tylko przez osobę lub system, który ma dostęp do klucza pseudonimizacji.

Podobnie jak w przypadku anonimizacji, pseudonimizacja **może być stosowana w naukach o Ziemi**, gdy badania zawierają dane, które mogą prowadzić do identyfikacji osób.

Np.

Badania społeczno-ekonomiczne: Jeśli badanie dotyczy wpływu zmian środowiskowych na społeczności ludzkie i wymaga gromadzenia danych osobowych, pseudonimizacja może być stosowana do ochrony tożsamości uczestników.

Dane lokalizacyjne: Jeżeli dane lokalizacyjne mogą prowadzić do identyfikacji osób lub organizacji, pseudonimizacja może być stosowana do zamaskowania tożsamości, jednocześnie umożliwiając naukowcom analizę danych.

Długoterminowe badania: W przypadku badań prowadzonych przez długi czas, gdzie potrzebna jest możliwość powiązania danych z konkretnymi jednostkami (np. rolnikami lub gospodarstwami domowymi), ale jednocześnie chcemy chronić tożsamość tych jednostek, pseudonimizacja może być stosowana.

Wymogi prawne, kodeksy postępowania

Ponowne wykorzystanie danych w naukach o ziemi może być regulowane przez różne przepisy, w zależności od jurysdykcji i specyfiki danych. Przykładem takich przepisów może być dyrektywa INSPIRE, obowiązująca w Unii Europejskiej.

Ochrona danych osobowych jest kluczowym aspektem prawa cyfrowego i jest ściśle regulowana w wielu krajach, w tym w Polsce. Regulacje dotyczące ochrony danych osobowych są zazwyczaj osobnym obszarem prawa, a nie są bezpośrednio uwzględniane w ustawach dotyczących ochrony baz danych.

Ustawa o własności przemysłowej jest jednym z głównych źródeł prawa dotyczącego ochrony praw intelektualnych w Polsce. Obejmuje ona różne formy własności intelektualnej, takie jak patenty, wzory przemysłowe, znaki towarowe, oznaczenia geograficzne i inne.

Jednakże, ustawa ta nie odnosi się bezpośrednio do ochrony danych badawczych jako takich. Ochrona danych badawczych może wynikać z innych przepisów prawnych, na przykład z przepisów o ochronie danych osobowych (jeżeli dane badawcze obejmują dane osobowe), z ustawy o zwalczaniu nieuczciwej konkurencji (jeżeli dane badawcze stanowią tajemnicę przedsiębiorstwa), lub z prawa autorskiego (jeżeli dane badawcze są wynikiem oryginalnej twórczości).

Ponadto, w pewnych okolicznościach, wyniki badań mogą być chronione jako wynalazki patentowane, jeżeli spełniają kryteria patentowalności, takie jak nowość, poziom wynalazczy i możliwość przemysłowego zastosowania. Taka ochrona patentowa jest regulowana przez Ustawę o własności przemysłowej.

W kontekście praw własności przemysłowej, istotną rolę w ochronie danych badawczych odgrywają **regulacje dotyczące tajemnic handlowych i nieujawnionych informacji**, które mogą zawierać wyniki badań i eksperymentów.

Udostępnianie i długotrwałe przechowywanie danych

Udostępnianie i długotrwałe przechowywanie

Udostępnianie – ASAP, ale...

- publikacje (dane powinny być udostępnione w momencie złożenia publikacji, wg polityki NCN dane powinny być udostępnione najpóźniej w dniu opublikowania pracy);
- doktoraty;
- embargo.

Okres przechowywania:

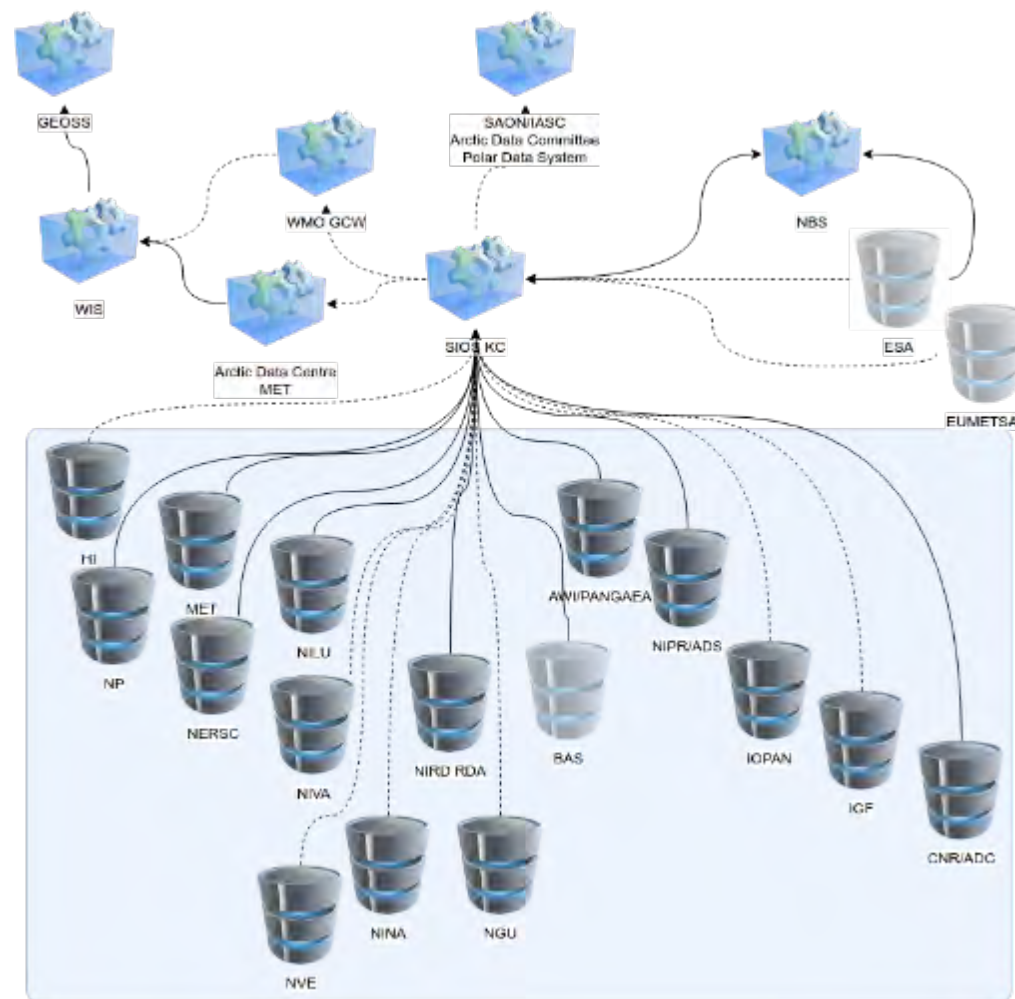
- FAIR data;
- minimum 5 lat¹⁾ (wg polityki NCN – dane przez min. 10 lat, metadane w sposób ciągły, ∞);
- przechowywanie danych opublikowanych, a danych zebranych?

¹⁾ Ministry of High Education and Science, *Danish Code of Conduct for Research Integrity*, listopad 2014



Kryteria wyboru repozytorium

- repozytoria dziedzinowe vs ogólne
- repozytoria o wysokiej reputacji dla dziedziny/dyscypliny
- repozytoria posiadające certyfikaty

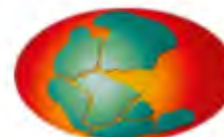
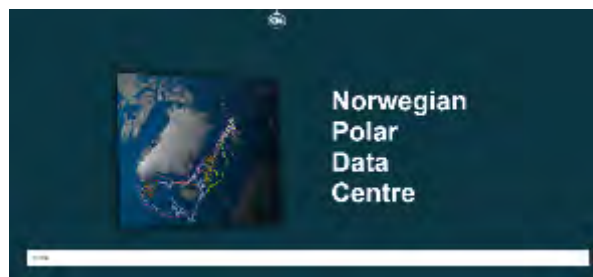




Kryteria wyboru repozytorium



Norwegian Satellite Earth
Observation Database for
Marine and Polar Research



PANGAEA.

Data Publisher for Earth & Environmental Science





Kryteria wyboru repozytorium

Repozytoria danych wymienione przez re3data (<https://www.re3data.org>)

lub\oraz certyfikowane repozytorium danych CoreTrustSeal (<https://amt.coretrustseal.org/certificates>)

1. ZENODO (external data exchange); <https://zenodo.org/> ; logowanie przez ORCID; 50GB na jeden zestaw danych; **DOI**;
2. RepOD; <https://repod.icm.edu.pl/dataverse/us> **DOI**;
3. PANGAEA - Data Publisher for Earth & Environmental Science; **DOI**; <https://www.pangaea.de/>;





Kryteria wyboru repozytorium

6. Repozytorium danych Społecznych RDS; <https://rds.icm.edu.pl/>;
7. MX-RDR –Macromolecular Xtallography Raw Data Repository; <https://mxrdr.icm.edu.pl/>;
8. ArrayExpress - Functional Genomics Data ; <https://www.ebi.ac.uk/biostudies/arrayexpress> ;Dane z wysokoprzepustowego sekwencjonowania;
9. PRIDE - PRoteomics IDentifications Archive; <https://www.ebi.ac.uk/pride/>; Dane z proteomu;
10. GenBank ®; <https://www.ncbi.nlm.nih.gov/genbank/> ; Dane dot. sekwencji DNA, RNA;
11. Open Science Framework (external data exchange); <http://osf.io/>;
12. Harvard Dataverse (external data exchange); <http://dataverse.harvard.edu/>;
13. Dryad Digital Repository (external data exchange); <http://datadryad.org/>;
14. Mendeley Data (external data exchange); <https://data.mendeley.com/>;
15. Figshare (external data exchange); <http://figshare.com/>.



Zadania związane z zarządzaniem danymi oraz zasoby

Zadania związane z zarządzaniem danymi oraz zasoby

Kto będzie odpowiadał za zarządzanie danymi?

Sytuacja wzorcowa:

1. Badacz:

- Opracowanie koncepcji zarządzania danymi (w tym PZD);
- Pozyskanie danych oraz zabezpieczanie ich archiwizacji na pierwszym etapie;
- Zapewnienie kopii zastępczej w czasie trwania pomiarów oraz po ich zakończeniu;

2. Data Steward:

- pomoc w przygotowaniu PZD oraz kontrola jego realizacji.

3. Dział IT:

- Zabezpieczenia miejsca repozycji danych oraz ich bezpieczeństwa.

Zadania związane z zarządzaniem danymi oraz zasoby

Kto będzie odpowiadał za zarządzanie danymi?

Sytuacja wzorcowa:

4. Władze:

- opieka prawna;
- zabezpieczenie techniczne.

Inne:

- format danych;
- koszty w czasie trwania projektu.



Wdrażanie i raportowanie Planów Zarządzania Danymi

Wdrażanie i raportowanie Planu Zarządzania Danymi

- DMP może zmieniać się w trakcie realizacji projektu (zalecane)
- Raport z wdrażania – należy opisać stan faktyczny na koniec realizacji projektu: planowano vs. zrealizowano.

Kontakt:
dariusz.ignatiuk@us.edu.pl



NARODOWE CENTRUM NAUKI

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.



Ministerstwo
Edukacji i Nauki

**Narodowe Centrum Nauki
Zespół ds. Otwartej Nauki
otwarta.nauka@ncn.gov.pl**

