UNIWERSYTET ŚLĄSKI
W KATOWICACH

INTERNATIONAL
LOVE DATA WEEK
FEBRUARY 10-14, 2025

RDM Do's & Don'ts

Bogumiła Gołek

LOVE DATA DAY 2025

10 II 2025

www.us.edu.pl

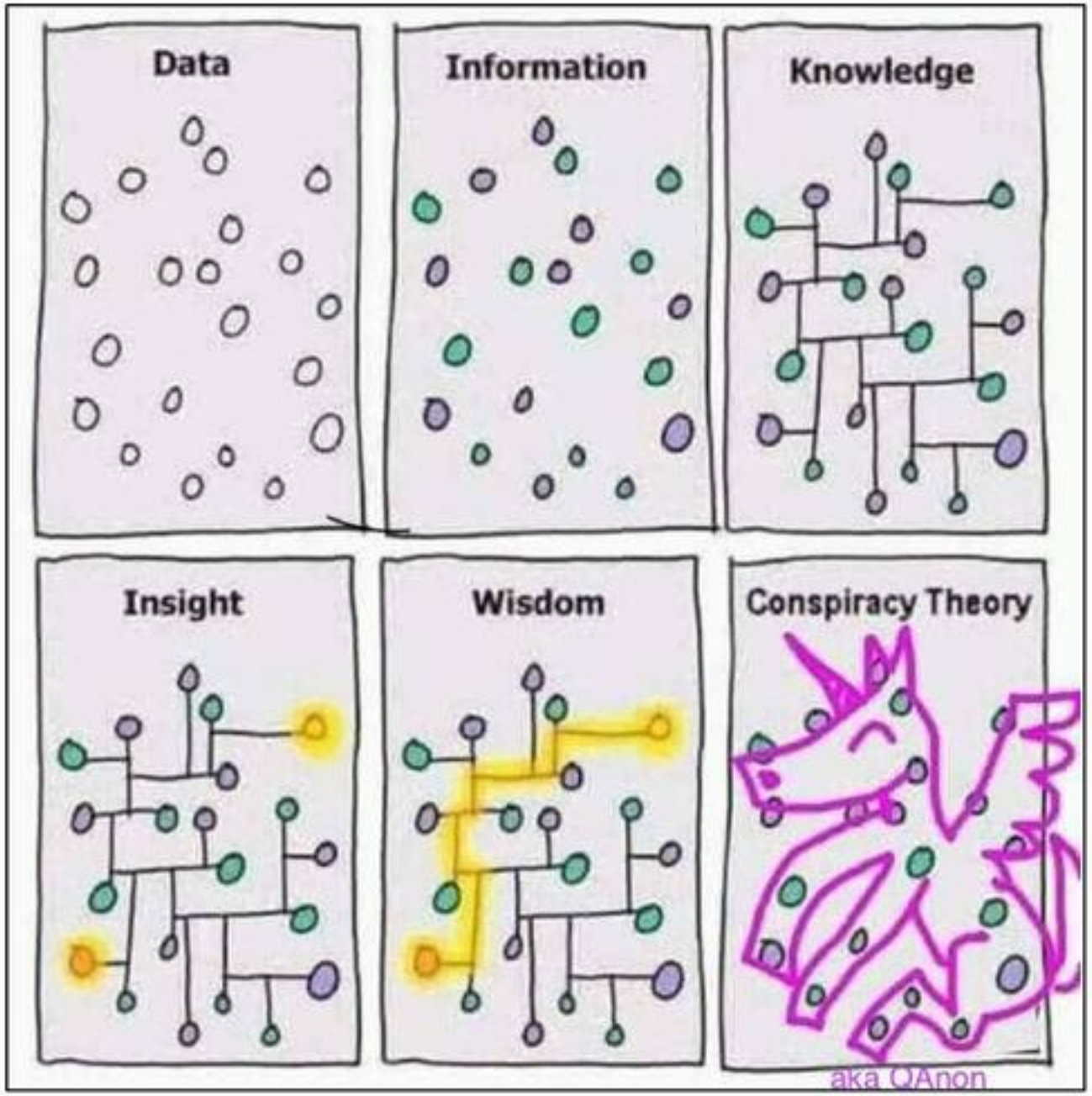
Znaczenie prawidłowego* zarządzania danymi

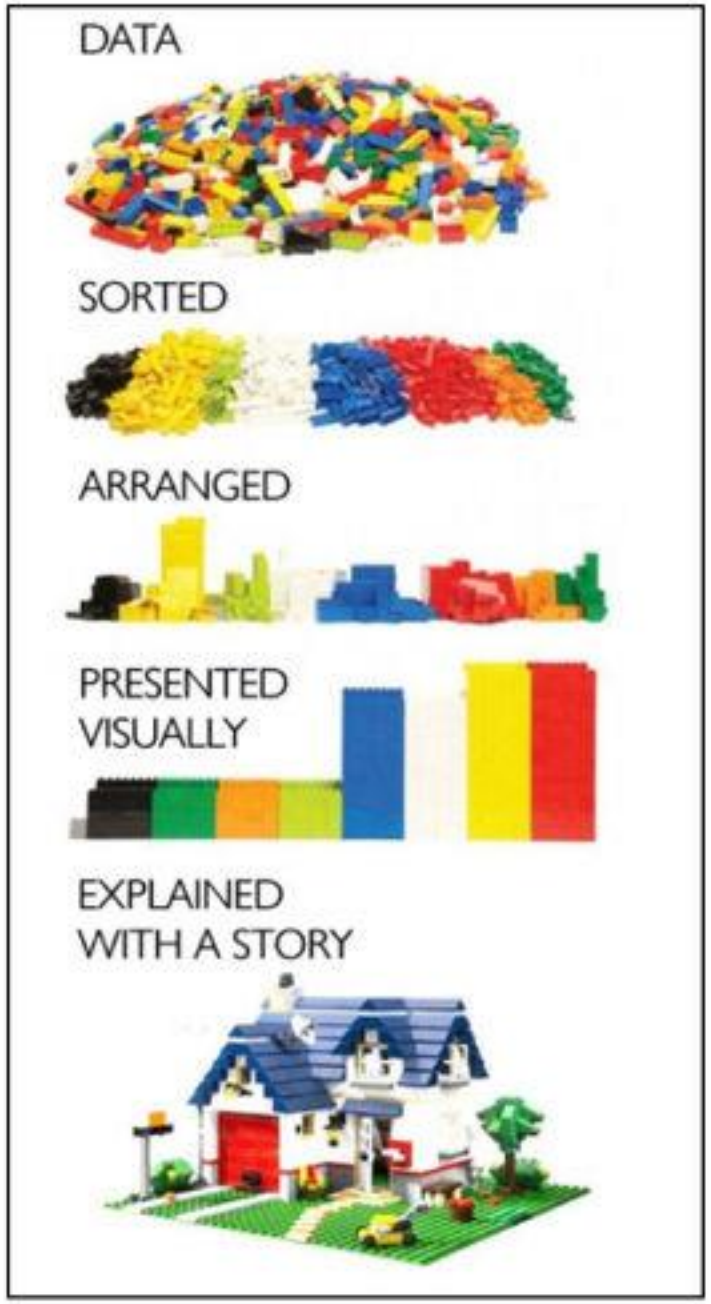
- Ochrona integralności danych i rzetelności badań
- Zgodność z wymaganiami NCN
- Ułatwienie (współ)pracy i ponownego wykorzystania danych

Znaczenie prawidłowego* zarządzania danymi

- Ochrona integralności danych i rzetelności badań
- Zgodność z wymaganiami NCN
- Ułatwienie (współ)pracy i ponownego wykorzystania danych

***stworzenia dobrego i wykonalnego planu**


Data
Information
Knowledge
Insight
Wisdom
Conspiracy Theory
aka QAnon


DATA
SORTED
ARRANGED
PRESENTED VISUALLY
EXPLAINED WITH A STORY

@Ross Waring

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Dewizy RDM 1/2

- **Przejrzysta organizacja i dokumentacja danych** to podstawa – stosowanie standardowych nazw, struktury katalogów oraz metadanych ułatwia zarządzanie i ponowne wykorzystanie danych.
- **Backupy i bezpieczeństwo danych** są kluczowe dla ich długoterminowego przechowywania – wdrażanie strategii 3-2-1 (trzy kopie, dwa różne nośniki, jedna lokalizacja zewnętrzna) minimalizuje ryzyko utraty danych.
- **Udostępnianie danych w odpowiednich repozytoriach** i stosowanie otwartych formatów wspiera otwartą naukę oraz zwiększa widoczność wyników badań.

UNIWERSYTET ŚLĄSKI
W KATOWICACH

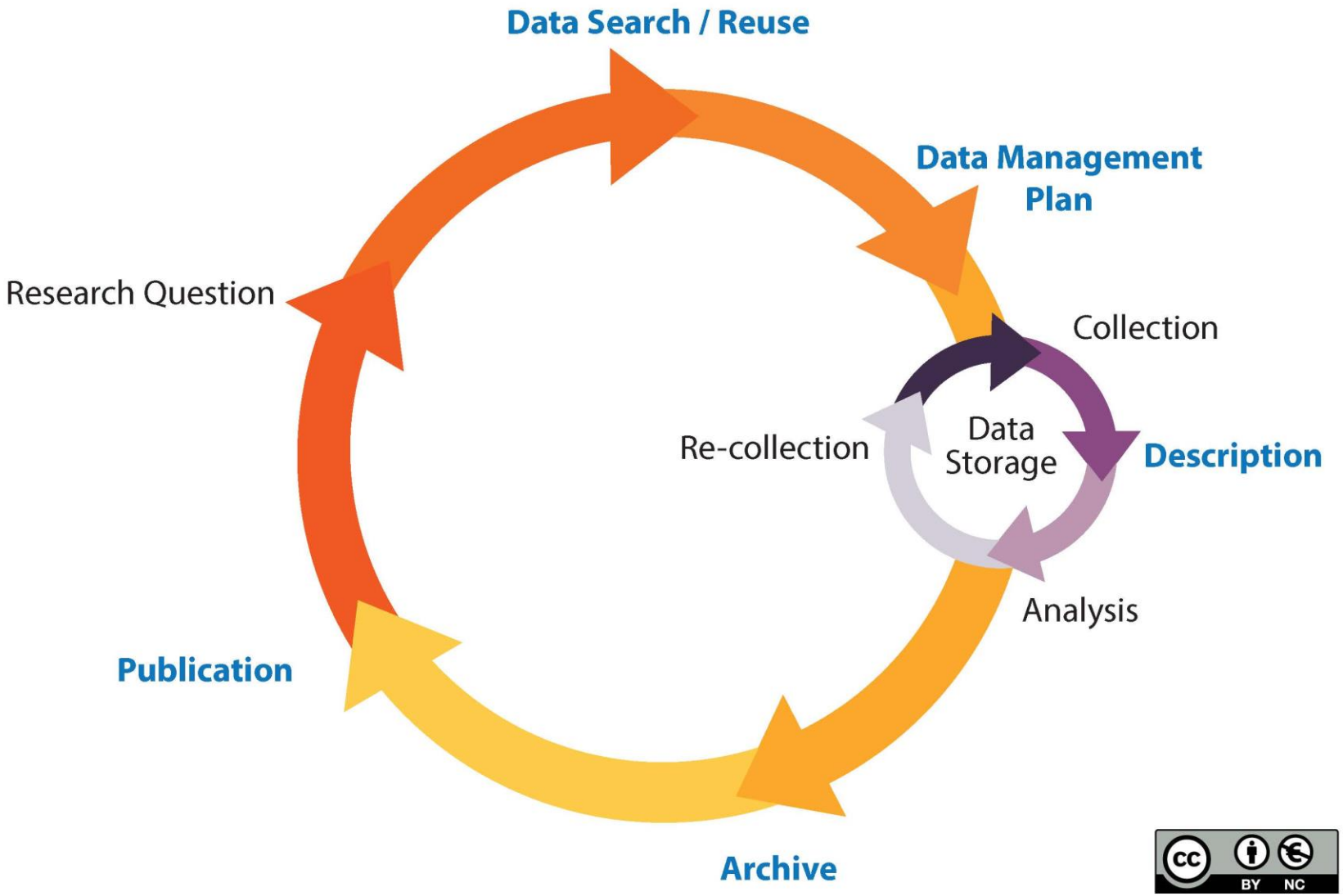
Dewizy RDM 2/2

Bezpieczeństwo danych wymaga odpowiednich zabezpieczeń, takich jak szyfrowanie, ograniczenie dostępu oraz korzystanie z bezpiecznych sieci i metod przechowywania.

Efektywna współpraca w zespole badawczym to przede wszystkim określony podział zadań, systemy kontroli wersji oraz zintegrowane środowiska cyfrowe zapewniające dostęp do aktualnych danych dla wszystkich członków zespołu.

Końcowy raport PZD powinien odzwierciedlać rzeczywiste praktyki zarządzania danymi – warto dbać o spójność dokumentacji i odpowiednie udostępnienie danych do archiwizacji.

The Research Data Management Lifecycle



CC BY NC

@University of South California

The Data Publishing
Life Cycle

@Environmental Data Initiative

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Generowanie danych

✅ **Do’s:**

- **Zapisuj surowe dane w oryginalnym formacie.**
- Chroń oryginalne dane, blokując je lub ustawiając jako tylko do odczytu.
- Odwołuj się do oryginalnych danych, jeśli pojawią się problemy.
- **Dokumentuj** swoje dane.

❌ **Don’ts:**

- Nie nadpisuj oryginalnych danych wersją oczyszczoną.
- Nie zostawiaj pustych komórek – stosuj wartości liczbowe wyraźnie poza zakresem (np. **„99999”** dla brakujących danych, **„88888”** dla danych nie dotyczy) i opisz je w dokumentacji.

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Katalogowanie i organizacja danych

✅ **Do’s:**

- Nadawaj plikom **czytelne nazwy** zawierające podstawowe metadane
- Używaj **spójnych nazw plików i katalogów** (np. „2024_ProjektXYZ_Pomiar1.csv”).
- Stosuj standardowe dla dyscypliny **formaty metadanych** (np. Dublin Core, DataCite).
- Utrzymuj **czytelną strukturę katalogów** (np. „Dane surowe”, „Przetworzone dane”, „Analiza wyników”).
- Twórz **pliki README** z opisem zawartości katalogów i struktury danych.

❌ **Don’ts:**

- Nie używaj nazw typu „Nowy folder (2)”, „dane_final_v3(ostateczne).xlsx”.
- Nie przechowuj niepotrzebnych duplikatów bez oznaczeń wersji.
- Nie pozostawiaj danych nieudokumentowanych (brak opisu kolumn, jednostek miar itp.).

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Przechowywanie i backup danych

✅ **Do’s:**

- **Przechowuj dane na bezpiecznych serwerach instytucjonalnych** (np. OneDrive/GSuite UŚ, NAS).
- Stosuj **co najmniej 3-krotne kopie zapasowe** (np. kopia lokalna, serwerowa, chmura).
- Zabezpieczaj wrażliwe dane poprzez **szyfrowanie i kontrolę dostępu**.
- Regularnie testuj backupy, by upewnić się, że można je przywrócić.

❌ **Don’ts:**

- Nie przechowuj jedynej kopii danych na pendrive'ach, laptopach, zewnętrznych dyskach bez backupu.
- Nie wykonuj kopii zapasowych ani nie przechowuj danych wrażliwych w komercyjnych chmurach (np. Dropbox, Google Drive).
- Nie zapisuj haseł w plikach typu „passwords.txt” – używaj menedżerów haseł.

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Zapewnienie bezpieczeństwa danych

✅ **Do’s:**

- Używaj **silnych haseł i uwierzytelniania wieloskładnikowego** (MFA).
- **Szyfruj dane wrażliwe** (np. dane osobowe, informacje objęte tajemnicą).
- Ogranicz dostęp do danych zgodnie z zasadą **minimalnych uprawnień**.
- Korzystaj z **VPN i zabezpieczonych sieci** do pracy z danymi na odległość.

❌ **Don’ts:**

- Nie przechowuj danych na prywatnych urządzeniach bez szyfrowania.
- Nie udostępniaj plików bez kontroli dostępu – korzystaj z haseł lub przypisz dostęp konkretnym użytkownikom.
- Nie używaj niezabezpieczonych nośników (pendrive'y, niezabezpieczone chmury).

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Współpraca w zespole badaczy

✅ **Do’s:**

- Wybieraj **bezpieczne środowiska** agregowania danych (np. OneDrive/GSuite UŚ, Git, dedykowany serwer).
- Określ jasne **role i odpowiedzialności** w zarządzaniu danymi.
- Dokumentuj zmiany w danych, korzystając z **systemów kontroli wersji** (np. Git, ELN, logbooki, ustalony sposób dokumentacji).
- **Podczas regularnych spotkań** zespołu omówiajcie kwestie danych.

❌ **Don’ts:**

- Nie zapisuj wielu wersji plików w sposób nieczytelny (np. „raport_v2_final_ostateczny.docx”).
- Nie ignoruj komunikacji – brak uzgodnionych procedur prowadzi do chaosu.
- Nie przechowuj plików wyłącznie na prywatnych komputerach członków zespołu.

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Udostępnianie i ponowne wykorzystanie danych

✅ **Do's:**

- Używaj **otwartych formatów plików** (np. CSV, TXT, JSON, ODF) dla większej kompatybilności.
- Udostępniaj dane w **repozytoriach naukowych** (np. tematyczne>dziedzinowe>ogólne,OSF, RepOD, PANGAEA).
- Stosuj **licencje umożliwiające ponowne wykorzystanie** (np. CC-BY, Open Data Commons).
- Określ **warunki dostępu** dla danych ograniczonego dostępu.

❌ **Don'ts:**

- Nie publikuj danych w zamkniętych, niestandardowych formatach (np. .xls zamiast .csv).
- Nie ignoruj kwestii prawnych dotyczących RODO i praw autorskich.
- Nie przechowuj danych badawczych na prywatnych kontach w chmurze (np. Google Drive, Dropbox bez zabezpieczeń).

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Rozliczanie grantu i raport końcowy PZD 

✅ **Do’s:**

- Weryfikuj zgodność z **pierwotnym PZD** i opisz ewentualne zmiany.
- Udostępnij dane w odpowiednich **repozytoriach** (jeśli wymagane przez NCN).
- Opisz, jak były przechowywane i zabezpieczane dane w trakcie projektu.
- Przygotuj notkę nt. **dostępności i ponownego wykorzystania** danych.

❌ **Don’ts:**

- Nie ignoruj wymogu archiwizacji danych przez minimalny wymagany okres (zwykle 10 lat).
- Nie zapominaj o dokumentacji – raport końcowy powinien być spójny z realizacją PZD.
- Nie odkładaj przygotowania raportu na ostatnią chwilę – wymaga analizy i podsumowania.

Last, but not least

✅ **Do’s:**

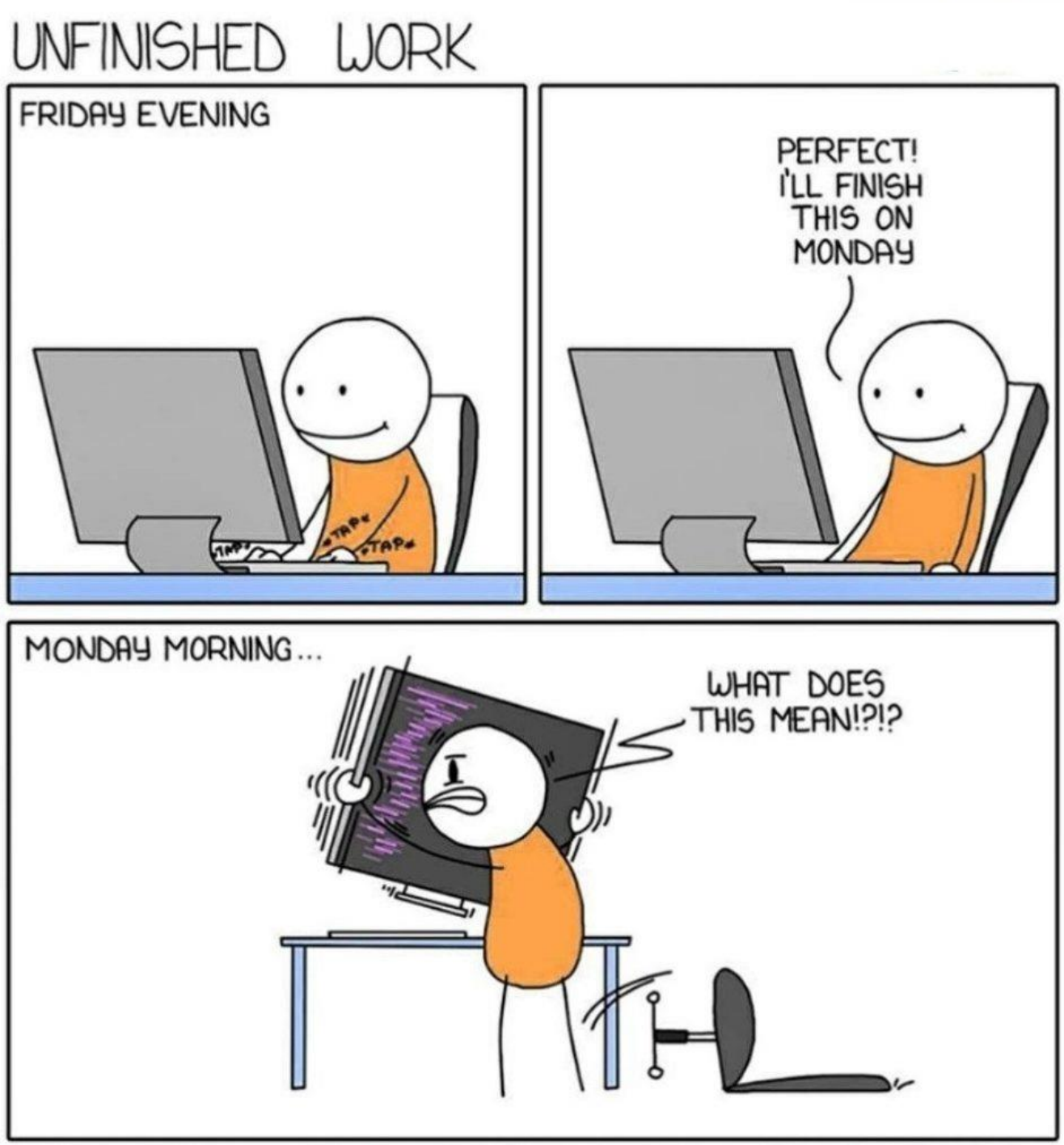
- Bądź konsekwentny i skrupulatny (nie daj się zwariować)
- Obserwuj naukowców wiodących w twojej dziedzinie (gdzie deponują dane, jak je opisują)
- **Rozmawiajcie** między sobą, dzielcie się wiedzą praktyczną i doświadczeniem – horyzontalny i wertykalny(!) transfer wiedzy
- Kto pyta/szuka odpowiedzi, nie błądzi ☺
- Porządek (także w danych) nie dzieje się sam – przeznacz odpowiednie zasoby czasowe na RDM

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Houston, chyba mamy problem...

- czyli o tym jak uniknąć błędów

www.us.edu.pl



@datasciencedojo.com

Brak spójnej struktury katalogów

Chaotycze nazwy i przypadkowe rozmieszczenie plików

Uzgodniona konwencja nazewnictwa i logiczna organizacja folderów (*np. Dane surowe, Analizy, Wyniki*).

Schemat tworzenia nazw plików zapisany w README

Niedostateczna lub bardzo uboga dokumentacja danych

Pliki README i metadane towarzyszące danym – opisujące format danych, ich źródło, jednostki miar i sposób przetwarzania

Protokoły badawcze precyzyjnie dokumentujące proces badawczy

Brak przyjętej strategii backupu

Nieregularne tworzenie kopii zapasowych

Stosowanie zasady 3-2-1: trzy kopie danych, dwa różne nośniki, jedna lokalizacja zewnętrzna

Regularnie testowanie możliwości odzyskania danych

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Nieprzemyślane lub niedbałe udostępnianie danych

Przemyślana struktura folderów
i konwencja nazewnictwa plików

Publikowanie danych
wraz z niezbędną dokumentacją
w repozytoriach naukowych

Licencje zgodne z polityką otwartego dostępu oraz wymogami NCN

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Brak zabezpieczeń podczas przechowywania i przesyłania danych

Szyfrowanie danych, stosowanie uwierzytelniania wieloskładnikowego (MFA) oraz korzystanie z zabezpieczonych serwerów i sieci VPN, określone zasady dostępu do danych

Problemy we współpracy zespołowej

Podział ról i wdrożone procedury dotyczące dostępu i zarządzania danymi

Zintegrowane środowiska i narzędzia kontroli wersji (np. Git, ELN)

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Brak zabezpieczeń podczas przechowywania i przesyłania danych

Szyfrowanie danych, stosowanie uwierzytelniania wieloskładnikowego (MFA) oraz korzystanie z zabezpieczonych serwerów i sieci VPN, określone zasady dostępu do danych

UNIWERSYTET ŚLĄSKI
W KATOWICACH

Kluczowe wnioski 1/2

- **Przejrzysta organizacja i dokumentacja danych** to podstawa – stosowanie standardowych nazw, struktury katalogów oraz metadanych ułatwia zarządzanie i ponowne wykorzystanie danych.
- **Backupy i bezpieczeństwo danych** są kluczowe dla ich długoterminowego przechowywania – wdrażanie strategii 3-2-1 (trzy kopie, dwa różne nośniki, jedna lokalizacja zewnętrzna) minimalizuje ryzyko utraty danych.
- **Udostępnianie danych w odpowiednich repozytoriach** i stosowanie otwartych formatów wspiera otwartą naukę oraz zwiększa widoczność wyników badań.

UNIWERSYTET ŚLĄSKI
W KATOWICACH

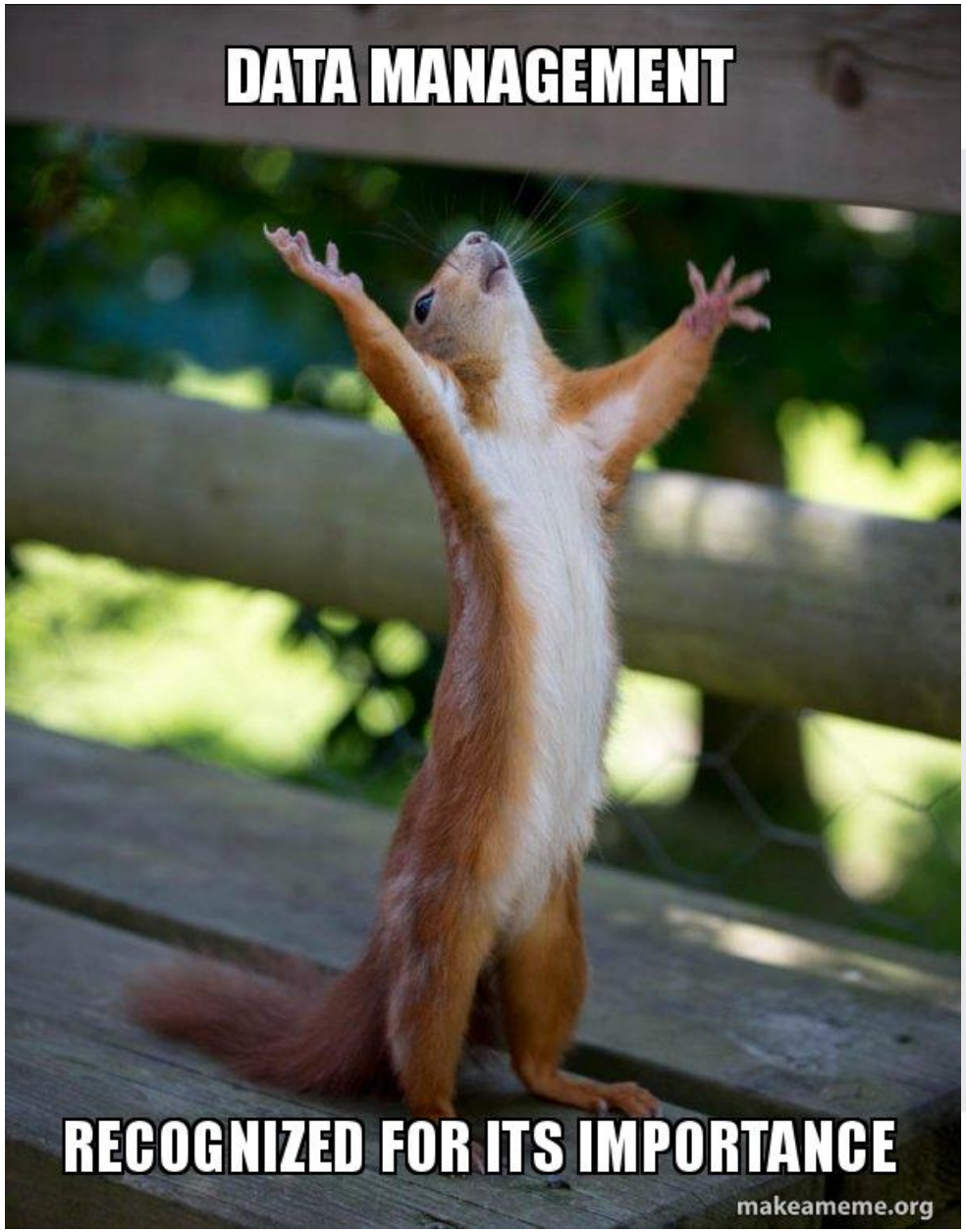
Kluczowe wnioski 2/2

Bezpieczeństwo danych wymaga odpowiednich zabezpieczeń, takich jak szyfrowanie, ograniczenie dostępu oraz korzystanie z bezpiecznych sieci i metod przechowywania.

Efektywna współpraca w zespole badawczym to przede wszystkim określony podział zadań, systemy kontroli wersji oraz zintegrowane środowiska cyfrowe zapewniające dostęp do aktualnych danych dla wszystkich członków zespołu.

Końcowy raport PZD powinien odzwierciedlać rzeczywiste praktyki zarządzania danymi – warto dbać o spójność dokumentacji i odpowiednie udostępnienie danych do archiwizacji.

UNIWERSYTET ŚLĄSKI
W KATOWICACH



Dziękuję za uwagę
☺