

# Dane badawcze w naukach społecznych

dr Danuta Życzyńska-Ciołek, IFIS PAN – [dzyczynska-ciolek@ifispan.edu.pl](mailto:dzyczynska-ciolek@ifispan.edu.pl)  
Magdalena Bielińska, IFIS PAN – [mbielinska@ifispan.edu.pl](mailto:mbielinska@ifispan.edu.pl)  
dr Marcin W. Zieliński, ISS UW – [marcin.zielinski@uw.edu.pl](mailto:marcin.zielinski@uw.edu.pl)



INSTYTUT FILOZOFII I SOCJOLOGII  
POLSKIEJ AKADEMII NAUK



Fundusze  
Europejskie



Rzeczpospolita  
Polska

Unia Europejska  
Europejskie Fundusze  
Strukturalne i Inwestycyjne



# Selekcja i przygotowanie danych społecznych do udostępnienia: SELEKCJA

Selekcja danych  
na różnych etapach  
projektu

Przechowywanie  
krótkoterminowe  
vs.  
długoterminowe

Dane, które / których...

...**trzeba** umieścić  
w repozytorium

...**warto** umieścić  
w repozytorium



...**nie warto** umieszczać  
w repozytorium

...**nie można** umieścić  
w repozytorium

# Selekcja i przygotowanie danych społecznych do udostępnienia: SELEKCJA

Trzeba

Trzeba umieścić w repozytorium (lub innym powołanym do tego miejscu) dane:

- do których przechowywania / udostępnienia jesteśmy formalnie zobowiązani
- które są podstawą publikacji, jeśli wydawca tego wymaga (cele weryfikacyjne)

# Selekcja i przygotowanie danych społecznych do udostępnienia: SELEKCJA

## Warto

Warto umieścić w repozytorium dane, które:

- są dobrej jakości
- są unikatowe lub trudne do zreplikowania
- mają istotną wartość dla dyscyplin(y)
- mają potencjał do tego, by były ponownie wykorzystane
- mogą zostać dobrze opisane i zaopatrzone w materiały kontekstowe
- mogą posłużyć do weryfikacji naszej pracy naukowej

# Selekcja i przygotowanie danych społecznych do udostępnienia: SELEKCJA

Nie warto

Nie warto umieszczać w repozytorium danych, które:

- nie spełniają większości lub żadnego z wcześniej wymienionych kryteriów
- w niektórych przypadkach: przejściowych, roboczych wersji danych badawczych

# Selekcja i przygotowanie danych społecznych do udostępnienia: SELEKCJA

Nie można

Nie można umieszczać w repozytorium:

- treści niezgodnych z prawem, w szczególności materiałów:
  - do których rozpowszechniania nie mamy uprawnień
  - których rozpowszechnianie w inny sposób narusza prawa osób trzecich, w tym ich dobra osobiste
- treści sprzecznych z dobrymi obyczajami, w tym godzących w zasady rzetelności naukowej
- w przypadku RDS – materiałów niebędących danymi badawczymi

# Selekcja i przygotowanie danych społecznych do udostępnienia: SELEKCJA

Przydatna  
literatura  
nt. selekcji  
danych  
badawczych

- Digital Curation Centre. (2014). Five steps to decide what data to keep: a checklist for appraising research data v.1 Edinburgh: Digital Curation Centre. Available online: [www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)
- Tjalsma, H. & Rombouts, J. (2011). Selection of research data: Guidelines for appraising and selecting research data. Data Archiving and Networked Services (DANS). Available online: <https://dans.knaw.nl/nl/over/organisatie-beleid/publicaties/DANSselectionofresearchdata.pdf>

# Selekcja i przygotowanie danych społecznych do udostępnienia: PRZYGOTOWANIE

## Co po selekcji?

1. Formaty plików
2. Nazewnictwo plików i folderów
3. Anonimizacja lub pseudonimizacja
4. Decyzja dotycząca licencji
5. Embargo?
6. Opis zbioru – metadane
7. Kwestie szczegółowe dotyczące danych jakościowych lub ilościowych



## 1. Formaty plików

- Rekomendowane formaty:
  - otwarte
  - bezstratne
- Akceptowane formaty: zamknięte (proprietary) bądź stratne, lecz będące w powszechnym użyciu
- Przydatne informacje: UK Data Service, <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>

Cel: zapewnienie łatwego odczytu danych i ich wysokiej jakości

## 2.

## Nazewnictwo

- W nazwach plików i folderów unikamy:
  - znaków diakrytycznych
  - spacji
  - znaków innych niż litery i cyfry, ewentualnie podkreślniki, którymi\_można\_zastąpić\_spacje
- Ewentualne numerowanie – 01, 001 itp.
- Nazwy krótkie, lecz znaczące
- Konsekwencja

Cel: uniknięcie problemów technicznych,  
przejrzystość zbioru

## 3. Anonimizacja lub pseudonimizacja

- Czy potrzebna?
- Anonimizacja – nieodwracalna
- Pseudonimizacja – odwracalna (osobno przechowujemy „klucz”)
- Osoby trzecie!

Cel: spełnienie wymogów prawnych,  
zapewnienie anonimowości osobom badanym

## 4. Licencje

- Różne licencje dla poszczególnych plików
- RDS:
- licencje CC (7 rodzajów), [creativecommons.pl](http://creativecommons.pl)
  - ograniczony dostęp:
    - wyłącznie w celach naukowych, bez możliwości udostępniania zbioru innym osobom i instytucjom
    - na innych zasadach

Cel: wyważenie między rekomendowaną otwartością a słusznymi ograniczeniami

## 5. Embargo?

- Embargo – kiedy warto?
- RDS: limit 36 miesięcy

Cel: wybranie optymalnej daty publikacji danych

## 6. Opis zbioru (metadane)

- Metadane: dane o danych, „ustrukturyzowane informacje opisujące zasoby informacji, np. zbiory danych badawczych”; „zawierają informacje o formie i treści zasobów, dzięki czemu pozwalają na ich wyszukiwanie i identyfikację oraz zarządzanie nimi” (Fenrich 2019: 6).
- Do zrobienia:
  - sprawdzenie wymagań repozytorium
  - zgromadzenie informacji „pod ręką”

Cel: uczynienie danych łatwymi do znalezienia i ponownego wykorzystania

# Selekcja i przygotowanie danych społecznych do udostępnienia: PRZYGOTOWANIE

Kwestie  
szczegółowe  
dotyczące  
danych  
jakościowych

- Nazewnictwo plików – sygnatury  
Przykład: zbiór „Biografia a tożsamość narodowa”, sygnatury plików: BTN\_01, BTN\_02...
- Struktura zbioru danych:
  - pliki podobne do siebie pakujemy w .zip
  - dołączamy spis materiałów, tzn. elementów zbioru (np. tabelę Excel)

Więcej informacji nt. archiwizacji danych jakościowych w RDS: [adj.ifispan.pl](http://adj.ifispan.pl)

# Selekcja i przygotowanie danych społecznych do udostępnienia: PRZYGOTOWANIE

Kwestie  
szczegółowe  
dotyczące  
danych  
ilościowych

## DOKUMENTACJA BADAWCZA:

### A. Informacje ogólne

kierownik/autor

tytuł badania

opis badania (cel)

### B. Informacje specyficzne dla typu danych

jednostka badania

technika zbierania danych

opis próby badawczej i sposobu jej doboru

okres zbierania danych (typ pomiaru)

rodzaj narzędzia badawczego

liczebność próby docelowej i zrealizowanej

opis próby rezerwowej i zasad jej wykorzystania (jeśli była)

minimalna liczba prób kontaktu

typ wagi (jeśli była) i sposobu jej konstrukcji

opis kontroli terenowej

### C. Narzędzia badawcze

### D. Książka kodowa (*codebook*)



# Selekcja i przygotowanie danych społecznych do udostępnienia: PRZYGOTOWANIE

## Kwestie szczegółowe dotyczące danych ilościowych

### ZBIÓR DANYCH:

---

#### Wymagane:

1. Określenie w nazwie zmiennej bądź jej etykiecie jednoznacznego odniesienia do odpowiadającego jej numeru pytania w kwestionariuszu.
2. Nadanie etykiet wszystkim zmiennym występującym w zbiorze
3. Nadanie etykiet wartości wszystkim wartościom wszystkich zmiennych (wyjątek to zmienne ciągłe lub bez przypisanej wartości)
4. Zastąpienie wszystkich systemowych braków danych (*system missings*) kodami specjalnymi wraz z opisem ich znaczenia
5. Przeprowadzenie procesu anonimizacji
6. SPSS, SPSS portable, STATA, ASCII+ddf

#### Zalecane:

1. Przyjęcie konwencji nazewnictwa zmiennych wg numerów pytań w kwestionariuszu
2. Zapisanie etykiet wartości odpowiedzi standardowo niedostępnych respondentowi wielkimi literami
3. Dla oznaczania wartości kodów specjalnych zalecane jest stosowanie wartości spoza skali zmienności danej zmiennej (-1; -2...)

# Prezentacja Repozytorium Danych Społecznych

Skąd dane  
w RDS?

RDS ([rds.icm.edu.pl](https://rds.icm.edu.pl)) to platforma dla zbiorów danych:

- opracowanych w ramach projektu DRODB
- dodawanych na bieżąco przez nowych użytkowników
- które były wcześniej udostępniane na platformie Archiwum Danych Społecznych ([www.ads.org.pl](https://www.ads.org.pl)) lub Archiwum Danych Jakościowych IFiS PAN ([adj.ifispan.pl](https://adj.ifispan.pl))

# Prezentacja Repozytorium Danych Społecznych

## Struktura RDS

