

WERYFIKACJI DANYCH WPROWADZANYCH do REPOZYTORIUM OTWARTYCH DANYCH

Tytuł i opis

Czy tytuł i opis precyzyjnie oddają istotę wprowadzonych danych: czy nie są generyczne, czy tytuł nie zawiera nadmiarowych informacji (np. nazwiska autorów, źródło finansowania)?

Opis powinien być możliwie szeroki, opisywać zarówno same dane, które trafiły do zbioru, jak i badanie, z którego się wywodzą.

Język

Najlepiej żeby wszystkie informacje były w jednym języku (wyjątek: tłumaczenie tytułu).

Nie mamy wielojęzycznych opisów, ale użytkownicy często kombinują w tym kierunku. Często afiliacje zostają po polsku, a reszta jest po angielsku – warto na to uważać i zmieniać na angielskie.

Słowa kluczowe

Czy są jakieś słowa kluczowe (nie są obowiązkowe, ale warto je sugerować, gdyż znakomicie ułatwiają znalezienie zbioru). Czy słowa kluczowe są wpisane są zgodnie ze składnią tj. kolejne słowo kluczowe po przecinku?

Informacje we właściwych polach formularza

Czy wszystkie informacje znajdują się we właściwych polach formularza (np. informacje o publikacjach w opisie nie powinny są tam, gdzie być powinny (np. informacje o publikacjach powinny się znaleźć w polu „Powiązana publikacja”, a nie w opisie).

Informacja o publikacjach

Czy powiązana publikacja:

ma identyfikator,

ma wskazany typ powiązania.

(bez tych informacji dane nie zostaną wprowadzone do Datacite)

Sprawdzenie wpisu dot. publikacji pod kątem: literówek, interpunkcja (kropki na końcu tytułu; brak kropek na końcu opisów).

Powiązana publikacja i zbiór: DOI w surowej postaci tam, gdzie jest wskazany identyfikator (ale w cytowaniu już lepiej jako link). URL najlepiej ten oparty na DOI.

Informacja o grantcie

Czy informacja o grantcie nie zawiera “nadmiarowych” informacji (np. w przypadku grantów europejski powinno być “H2020”, “FP7”, bez żadnych dodatków). Nie powinno się dodawać dopisków dotyczące podprogramów; w numerze grantu nie powinien być dopisywany tytuł projektu.

Format danych w plikach tekstowych/tabelarycznych (CSV)

Kodowanie tekstu: format UTF-8

Separator tekstu: cudzysłów

Nie dodawać zmiennych liczbowych (przy automatycznej analizie zmienna automatycznie zostanie potraktowana jako tekst).

Separator dziesiętny: kropka

Separator pól: przecinek

Pierwszy wiersz zarezerwowany dla nazw zmiennych.

Nazwy zmiennych nie powinny zajmować kilku wierszy.

Wszystkie niepuste kolumny muszą mieć nazwy.

Format danych w plikach Excel (XLSX)

Jeden plik, jeden arkusz. Jeden arkusz, jeden zestaw danych.

Nazwy plików: tylko a-Z, 0-9 i _.

Bez kropek innych niż ta rozdzielająca nazwę od rozszerzenia, bez polskich znaków, bez myślników (jest ich wiele rodzajów), bez znaków specjalnych.

Również wewnątrz archiwów.

Archiwum tylko w formacie ZIP

Wszystkie skróty w danych lub metadanych powinny być wyjaśnione (najlepiej w formie legend).

Unikać przekazywania informacji kolorami czy formatowaniem tekstu

Licencje – czy są zgodne w różnych miejscach?