

Zalecane formaty plików i wytyczne dotyczące przygotowania danych tabelarycznych //

Recommended file formats and guidelines for the preparation of tabular data

Tekst // Text

txt

odt

html

xml

oraz formaty natywne dla danego języka programowania, jeśli plik zawiera fragment kodu //

and programming language native formats, if the file contains code

Jeśli wykorzystano płatne biblioteki, należy – o ile to możliwe – przekonwertować plik do „waniliowej” wersji. Jeśli to niemożliwe, należy w opisie zbioru wymienić stosowane biblioteki. Można również dodać kopię pliku zapisaną jako plain text (.txt) //

If paid libraries are used, convert the file to a vanilla version. If this is impossible, the libraries used should be listed in the dataset description. A copy of the file saved as plain text (.txt) can also be added.

Obrazy // Images

png

jpeg2000

tiff

Audio // Audio

wav

Wideo // Video

wolne // free: otwarte, ale nie wolne (dopuszczalne) // open, but not free (acceptable):

mkv mp4

ogg mov

ogv

Dane filmowe zasadniczo są skompresowane i jest to zjawisko pożądane. Jeśli w danym obszarze badań jest potrzeba zachowania danych filmowych nieskompresowanych, zazwyczaj używa się autorskich formatów. Takie przypadki są weryfikowane indywidualnie //

Video data is usually compressed and this is a desirable feature. If there is a need to keep video data from being compressed, proprietary formats shall normally be used. Such cases are reviewed individually.

Archiwa // Archives

zip

CAD

formaty: .sldprt oraz .igs (CADowskie) można konwertować na .step (otwarty standard formatu pliku używany do wymiany danych między różnymi programami CAD/CAM/CAE) //

The .sldprt and .igs (CAD) formats can be converted to .step (an open file format used for data exchange between different CAD/CAM/CAE programs).

Dane tabelaryczne, arkusze kalkulacyjne // Tabular data, spreadsheets

csv

tab

ods

rdata

sav/spv (wyjściowe z SPSS // output SPSS files)

Przygotowanie plików CSV - wskazówki:

- należy stosować kodowanie typu UTF-8. W konsoli Linuksa można to sprawdzić: `file -i nazwaPliku.csv`,
- separator tekstu: cudzysłów. Zmienne liczbowe nie powinny być w niego ujęte, bo wtedy przy automatycznej analizie zmienna zostanie potraktowana jako tekst. Więcej o analizie danych tabelarycznych poniżej,
- separator dziesiętny: kropka,
- separator pól: przecinek lub średnik,
- pierwszy wiersz zarezerwowany powinien być dla nazw zmiennych, nie powinny one zajmować innych wierszy,
- wszystkie niepuste kolumny muszą mieć nazwy.

CSV file preparation - tips:

- use UTF-8 encoding. You can check this in the Linux console: `file -i filename.csv`
- text separator: inverted commas. Numeric variables should not be enclosed in it, because then the variable will be treated as text during automatic analysis. More on the analysis of tabular data below,
- decimal separator: full stop,
- field separator: comma or semicolon,
- the first row should be reserved for variable names, they should not occupy other rows,
- all non-empty columns must have names.

Automatyczna analiza // automatic analysis

Pliki csv, tab, xlsx, sav/spv przechodzą próbę automatycznej analizy – jeśli będzie ona udana, plik będzie dostępny dodatkowo w formatach tab, rdata i sav, będzie też możliwość jego przeglądania online. [Dowiedz się więcej](#).

Dla plików xlsx, które z różnych względów nie przejdą analizy, zostanie wygenerowana kopia w formacie ods //

The csv, tab, xlsx, sav/spv files undergo an automatic analysis test — if successful, the file becomes additionally available in tab, rdata and sav formats, and is possible to be viewed online. [Learn more](#).

For xlsx files that do not pass analysis for various reasons, a copy in ods format is created.

Aby zapewnić pomyślną analizę z trafnymi wynikami, plik danych tabelarycznych powinien:

- składać się z jednego tylko arkusza danych,
- zawierać tylko jedną tabelę w układzie pionowym, tj. każda użyta kolumna powinna zawierać jedną zmienną, a poszczególne wiersze — wartości zmiennych dla jednego przypadku,
- nie zawierać komórek scalonych (program analizujący dokona próby rozdzielenia scalonych komórek, ale robi to źle),
- nie zawierać pustych kolumn ani wierszy, w tym „marginesów” górnych ani lewych: tabela powinna zaczynać się od kolumny A i od wiersza 1,
- w wierszu 1. zawierać wyłącznie kompletne nazwy zmiennych,
- od wiersza 2. aż do końca zawierać tylko wartości zmiennych,
- nie zawierać komentarzy, objaśnień stosowanych jednostek i skrótów, opisów warunków pomiaru itp. — te powinny się znaleźć w opisie pliku, opisie zbioru i/lub w specjalnie przygotowanym pliku readme.txt ([wzór](#)).



Pliki niespełniające tych warunków nie będą odrzucane w procesie weryfikacji, ale mogą go znacząco wydłużyć. Pliki, które nie przeszły analizy, są trudniejsze w odbiorze i trudniejsze do ponownego wykorzystania//

To ensure a successful analysis with relevant results, any tabular file should:

- consist of only one sheet,
- contain only one table in a vertical layout, that is, each column used should contain one variable and the individual rows should contain the values of the variables for one case,
- not contain merged cells (the analysis program will attempt to split the merged cells, but will do so incorrectly),
- not contain blank columns or rows, including top and left 'margins': the table should start from column A and from row 1,
- contain only valid variable names in row 1,
- from row 2 downwards contain only the values of the variables,
- not contain comments, explanations of units and abbreviations used, descriptions of measurement conditions, etc. — these should be included in the file description, the dataset description and/or in a dedicated readme.txt file attached to the dataset ([template](#)).

Files not fulfilling these conditions will not be rejected in the verification process, but may significantly delay it. Files that do not pass the analysis are more difficult to read and more problematic to reuse.