Surveying Open Data Practices (SODP) - 2020

Anne Treasure

04/03/2021

Survey setup

The survey was developed in English and translated to French and Portuguese. Each of the three translations were loaded as individual surveys in the JISC Online Survey tool.

Preparing the translations for use in analysis

In order to analyse data from all three translations combined, we downloaded the key files in Comma-Seperated Value (CSV) format for each translation from the byline survey tool.

Each question and associated options appear as a row in the key CSV file. The file contains two columns:

- URN: Either the question number (e.g. Q1, Q25, etc) or the option number for the question (e.g. 1, 10, 12, etc) - *(Note that option numbers aren't uniquely associated with their questions, which may cause problems in analysis, merging, and more.)*
- Unique Response Number: The actual question (e.g. "What is your age?") or option (e.g. "18 - 24", "female", "I do not work in an institution conducting research", etc.)

```
#{r key_file_excerpt, echo=FALSE, fig.cap="An excerpt from the English version key
file."} #knitr::include_graphics(here("reports", "draft_images","key_file_screenshot.png"))
#
```

Due to the non-unique nature of the URN column, we added a unique ID column that could be used to merge the three translation files with. This was done to ensure the questions and options did actually correlate between the files. When we did this, we realised that the French survey missed the final option for Question 19.

- Question: What practical problems influence your sharing of datasets (please select all that apply)?
- Missing option: I have no practical problems sharing my data

To compensate for the missing option, we inserted a row with the option number "12" under URN and no text in the Unique Response Number-field.

The three files were then merged using the unique ID numbers for each row. Data was checked to confirm the data for each translation corresponds to the correct question and associated options.