
Citing research data

When using third-party data in your research, it is necessary to properly cite them. Similar to citing publications, it is important to provide enough information so that the data can be easily found and identified. The [Joint Declaration of Data Citation Principles](#) (Data Citation Synthesis Group 2014) specifies the basic principles of data citations. As many data repositories, publishers, and institutions adhere to these principles, often a recommended citation format is provided for published data that aligns with these principles. If there is no recommended citation format along the data and the author does not provide it in the documentation, you should include at least the following information:

- Author of the dataset
- The date when the dataset was published
- The name of the dataset
- The place of publication of the dataset (a publisher or a repository)
- Persistent identifier (if it was assigned to the dataset)
- The specification of the version or the subset of the data that you used (if applicable)

The screenshot shows a dataset page for 'Czech word and MWE lists' (Version 1.0). The citation information is: Cvrček, Václav, 2020, "Czech word and MWE lists", <https://doi.org/10.18710/PGDWXC>, DataverseNO, V1. A 'Cite Dataset' button is visible with a dropdown menu showing options for EndNote XML, RIS, and BibTeX. A 'Dataset Metrics' box shows 31 Downloads. The main text of the dataset description is: 'This post contains word and MWE (multi-word expression) lists used for the operationalization of some of the linguistic features in the multi-dimensional analysis (MDA) of Czech project carried out at the Czech National Corpus. The MDA procedure requires identifying and operationalizing linguistic features relevant for register variation in the language under scrutiny. In the Czech MDA project, some of these features were operationalized by compiling lists of words and multi-word expressions, which can then be matched against a text to identify occurrences. Compiling such a list can be tedious and error prone work, which is why we provide ours as a resource for other linguists either to adopt wholesale or at least use as a starting point to build on top of. (2020-04-06)'

Useful resources

Ball, Alex & Monica Duke. 2015. How to cite datasets and link to publications. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Dostupné z <https://www.dcc.ac.uk/guidance/how-guides/cite-datasets>

Data Citation Synthesis Group. 2014. Joint declaration of data citation principles. Martone M. (ed.) San Diego CA: FORCE11. <https://doi.org/10.25490/a97f-egyk>