Data Documentation

Scientific production on data repositories and open science published in the Web of Science database – Bibliometric conceptual analysis

Version 1.0 – May 2nd, 2024

Sinval Adalberto Rodrigues-Junior[1]; Marcelo Votto Texeira[1]

[1] School of Librarianship, Community University of the Chapecó Region – Unochapecó;

Abstract

This document describes data collected from the Main Collection of the Web of Science database. Records of published studies addressing the intersection of Open Science and data repository were searched up to January 15th, 2024, and the final dataset was comprised of 545 records for bibliometric analysis.

Citing the dataset: Rodrigues-Junior, Sinval, 2024, "Scientific production on data repositories and open science published in the Web of Science database – Bibliometric conceptual analysis", https://doi.org/10.7910/DVN/MZ1EUP, Harvard Dataverse, V1.

Motivation

Open science is a paradigm shift in the way research is produced and publicized, and data repositories are part of the infrastructure involved in open data. The understanding of what has been published on the issue may help clarifying the reach of the open science movement so far.

Methodology

The Main Collection of the Web of Science database was searched for studies addressing the theme ‘data repository’ and ‘Open Science’. The search strategy was built based on the term “data repository” and the terms that represented the context of Open Science based on the taxonomy defined by Pontika et al. (2015). Early access studies were removed, due to the lack of publication date. Also, proceeding papers, data papers and editorial material were excluded. Therefore, the dataset was composed by published original articles and review articles.

"data repository" OR "data repositories" (All Fields) AND "open science" OR "open access" OR "open data" OR "open science policies" OR "open science tools" OR "open repositories" OR "open data repository" OR "open data repository ecosystem" OR "open data repository eco-system" (All Fields) and Early Access (Exclude – Document Types) and Proceeding Paper or Data Paper or Editorial Material (Exclude – Document Types)

Data were analyzed with the package Bibliometrix 4.0 for R (Aria and Cuccurullo, 2017), through its web-based user-friendly version biblioshiny.

Components of the dataset

The data file wos_545_15-01-24.bib contains 545 records of studies addressing the theme proposed up to January 15th, 2024. The file contains the following information:

C1: Affiliation

AU: Author

RP: Corresponding author

DT: Document type

SO: Journal

LA: Language

PY: Publication year

TI: Title

TC: Total citation

AB: Abstract

DI: DOI

ID: Keywords Plus

DE: Keywords

CR: Cited references

WC: Science categories

Information about C1, AU, RP, DT, SO, LA, PY, TI and TC are excellent, meaning that there is no missing information. AB and DI miss 0.73 and 1.47% of the information, respectively. ID is rated as acceptable, since 19.27% of the information is missing, and DE misses 32.84%, and is rated as poor. CR and WC are completely missing.

About the project

The project had no funding. The authors declare that they have no conflict of interest of any nature or kind.

References

Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive Science mapping analysis. *Journal of Informetrics,* v. 11, n. 4, p. 959-975, 2017. Doi: https://doi.org/10.1016/j.joi.2017.08.007.