

# Cochrane Data Extractor: Automated Dataset Harvesting from Cochrane Systematic Reviews

Is automated extraction the way forward for extracting from the large-scale Cochrane datasets that Cochrane has started publishing? We developed an automated Python pipeline to download and organise the study-level data that Cochrane has put up on its website (pairwise, DTA, NMA). We designed it to check the data links and then download, validate, and organise the data into standardised datasets with covariates. The test set showed 11 pairwise datasets with 1,295 data rows at high accuracy. Uninterrupted runs were then carried out until 501 Cochrane reviews had been downloaded and their data organised into the Pairwise70 repository. We converted the Cochrane open data, which is distributed and hard to use, into a single data artefact to support meta-analysis research. This tool is limited to recent Cochrane reviews and does not work on other databases.

## References

1. Pairwise70 open meta-analytic dataset. <https://github.com/mahmood726-cyber/Pairwise70>