**Manual cross-validation data for the article: “Comparison of bibliographic data sources: Implications for the robustness of university rankings”**

Chloe Brookes-Kenworthy, Chun-Kai (Karl) Huang, Cameron Neylon, Katie Wilson, Alkim Ozaygen, Lucy Montgomery, Richard Hosking

Centre for Culture and Technology, Curtin University, Bentley 6102, Western Australia

**Abstract**

These are sets of data collected from the manual cross-validation of DOIs (and related research outputs) that are sampled from Web of Science (WoS), Scopus and Microsoft Academic (MSA). For each of the 15 universities, we initially collect all DOIs indexed by each of the three bibliographic sources. Subsequently, we randomly sample 40, 30 and 30 DOIs from sets of DOIs that are exclusively indexed by WoS, Scopus and MSA, respectively, for each university. A manual cross-validation process is then followed to validate certain characteristics across the data sources. This cross-validation process was carried out by a data wrangler, on a part-time basis over a few months, for which online data was accessed from 18 December 2018 to 20 May 2019.

**Content**

1. One Excel file (with 3 sheets) for the results collected for each university (i.e., 15 files, each named according to GRID ID and name).
2. A summary Excel file of all data collected.