OPEN ACCESS DASHBOARD
DATA DESCRIPTION

9 NOVEMBER 2020

DATA SOURCES AND OUTLINE

Data were collected from four main sources. Crossref Metadata was used to collate the population of outputs that were considered for analysis. Each DOI contained in this collection provided information on funding and publication dates. The FundRef dataset then provided further information on each funder, such as their type and geographical location. Bibliographic information was then obtained for Microsoft Academic, which allowed us to link objects with DOIs with institutions. Finally, the GRID dataset offered metadata about institutions, primarily for this analysis the institutions addressed, and therefore country.

LIMITATIONS

Research outputs that do not have an associated DOI are not included in this analysis. While this means we did consider the contribution of over 100 million outputs, there is still a substantial contribution to the scholarly record not currently covered by this identifier system. Additionally, Funder data only exists from the commencement of the FundRef initiative and is not complete, with quality diminishing the further back in time you go. Finally the bibliographic data source (Microsoft Academic) has substantial biases and limitations with respect to affiliation sources. These are minimised and less pronounced when working at aggregation at the country level as opposed to individual institutions however.

DATA GATHERING

For this analysis, we consider the set of all research outputs with Crossref DOIs as this target population. This is identified as the most practical approach that allows tracking and disambiguation of research objects using persistent identifiers.

Queries were run against a Microsoft Academic snapshot to result in a comprehensive set of outputs for each GRID and aggregated to the country level. Subsequently, these are filtered down to include only objects with Crossref DOIs. Objects were then aggregated up to the country level based on metadata about those institutions obtained from the GRID database.