

## **Evaluating the possibilities of DataCite for developing 'Open Data metrics' on the production and usage of datasets worldwide**

Nicolas Robinson-Garcia, Phillipe Mongeon, Wei Jeng, Rodrigo Costas

### *Abstract*

DataCite is an international not-for-profit organization aiming at improving data citation practices by facilitating access to data. The main approach to achieve these objectives relies on the assignation of persistent identifiers, namely DOIs, to datasets and recommending the adoption of citation standards in a similar vein to what happens with journal articles. So far, DataCite records have not been extensively considered for the production and development of metrics on Open Data, (in terms of production and impact). This paper aims at: 1) exploring DataCite as a potential data source for analyzing data sharing and developing open data indicators worldwide and 2) studying to what extent citations and altmetrics are made to data sets, particularly focusing on the presence and use of DOI standards.

**Methodology.** The whole DataCite database was downloaded in April 2016 and the data parsed and organized in a preliminary relational SQL model. More than 7 million records were obtained. Metadata on creators, publication dates, permanent identifiers, record types, etc. have been considered. The DOIs were matched with the DOIs provided in the cited references of the publications covered in the Web of Science and in the DOIs collected by Altmetric.com.

Preliminary results, in line with previous studies, show a very low presence of datasets in both citations and altmetrics as measured by the use of the DOI of the users in their references or altmetric mentions.

The results of the analysis will be contextualized with previous studies on Open Data, the Data Citation Index and other altmetric and bibliometric-based approaches. We will discuss to what extent are data citation practices common and offer some recommendations for the development of Open Data indicators. We conclude suggesting further research lines in the understanding and conceptualization of metrics for Open Data, Data sharing/dissemination and data reuse and citation.