

### ***Nature Communications*: Note to accompany the dataset**

There has long been a discussion over whether open research contributes to increased article usage and citations. In this debate, Nature Publishing Group has a good test case in *Nature Communications* – a born-hybrid journal providing a large sample size, where all articles are high-quality, original research and receive similar standards of service, regardless of whether or not they're open access.

In May 2014 Nature Publishing Group decided to partner with the Research Information Network (RIN), and asked them to undertake an independent analysis of this citations dataset. This report is intended to be an exploratory white paper, analysing machine-generated data on a bounded set of articles. The report seeks only to present the raw normalized data, and does not investigate the reasons for the observed differences.

The reasons may be complex and due to a number of factors, and more research is needed to understand them fully, but none of these factors were part of the original scope of the RIN report. Therefore, we realise that this research doesn't definitively answer the question of whether open access articles are viewed and cited more than subscription articles, but we think this contribution adds to the debate.

The data covered 2,878 articles published between April 2010 and December 2013, with citation data for all articles from publication up to April 2014.

RIN removed 4 articles because Web of Knowledge was not able to provide any data about the number of citations they had received. On reflection, they then removed 866 articles which were published on or after 1 July 2013. This was to ensure that articles with little time to gather citations did not skew the data for 2013. Data for the final question on web-based activity are a sub-set of the original dataset, comprising only the 722 articles published in the first six months of 2013.

It is important to note in presenting any analysis of citations of OA and subscription articles that RIN could not control for all possible factors, including the possibility that many of the articles published on a subscription basis may have been deposited and made accessible via institutional or subject-based repositories, or pre-print services such as ArXiv. Nor has this data enabled RIN to control for factors such as the number of authors of each paper, or their geographical location, or the possibility that authors may choose to make their 'best' papers OA. *Nature Communications*, launched in 2010, is a young journal which is still developing its brand and its author and institutional subscriber base. Factors such as these may well have had an impact on the citation scores. These results should be considered as a snapshot in the journal's development rather than definitive.