

# Tracking and Mitigating Climate Change - How open is the research?

Bianca Kramer<sup>1</sup> and Cameron Neylon<sup>2</sup>

1. Utrecht University
2. Centre for Culture and Technology, Curtin University

v0.9 10 March 2022

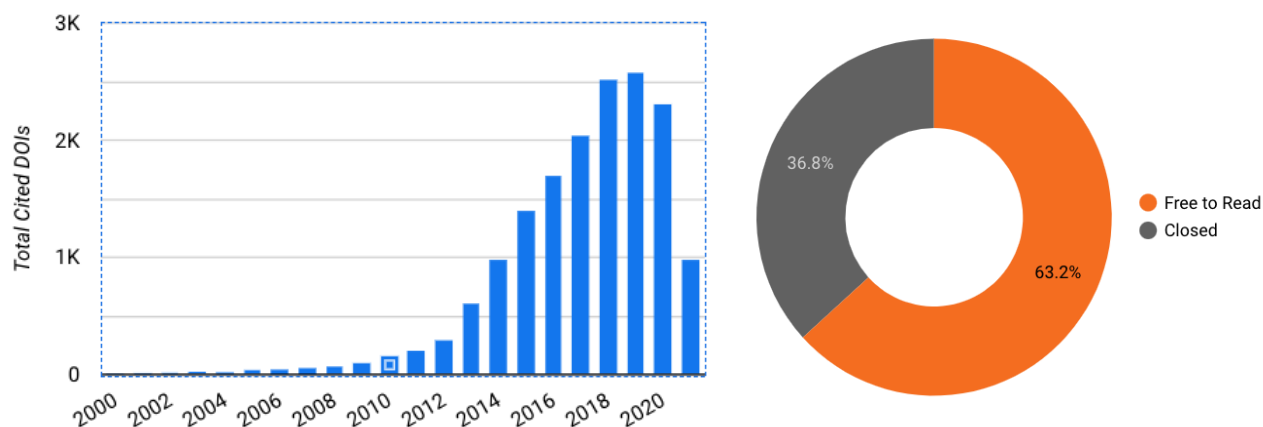
*Update - The [linked dashboard](#) now also contains the references from the 2021 WG1 Report on the Physical Sciences Basis. The [original dashboard](#) is still available.*

On Monday 28 February, the Intergovernmental Panel on Climate Change released the [Working Group 2 contribution to the Sixth Assessment Report](#). Coming in at 3,675 pages it shows the serious challenges ahead for preventing and mitigating the effects of the Climate Crisis.

So what can we tell about the underlying science behind the report? We extracted 17,420 unique DOIs (of which 16,325 could be matched to Crossref) that were cited in the report and linked them to the dataset we've built at the Curtin Open Knowledge Initiative to analyse. Following our original release we were also able to add [DOIs from the WG1 report thanks to previous work](#) by [Alexis-Michel Mugabushaka](#). The [resulting dashboard](#) provides a snapshot of research and analysis that is relevant to the issues tackled in the WG2 analysis, which looks at "Impacts, Adaptation and Vulnerability".

## The Highlights

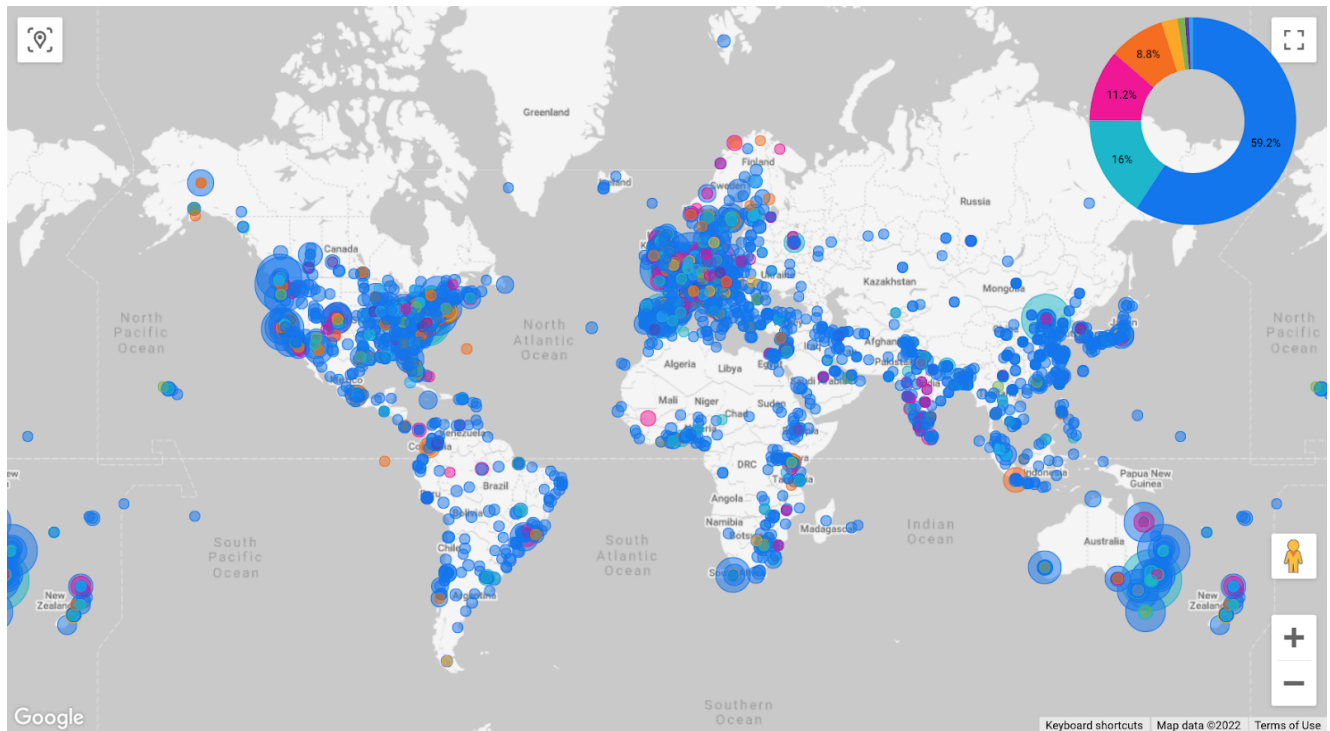
The majority of the cited research is recent, dating from 2014 onwards, and a substantial majority (63%) is freely accessible to read online. This is significantly higher than the global average of all literature (or at least things with Crossref DOIs) which is 43% for articles published from 2014 onwards.



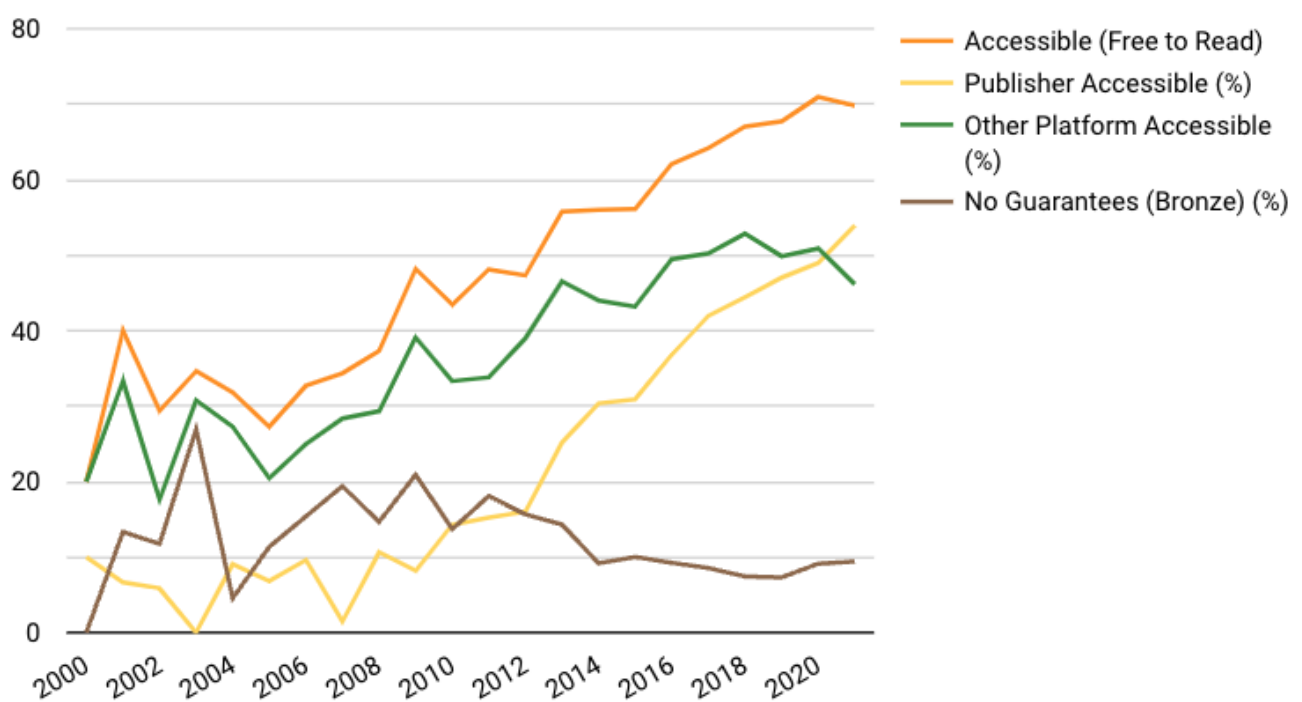
Less positively, for a report focused on impacts, adaptation and vulnerability, the research is largely the product of wealthy

institutions in rich nations. As is now commonly observed, contributions from African institutions are scarce, and many of the places with

significant risks of climate change effects appear under-represented. This is of course where the research is being *done* rather than where it is *about*, but if the climate crisis is a global challenge then engaging more researchers with experience of its effects on the ground can only help.



Also, less positive are patterns of how access is being delivered. While there has been an increase in access provided by publishers (sometimes referred to as “gold” OA) over the past ten years, the deposit of this literature into other repository platforms (including disciplinary, institutional and preprint repositories) appears to have stagnated over the past 5-7 years. Given that repositories are an important route to visibility access for academic and non-academic communities (watch this space for some more forthcoming evidence on this), a focus on improving deposit rates for this literature would provide significant dividends at low cost. It is nonetheless positive that over 50% of this literature is being published open access with at least some re-use rights with 6,195 of these outputs having a Creative Commons license.



Finally, one observation is that this analysis was harder than it needed to be. The references aren’t provided in a structured, machine-readable form and the referencing format differs from chapter to

chapter (every chapter except Chapter 1 and 6 included DOIs in the references at least, but in many chapters, DOIs are not consistently added to references). It would be great to see reports like this released with high quality structured data on the material being cited. Presuming that the authors are using some kind of reference manager this shouldn't be super difficult and has the potential to create greater confidence through transparency.

## The Approach

The PDFs for the 18 Chapters and 7 Cross-Chapter Papers of the WG2 report were downloaded from the IPCC site and uploaded to Scholarcy to extract citations, which were exported as RIS files. From these, DOI strings were identified by pattern matching. As not all reference information was processed correctly, DOIs were extracted from all fields, rather than just the DO field in the RIS files.

Scholarcy was successful at identifying 17,564 unique DOIs cited in the report. Lists of DOIs for each chapter and cross chapter paper were processed using a custom Python script to generate a pandas DataFrame which was saved as CSV file and uploaded to Google Bigquery.

We used the main object table of the Academic Observatory dataset, which combines information from Crossref, Unpaywall, Microsoft Academic, Open Citations, the Research Organization Registry and Geonames to enrich the DOIs with bibliographic information, affiliations, and open access status. A custom query was used to join and format the data and the resulting table was visualised in a [Google DataStudio dashboard](#).

A total of 16,320 DOIs could be matched to Crossref metadata. Most of the unmatched DOIs appeared to be truncated or otherwise incorrect but we have made no effort as yet to correct mismatches. Matching systematically biased our dataset towards journal articles. Compared to the output types identified by Scholarcy (76% journal articles, 14% books) 99% of entries in the matched dataset are identified as journal articles. We would welcome contributions to improve the identification of references, in terms of both improving reference capture from the reports and in improving the linking to DOIs.

## Data and Code Availability

RIS files of extracted references for each Chapter and Cross chapter paper, as well as a csv file of extracted DOIs are available from Zenodo and [Github](#), together with the code used to process and link the extracted DOIs to the COKI dataset.

- Alexis-Michel Mugabushaka. (2021). An open dataset of IPCC reports 'references (6th Assessment Cycle) – Version 1 (1.1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5475442>

### Additional outcomes:

- [Code and Data at Zenodo](#)
- [Code and Data at Github](#)
- [Dataset on BigQuery](#) - Requires a google account for access and bigquery account for querying
- [Data Studio Dashboard](#) - Interactive analysis of the generated data