Climate Justice in Electronic Publishing: A New Approach Supporting Global South Participation

Author copy: Submitted 21 Dec 2024 FINAL DRAFT SW

## Abstract

How current scholarly electronic publishing is carried out is inherently a climate injustice as it is negligently wasteful of the earth's resources and it unnecessarily hinders reading and participation by the Global South. Mega-publishers and societies are responsible for the state of electronic publishing and the resulting climate injustice.

A new publishing model for electronic publishing is proposed, informed by the semantic web, hypermedia, and the earlier technological visionaries who built and promoted global access to knowledge through granular indexing and linking. This new type of publishing remains unsupported in mainstream scholarly publishing, which renders the knowledge it contains almost unnavigable, especially in complex, fast moving research domains such as climate.

The authors of the paper are members of the #semanticClimate open research group, led by young Indian scientists, and together we are working to implement this new model for the IPCC climate reports. The Intergovernmental Panel on Climate Change (IPCC) report, the *IPCC* – *Sixth Assessment Report* is the most authoritative summation of climate change scientific knowledge and influences the acknowledgement of climate justice globally in policy for addressing climate change. As do other research bodies the IPCC uses conventional electronic publishing, but this holds back its potential wider reach. Our research presented here intends to demonstrate how such reports could be made more accessible.

The research presented here is a semi-automated literature search on the topic of ‘climate justice’ and asks the question – what does the open access scholarly corpus know about the topic and what is the shape of the discourse as it exists in this literature, in which papers and journals has the topic appeared, and reaching back several decades, how often has it occurred? This is a demonstration of the open-source tools that will be used for future work to create the Climate Knowledge Graph (ClimateKG) which aims to make the IPCC Report globally accessible.

The research covers the steps from an experiment already carried out by the team when the open access scholarly research corpus is searched and retrieved computationally from Europe Pubmed Central’s 6.5 million open access papers as well as all 70 chapters of the *IPCC* – *Sixth Assessment Report.* The outcome of the work completed is a first round scoping exercise to create the Climate Justice Dictionary. The dictionary represents terms associated with climate justice collected from the open scholarly corpus over 20 years as well as from the IPCC report.

---

**Keywords:** climate justice, climate change, IPCC, UNFCCC, UN Climate, knowledge graph, semantic web, hypermedia, dictionary, linked open data, Wikidata, 5-Star, linked open document

## Introduction

### Electronic scholarly publishing and IPCC’s publishing

#### Electronic scholarly publishing

This journal, the *Journal of Electronic Publishing*, focuses on scholarly electronic publishing and the argument that we are going to make is that scholarly electronic publishing, the publication process itself as currently practiced, generates climate injustice. Because the publication process is something which is not democratic, but instead is determined by a small number of mega-publishers and some mega-societies who make their own rules about electronic publishing. What we are going to suggest here, is that this is a process which is actually fundamentally unjust, particularly for the Global South.

We have a number of things to address. Firstly, the actual resources used in electronic publishing. Secondly, whether electronic publishing is an injustice for climate discussions. And thirdly, since climate injustice is often imbalanced between the Global North and the Global South, how electronic publishing disadvantages the Global South.

We have a number of aspects which are unjust here. We have the fact that mega companies are creating the agenda for academic publishing and that the people who are consuming or writing about climate justice are not able to influence this agenda. The agenda includes the fact that these companies and publishing societies: a. Dictate the technology of publishing, and; b. They create surveillance publishing. (Pooley 2022) In some cases this leads to massive resource depletion, especially when it comes to the introduction of artificial intelligence models. (Tamburrini 2022)

Open access publishing is also increasingly controlled by mega publishers. (Butler et al. 2023) Here publishing is not done as a public good, but to extract profit for publishing corporations. The costs of fee-based commercial OA models are huge for the Global South, rendering mainstream scholarly publishing fundamentally unjust. (Raju and Badrudeen 2022)

Looking specifically at climate research, what this means is that the Global South is not an equal contributor to discussions on climate justice and they have to accept the agenda that the Global Northern mega publishers create. This agenda is English language centric,(Chowdhary 2024) which means that many of the countries in the Global South who suffer already from the effects of climate change do not have a say in how knowledge is reported about their problems and potential solutions. We believe that this is one major aspect of climate injustice in the publishing system.

The second injustice mentioned was that the surveillance aspect that publishing companies currently practice has a hidden agenda of monitoring authors and readers and collecting or extracting data from them. While this is not immediately a climate issue, this form of data harvesting requires large-scale data computing, further contributing to the depletion of the world's resources. The generation of excess heat from data farms requires the use of fossil fuels and is only compounded by supporting artificial intelligence systems. (Tamburrini 2022) These resource use issues are unknown in terms of scale, are undemocratic, and lead directly to misuse of the world's resources.

Furthermore, the process of publication is not aimed at knowledge transfer but at revenue generation for publishing corporations. The technology of publishing, based on PDF, has remained constant for twenty-five years. This is an inefficient way of transmitting knowledge and also leads directly to waste and inefficiency as PDF is used as a mono-directional form of information transfer, created by publishers and forced upon readers. Readers have little say in how they consume climate research and it is extremely difficult to transform PDFs into any other format, which does more than only supporting reading with human eyes. Turning PDFs into semantic form, making them searchable, translating them into other languages, adapting them for different classes of reader, is almost impossible. Therefore if you cannot easily read PDFs in English, you cannot have access to scholarly climate knowledge. In the open research project #semanticClimate[[1]](#footnote-1) we address a number of non-academic constituencies: high school students (very possibly without English as a first language); city planners who need climate research to shape regional climate plans, but it is not available in the kind of form that they can easily consume; people with limited eyesight, etc.

In general, PDFs are not structured for ease of reading. They do not take account of modern mobile devices and the fact that in many countries the dominant form of research consumption is mobile devices rather than laptops or desktops. (Sobral 2020) On mobile devices PDFs do not display well or at all, which becomes an issue when displaying for example large tables which need to be rotated by 90 degrees, as well as tables and figures which are split over more than one page, which again are extremely difficult to read on a mobile device. Modern software allows research to be consumed in a much more efficient manner than PDFs. In many cases, readers have to download the whole paper and print it out, leading to further paper waste.

### IPCC reports and climate justice

The IPCC has an objective set out in its mandate to ‘provide governments at all levels with scientific information that can be used in climate policies’ (IPCC 1988) and from this is charged with assessing pathways for mitigating climate change. (Hughes 2024) The pathways represent likely futures for the earth's climate and the state of climate change, and they also represent the effects of mitigating actions. (‘IPCC Glossary Search’, n.d.) The actions to achieve a pathway, such as the 1.5C Emission Pathway,[[2]](#footnote-2) are then taken into a treaty process called the UNFCCC from which member states agree on actions that directly affect people's life outcomes. The conjoined agencies of the IPCC and UNFCCC operate under a model of Shared Socioeconomic Pathways (SSPs)(Wei et al. 2018) which inherently brings climate justice to center focus as a global issue between rich and poor nations of the Global North and Global South.

The approach of SSPs has its opponents in climate change denial. (Jylhä and Hellmer 2020; Milfont et al. 2018) In 2017 the first Trump US administration announced its departure (Wikipedia 2024b) from the Paris Agreement and the decision to take a different approach to climate change, which is to abandon the idea of a shared future for humanity, as succinctly hypothesised in Bruno Latour's 2018 book *Down to Earth.* (Latour 2018)

The IPCC reports are much more than a report on climate research or the science of climatology and the current status of the planet's climate. Instead they are a *knowledge and policy instrument* that forms a vital organ for bringing climate justice into legal frameworks and justice systems. (UNDP 2023)

While repeating the assessment cycle over its 32 years since the first *Assessment Report* was published, the IPCC has evolved its report publication process. (IPCC 1992) The open access book *The IPCC and the Politics of Writing Climate Change* by Hannah Hughes covers in detail the working of the IPCC as a knowledge endeavor of herculean scale that is dealing with a dynamic field of study, and where the interface of scientific knowledge and society create political tensions. (Hughes 2024) Hughes’ work is an example of a study of the IPCC work that the Climate Knowledge Graph, which we will outline more in depth in what follows, will no doubt extensively make use of as it is one of the few places that shed light on the IPCC knowledge creation process.

The authoring processes of the IPCC as a scientific working group also had to respond and change in relationship with and response to climate justice questions. As an example, in 1996 the Second Assessment Report, Working Group III (WG3) with an emphasis on economic thinking, lacked a protocol to ensure proper consultation and as a result the following was submitted in its report ‘suggesting a cash value of $1.5 million to a human life in the OECD against $150,000 in developing countries’. (Pearce 1996, 177) The chapter was not approved by the subsequent plenary session. As a result, following assessment reports procedures and rules were tightened and codified, for example assigning the partnering of chapter lead authors from developing countries. (IPCC 2013)

Chapter 6 of the WGIII SAR used controversial assumptions to calculate the ‘social costs’ of climate change, suggesting a cash value of $1.5 million to a human life in the OECD against $150,000 in developing countries (Pearce et al. 1996). As a result of developing country objections the final report did not make it through plenary approval and an additional session had to be scheduled. (Agrawala 1998, 626)

*The IPCC and the Politics of Writing Climate Change* (Hughes 2024, 70 Footnote 20)

In terms of the IPCC reports themselves being part of a climate justice system, of sorts, the current evolution of the mechanisms of its functioning is briefly outlined below. There are many other agreements and treaties that play significant roles in the process of moving science knowledge into policy and legal frameworks, but for reasons of brevity and for illustration the following have been highlighted. The initial IPCC *Assessment Report* in 1992 triggered UN member states to take action. The treaty body was formed, the United Nations Framework Convention on Climate Change (UNFCCC), and a key task of this body was to establish a framework for how nations would organise themselves to respond to the cause of climate change, the production of Greenhouse Gases (GHG), this was the Kyoto Protocol. (UNFCCC 1997; n.d.-b) Then in 2015 all UN countries negotiated the Paris Agreement which set a target of global surface temperature preferably to be 1.5 °C. (UNFCCC 2016) Ongoing treaty negotiations take place in the decision-making body called Conference of the Parties (COP). (UNFCCC, n.d.-a) These multilateral agreements are then enacted in member state legislation. To illustrate how this mechanism fully acknowledges climate justice see the example COP treaties clause below:

*Also acknowledging* that climate change is a common concern of humankind, Parties should, when taking action to address climate change, respect, promote and consider their respective obligations on human rights, the right to health, the rights of indigenous peoples, local communities, migrants, children, persons with disabilities and people in vulnerable situations and the right to development, as well as gender equality, empowerment of women and intergenerational equity,

FCCC/PA/CMA/2021/10/Add.1. Decision 1/CMA.3. Glasgow Climate Pact (UNFCCC 2022)

The IPCC authors, including the chair Jim Skea, have made a submission to the International Court of Justice on an ‘advisory opinion’ tasked from the court by the United Nations General Assembly resolution 77/276 on ‘Obligations of States in respect of Climate Change’. (ICJ 2024b) It is in these contexts that climate justice is codified into law.

39. China submits that developed countries have an obligation to bear their historical responsibilities. IPCC reports80 reveal that historical emissions from developed countries are the primary cause of the current climate crisis and injustice.’ (ICJ 2024a)

#### The IPCC’s digital publishing endeavours

In this section we will argue that the IPCC’s reports, or the way in which they are made available through digital publishing, is influenced by two key factors: the academic publishing industry and governmental Grey Literature practices. (GreyNet 2024)

Firstly, IPCC’s digital publishing implicitly follows the norms dictated by the academic publishing industry. The scientific work, research, and data science carried out by the IPCC, makes up the most authoritative climate science available and is carried out to the highest standards. The data science work of the IPCC has been modernised and is carried out using modern open science methods using FAIR Data Principles and data is openly stored on GitHub. (‘FAIR Principles’, n.d.; ‘IPCC-WG1’, n.d.)

Secondly, the IPCC’s remit is to ‘provide governments at all levels with scientific information that can be used in climate policies’. (IPCC, n.d.-a) The primary focus point for the IPCC to achieve this is by adding clauses to its Summary for Policymakers (SPM), which are IPCC Assessment Report (AR) summaries that have to be agreed on unanimously by all signatories, nation states and parties (and AR6 included 195 member). As an example here is a clause in the section Mitigation and Adaptation Options across Systems from *The Climate Change 2023: Synthesis Report (SYR)*: (Calvin et al. 2023)

C.3Rapid and far-reaching transitions across all sectors and systems are necessary to achieve deep and sustained emissions reductions and secure a liveable and sustainable future for all. These system transitions involve a significant upscaling of a wide portfolio of mitigation and adaptation options. Feasible, effective, and low-cost options for mitigation and adaptation are already available, with differences across systems and regions.  
(high confidence) {4.1, 4.5, 4.6} (Figure SPM.7)

*The Climate Change 2023: Synthesis Report (SYR)* [IPCC\_AR6\_SYR\_SPM.pdf#page=34.35](https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf#page=34.35) (Calvin et al. 2023)

#### Climate justice and IPCC *Sixth Assessment Report*

The IPCC in their *Sixth Assessment Report (AR6)* has an emphasis on climate justice while addressing climate change. The different chapters in AR6, Working Group I: Climate Change 2021: The Physical Science Basis; Working Group II: Climate Change 2022: Impacts, Adaptation and Vulnerability, and; Working Group III: Climate Change 2022: Mitigation of Climate Change (WG1/WG2/WG3) (Intergovernmental Panel On Climate Change (Ipcc) 2023a; 2023b; 2023c) describe the importance of equitable climate action, integrating local and Indigenous knowledge with scientific approaches for effective climate adaptations and mitigations. The concern about climate justice in the IPCC reports has been raised due to unequal losses and damage to vulnerable populations. There are disproportionate adverse impacts of climatic hazards on people who have contributed less to climate change. Climate vulnerability has also increased due to gender inequalities, so gender considerations are crucial for climate justice AR6, WG2, Chapter 08 argues. (Intergovernmental Panel On Climate Change (Ipcc) 2023b) The different findings from the reports have shown the need for climate justice while making adaptation strategies. The Paris Agreement and the 2030 Sustainable Development Agenda has widened the scope of adaptation governance by linking adaptation to development and climate justice in AR6/WG2, Chapter 18. (UNFCCC 2016, UN 2015, IPCC 2022) (UNFCCC 2016; UN 2015; Intergovernmental Panel On Climate Change (Ipcc) 2023b) Many environmental and climate justice activists have drawn attention to considerations of economic and environmental inequalities to increase awareness and also advocate for stronger climate mitigation efforts in AR6/WG3, Chapter05. The Civil Society and Social Movements section mentioned in the AR6/WG3/Chapter 14 (IPCC 2023) have focused on protecting rights, adopting responsibility-based approaches to climate finance, adaptation and mitigation. They have also raised questions about the equitable allocations of future climate budgets to developed and developing countries. The Climate Change 2023: Synthesis Report (SYR) (Calvin et al. 2023) has mentioned the disproportionate risks from climate hazards such as heatwaves, droughts, and flooding to poor and marginalized communities who are less responsible for this adverse climate change. The SYR report has urged the organization and government to design and structure adaptation and mitigation strategies that should consider equity and justice. Climate justice recommendations mentioned by IPCC reports in the different chapters assist in the development of policies that will consider each and every section of society.

### A new publishing model

To further contribute to the climate justice endeavors proposed within the IPCC reports, we will in the next section outline a new publishing model that could be applied to the publication process of the IPCC reports. The new publishing model is informed by the semantic web, hypermedia, and the work of earlier technological visionaries who built and promoted global access to knowledge through granular indexing and linking – from Tim Berners-Lee, Alan Kay, Ted Nelson, Vannevar Bush, J. C. R. Licklider, Donald Knuth, to Paul Otlet a century before the semantic web, and reaching even further back to annotation and hermeneutics. (‘Tim Berners-Lee’ 2024; Kay 1972; Nelson 1974; Bush 1945; Licklider 1965; Wright 2014; Krajewski and Krapp 2011) Even with such a long history, semantic publishing is still waiting in the wings, and has not reached the mainstream practice of scholarly publishing. The publishing model we will describe next therefore represents a new type of publishing that technically enables multi-linguality, better global discoverability, more accurate scientific categorisation and indexing, and re-use and republishing. In addition to that, the perspective on research publishing taken here is to focus on the knowledge being available in academic repositories or corpora as a whole as opposed to prioritising individual papers or journals. To test this model, we conducted a semi-automated literature review on the topic of climate justice and asked the question – what does the scholarly corpus know about the topic and what is the shape of the discourse as it exists in this literature?

The literature search on climate justice is a demonstration of the open-source tools that will be used to create the Climate Knowledge Graph ([ClimateKG](https://climatekg.semanticclimate.net/)) – and some of the specifics of knowledge graph capabilities. Knowledge graphs is a data store that has nodes and properties, and the term Graph is a mathematical term for relation. (Wikipedia 2024a) ClimateKG is an online graph database of a literature corpus connected to a search and publishing engine. The corpus can be searched and the search results saved online as well as being collated into a multi-format publication. ClimateKG could be used for frictionless dissemination of the IPCC reports, and include: cataloguing the literature corpora; word search; content publishing and reuse. For example, we searched and retrieved papers and data computationally from the open access repository Europe Pubmed Central (Europe PMC) and its 6.5 million open access papers as well as all 70 chapters of the *Intergovernmental Panel on Climate Change* – *Sixth Assessment Report (AR6)(IPCC AR6).* (Calvin et al. 2023) The output of this search showed *where* “climate justice” (in which papers and journals) is being used and creates tables using JQuery [DataTables](https://datatables.net/) software to browse the collected results that are stored in a mini-corpus. Data analysis was then carried out to find associated terms, which we used to create the Climate Justice Dictionary, which is a word list, which includes all the associated terms from Europe PMC and the IPCC and annotations from Wikipedia, Wikidata, and Wiktionary stored as a linked open data document.[[3]](#footnote-3)

Why are knowledge graphs important? And why are they important to improve access to Climate Justice literature? The scale of scholarly literature and its organisation is too complex for a person, group, or institution to easily navigate: to drill down, filter, retrieve, collaboratively review, reuse, or even cite. The ClimateKG project's primary mission is to make IPCC Reports accessible globally. As argued earlier the IPCC Reports are authoritative climate science and policy literature that regularly report on the status of climate change including ways that the dangers from climate change can be addressed. Currently the IPCC Reports are published as PDF or with unstructured web versions which makes them virtually unusable in modern information systems like the web, as a consequence the promulgation, reach, and efficacy is limited. The most recent *IPCC AR6* report is a substantial corpus itself – and this is the sixth assessment report since the Intergovernmental Panel on Climate Change (IPCC) creation 36 years ago, in 1988. A knowledge graph effectively creates multiple layers of indexing on top of the untouched ‘version of record’ publication. Such a layered approach was adopted in geospatial cartography in the 2000s by the Open Street Maps project. (‘OpenStreetMap’ 2024) Such indexing is technically called semantic annotation and uses ontologies, taxonomies, and controlled vocabularies. Concepts within the publication can be tagged, for example ‘[climatology](https://www.wikidata.org/wiki/Q52139)’ and ‘[climate justice](https://www.wikidata.org/wiki/Q1291678)’. When tagged using Wikidata, access to a host of trusted source IDs and language translation are available – 97 language translations and 50 identifiers for the former and with 40 translations and 8 identifiers for the latter. Many other semantic and syntactic questions can be added as indexed and linked layers to the knowledge graph: IPCC pathway frameworks[[4]](#footnote-4), use of climate models, data and supporting software used, use of other syntheses papers and their data. The list goes on.

Currently this syntactical and semantic structuring and indexing has to be done retrospectively. That is until the new publishing model being described here and by many others is carried out at the time of authoring and publishing of the reports. (Stocker et al. 2024; Capadisli 2019; ‘Papers with Code - The Latest in Machine Learning’, n.d.)

Back in 2011, Tim Berners-Lee proposed a 5-Star deployment scheme for data (Berners-Lee 2011) (see Table 1). Since then the open-science movement has made the infrastructures needed to enable this scheme’s implementation. In data management this has been taken up almost universally with the FAIR Principles.[[5]](#footnote-5)

|  |  |
| --- | --- |
| ★ | make your stuff available on the Web (whatever format) under an open licence |
| ★★ | make it available as structured data (e.g., Excel instead of image scan of a table |
| ★★★ | make it available in a non-proprietary open format (e.g., CSV instead of Excel) |
| ★★★★ | use URIs to denote things, so that people can point at your stuff |
| ★★★★★ | link your data to other data to provide context |

Table 1: 5-Star deployment scheme for data

But, in scholarly publishing, publishers have been the implementation bottleneck where it concerns research publications. Even if researchers want to provide linked open data (LOD)[[6]](#footnote-6) formats, interoperable formats, embedded LOD, or computational forms with literate programming – in most cases academic publishers across all fields have not implemented any infrastructure to accommodate this. (Stocker et al. 2024; ‘Calls for Linked Research’, n.d.; Adema 2023)

One thing that would be helpful in this context is a Linked Open Document – Checklist (see Table 2), which would support implementation of a 5-Star model for research publications. This would at least allow authors to deposit a preprint-with-code (LOD) of their publication in a repository, crucial where LODs are really needed in complex, data driven, fast moving knowledge domains.

| **Linked Open Document –** **Checklist (preprint-with-code)** | | |
| --- | --- | --- |
| **No.** | **Item** | **Check** |
| 1. | Add to Wikidata to connect to Linked Open Data Cloud: publication metadata, identifiers to significant properties and component, and Identify and link concepts |  |
| 2. | Regardless of target format create and display a semantic version openly, in a validated interoperable format: HTML, JATS, etc |  |
| 3. | Use W3C Typesetting: Web Publication manifest and CSS Paged Media |  |
| 4. | Validated interoperable format: HTML, JATS, etc |  |
| 5. | Open licensing |  |

Table 2: Linked Open Document – Checklist (preprint-with-code)

The work we are doing with the Climate Knowledge Graph has been produced as part of the open research community that organizes itself under the banner of #semanticClimate (Yadav et al. 2024)(Bhadra, Kumari, and Murray-Rust 2024; Yadav et al. 2024). In this article we are conceptualising the ClimageKG and in 2025 the software development phase will start. This community is creating software, organising events, and is contributing to creating the ClimateKG, which, as explained earlier, will be a web service for search and publishing, to enable global dissemination of the knowledge contained in the UN IPCC reports. Future work (which will take place during 2025) with partners including the Leibniz Joint Lab Future Libraries & Research Data, Open Science Data Services, Open Research Knowledge Graph ([ORKG](https://orkg.org/)), and Lab Knowledge Infrastructures at TIB – Leibniz Information Centre for Science and Technology (TIB), alongside the #semanticClimate community.

## Our research carried out for the semi-automated literature search

With our research, we want to explore and demonstrate, using the open-source software tools used by #semanticClimate, the use value of ‘knowledge graph’ technology to establish the extent of a given topic in scholarly literature corpus. By having the ability to have data on where a term occurs, it permits logical reasoning to be carried out over a corpus, knowing the degree to which something occurs, either in a single paper, over corpora, or in a temporal aspect which allows data analysis to reveal patterns. A temporal word frequency N-Gram is such an example. As explained before, we used ‘climate justice’ as a term to search scholarly literature corpus repositories.

In scope are open access research articles that are slanted towards bioscience, but the searched literature included other disciplines that were available in open repositories favorable to preprints, and that permitted ‘text and data mining’. Preprints were highlighted so that a representative view can be gained of contemporary, up-to-the-minute, research.

The main research question we asked was — how we could efficiently and usefully locate the discourse on the topic of climate justice in the open access scholarly literature corpus and the *IPCC Sixth Assessment Report* in a way that is open science compatible and informed by the *UNESCO Open Science Core Values and Guiding Principles*? (UNESCO 2021) We further asked the following sub questions:

What is the frequency of the discussion of climate justice in scholarly literature, both by temporal parameters — e.g., 3 year periods — and in consideration of geographic aspects. Word frequency is determined using the natural language processing software library called [spaCy](https://spacy.io/). (‘spaCy · Industrial-Strength Natural Language Processing in Python’, n.d.) Using work frequency helps reveal patterns and gain insights in the literature.

What associated terms do authors use for climate justice? The terminology is important as it encapsulates the discourse around the topic of climate justice. Terms we found include:

* + Organizational Justice
  + Climate justice movement
  + Environmental Justice
  + Resilient systems and restorative justice
  + Procedural Justice
  + Justice in sustainability science
  + Equity and justice

The terminology can then be used in review processes by participants in evaluating the results and in further data analysis that looks at aspects such as word frequency analysis, generating vector embeds, machine learning, NLP, etc.

What is the use of the term climate justice and its associated terms as they appear in UN Climate literature: the *IPCC Sixth Assessment Report*?; (IPCC 2023) and the *IPCC Glossary*? (IPCC, n.d.-b) The searches are semi-automated, which means researchers review the data, and then from the data on term-use the researcher is either aided in finding relevant research or it can help in understanding or interpreting patterns.

How does the term climate justice and its associated terms appear in Wikipedia, Wikidata, and Wiktionary? We selected these resources as these are important open public resources for climate information, but they are also important for knowledge graph creation, for annotation — indexing, cataloguing, and multilinguality, etc.

## The Climate Knowledge Graph (Conceptual modelling and future work)

Knowledge Graphs are part of the modern implementation of the Semantic Web — which enables automated knowledge management using data about definitions and relationships. In the context of the web, semantic refers to defining meaning in a machine readable way. But the web has a design fault — it has no cataloguing or indexing system. The web’s creator, Tim Berners-Lee, proposed a solution to this, which he coined as the ‘Semantic Web’. But his intervention came too late, after the web had already developed in another direction, and since then those interested in the opportunities that the semantic web has to offer have had to try and continue implementing his vision under difficult circumstances. The tech giants exploited the lack of a central cataloguing or indexing system and created much of their wealth by making their own private knowledge graphs. An example of the Semantic Web is [Google Knowledge Graph](https://en.wikipedia.org/wiki/Google_Knowledge_Graph) (‘Google Knowledge Graph’ 2024) which has 500 billion facts (2020) and generates the infoboxes on search results.



Figure 1: ClimateKG tech stack schematic.(Graphvis [source file](https://github.com/semanticClimate/JEP-article/blob/main/graphviz/Knowledge_Graph.dot)) (‘Graphviz’, n.d.)

Our work on the ClimateKG involves building aknowledge graph for the UN [IPCC Assessment Reports](https://www.ipcc.ch/report/sixth-assessment-report-cycle/). The knowledge graph would be used for dissemination of the reports, and would include the following functionality: cataloguing the literature corpora using ontologies, taxonomies, controlled vocabularies; word search interfaces as used on academic repositories; and content publishing and reuse with the ability to save search results collated into new full text editions and output as as multi-format, Web, ePub, PDF, etc.[[7]](#footnote-7)

The IPCC reports are the gold standard of climate science, but they are published in a rudimentary way as they are only being released as PDFs while some are also published as basic web pages. IPCC Reports are ripe for semantification as many pre-existing parts just need connecting — data, citations, author list, and glossaries, etc. The ClimateKG goal would be to have the IPCC *Sixth Assessment Report (AR6)* available online with the added functionality afforded by semantic publishing. The knowledge graph allows for queries that can collate and export as multi-format publications.

The #semanticClimate community has been developing software tooling since 2019 for searching and semantifying climate literature. Over 2023/2024 numerous hackathons and developments have been made resulting in several forms of semantification of the *UN IPCC Sixth Assessment Report*, including Text and Data Mining conversion to HTML of all [70 chapters](https://github.com/semanticClimate/ipcc/tree/main/cleaned_content), [2000 climate terms aligned with Wikidata](https://github.com/petermr/semanticClimate/blob/main/ipcc/ar6/test/total_glossary/glossaries/total/acronyms_wiki.csv), and the extraction of a [Glossary of 600 terms](https://vivliostyle.vercel.app/#src=https://raw.githubusercontent.com/semanticClimate/glossary-demo/main/ipccglossary.jsonld). In 2025 #semanticClimate will be working with TIB and National Institute of Plant Genome Research (NIPGR) to build the ClimateKG as a web service. Being a library web service means that there is an institutional guarantee that ensures that material is authoritative, that availability is maintained, and there is support for the different publics and user groups.

The ClimateKG, which we will be building in the future, will be a modern catalogue and index of climate science literature with a subject-focus on UN Climate and UN IPCC corpora. As explained in the methods section, it will be open-source and is based on open science principles. The ClimateKG technology allows for search, semantification (adding meaning), data analysis, machine learning, and machine learning and AI functions of all types (scientifically validated, open science compliant, and ethically based in the Sustainable Development Goals (SDGs) which is necessary for climate justice), and re-publishing. Re-publishing here means supporting frictionless and authoritative reuse of content, our first example planned product will be a search function that collates the search result sections into a reader that can be accessed online and outputted as a multi-format publication. See the following prototype as an example, *IPCC Reports and City Climate Change Plans: Proof of concept prototype – Open Climate Reader*. (#semanticClimate 2023)

The knowledge graph creates a map of the literature included in the corpora as a data set that can be used for granular search and can retrieve results down to a sentence level. It also assigns meaning to content by using a combination of ‘definitions’ and ‘relationships’. Definitions use controlled vocabularies, such as, *type of thing* ‘[Organization](https://schema.org/Organization)’ — e.g., the IPCC. The IPCC has expertise in the following, so has the *property* of ‘[knowsAbout](https://schema.org/knowsAbout)’: ‘[climatology](https://www.wikidata.org/wiki/Q52139)’ and ‘[climate justice](https://www.wikidata.org/wiki/Q1291678)’, which reflects its relationships.

With only these few bits of data (The organisation UN IPCC and its expertise) we can tap into Wikidata linked open data on ‘climatology’ and retrieve 97 language translations and 50 identifiers on ‘climate justice’ with 40 language translations and 8 identifiers. The identifiers mentioned here are authority control systems and are used to disambiguate, validate, or carry out queries on terms, such as: Stanford Encyclopedia of Philosophy ID, OpenAlex ID, JSTOR topic ID (archived), etc.

## Methods

Our work on the ClimateKG up to this point has been based on various methods. First of all, it has been organised as an open research community. The#semanticClimate Community, who make open-source software, is the umbrella context for the research that we are discussing here. Community is the ‘how’ in how we make software, as we choose to work as an open and public community to develop this software. Community driven software development is run through, amongst others, conducting hackathons, prototyping and demoing, running intern programmes, learning programmes, and experts and partners organisation contributions. Secondly, we have employed open-source coding and creating software. The code made by the #semanticClimate community is literally the ‘why’ and ‘what’ of the project.[[8]](#footnote-8) Open-source methodologies are inherently social and facilitate community building, these include as examples: using open intellectual property, creating learning communities, open review processes, and knowledge sharing. Participants are immersed in a continuous learning environment that has room for novices and experts. In addition to that we have made use of open channels for communication that supportsynchronous and asynchronous working (e.g., our weekday meetings are held 13:30 IST on Slack. We have also shared all our research in development employingopen-notebook science, an open science practice which means making all research activities open and made available as they are created to ensure there is no insider knowledge. (Wikipedia 2023) We have also followed*UNESCO Open Science Core Values and Guiding Principles* and FAIR Principles. (UNESCO 2021) A further significant component is that ourresearch outputs were versioned using storage on Git and deposits as Digital Object Identifiers (DOIs) minted on the CERN hosted online service Zenodo and other Persistent Identifiers (PIDs)[[9]](#footnote-9) systems with a guidance implementation checklist for researchers including review.Finally, we have employed the Linked Open Document Checklist we mentioned before to follow guidelines and submit a preprint-with-code and have usedsoftware citation in GitHub using Citation File Format (CFF) (Druskat, n.d.) in connection with Zenodo.

## Work programme

The work programme whe devised outlines the research carried out, the steps taken, and results as data and documentation of findings. The programme is a demonstration of the same processes of semantic enrichment of literature that will make a significant contribution to the future work of creating a knowledge graph for UN Climate literature, as well as the expected results that infomatic searches can perform on such textual corpora. We are not presenting final results in this article, in other words, this article is not a representative analysis of the state of the art of the frequency of research on the topic of climate justice. Instead this article reflects an indicative scoping exercise and conceptual modelling to show what the tooling is capable of, the steps of what’s involved, its scale, and terms.

An open-notebook science log has been made for the complete work programme on [GitHub Discussion](https://github.com/semanticClimate/ClimateKG/discussions/1), [GitHub Project (Issues)](https://github.com/orgs/semanticClimate/projects/8), and the [Jupyter Notebook](https://colab.research.google.com/drive/1OFNPR9bQCOWtbOzOCmNEImF0yQQEW1is?usp=sharing) *‘#semanticClimate Tools’*. (Bhadra, Kumari, and Murray-Rust 2024) All data was stored on GitHub and made available according to FAIR Principles. The work programme covered the following steps to address the four sub questions mentioned earlier. First of all, we created a GoogleColab Jupyter Notebook to search Europe PMC for the climate justice term and retrieve literature in 3 year periods from 2004 to 2024. We retained the literature and data and reported on the findings. Secondly, we carried out data analysis on this mini-corpus (the papers downloaded from Europe PMC) to find associated terms. Thirdly we searched semantified versions of IPCC reports held on by #semanticClimate on GitHub: *IPCC AR6* and the *IPCC Glossary* and again reported on findings. Fourthly we searched Wikipedia, Wikidata, and Wiktionary for all associated terms and annotated all terms with results. And finally, we created the Climate Justice Dictionary, which is a word list, which includes all the associated terms from Europe PMC and the IPCC and annotations from Wikidata stored as a linked open data document.

A fully reproducible step-by-step guide as well as associated software and supporting data can be found in Appendix II: Work programme in detail.

The research, code development, and literature search was carried out with participation of all team members who have contributed to authoring the article.

## Conclusion

### What were the outcomes of the scoping exercise we carried out of a semi-automated literature review on climate justice, and why is this so important as a new model to publish and to make climate research more accessible? Below are some of the findings that demonstrate the use of the #semanticClimate tooling and working process. They should not be taken as definitive results on the use of climate justice in the corpora being examined but instead as a modelling exercise and an example of the indicative capabilities of the process.

The semi-automated literature review can be used on any topic and in different domains. See: ‘Appendix II: Work programme in details’, for how the results were gained. In relation to the ‘climate justice’ term search these are the results: in the open access scholarly literature corpus from Europe PMC the data showed that there was no publication on climate justice from 2004 to 2009. From 2010 onwards, the publication for climate justice articles has increased exponentially. In the most recent 3 year period, 2022-24, a significant increase occurred to 574 papers. (Table 4)

Carrying out data analysis using word frequency analysis the researchers looked for associated terms for climate justice on Europe PMC.DataTables are used to display a list of retrieved articles that use the term ([list of 100 articles from 2010-2024](https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/literature_search2010-2024/datatables2010-2024.html)) (Figure 4) this allows the researcher to have an overview of journal names, authors, etc., which helps the researcher evaluate the results in a convenient way. Additionally, the DataTables include a list of countries occurring in the articles for the year 2010-2024. (Figure 5)

A word list was generated from the Europe PMC search of climate justice related terms in Europe PMC 100 articles from 2010-2024, of 25 words. See the list here: [words\_from\_lit\_corpus.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_from_lit_corpus.txt) (‘Appendix III: Supplementary material ‘Europe PMC semi-automated literature search Jupyter Notebook ‘Climate Justice Demo #semanticClimate Tools’), for a step-by-step reproducible process of the word list generation. This word list of associated terms helps improve the view on the wider discourse around climate justice.

Searching the IPCC *AR6* semantified version held by #semanticClimate on Github and the *IPCC Glossary* GitHub version, identified the following list chapters that include the term ‘climate justice’, (Figure 6). A word list was generated [‘*Word list IPCC Report – 51 terms’*: *Climate justice*](https://github.com/semanticClimate/JEP-article/blob/main/data/words_IPCC_chao08_chap18.txt) related terms from the IPCC *AR6*/WG2 Report – Chapter [08](https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/test/resources/ipcc/cleaned_content/wg2/Chapter08/html_with_ids.html) and [18](https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/test/resources/ipcc/cleaned_content/wg2/Chapter18/html_with_ids.html))’. See: Word list: [words\_IPCC\_chao08\_chap18.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_IPCC_chao08_chap18.txt).

Finally the *Climate Justice Dictionary* was made combining the two word list to create 76 terms. These were used to search Wikipedia, Wikidata, and Wiktionary for all associated terms; and, then creating the annotations stored as linked open data: (Figure 8) and (Figure 9). [The Climate Justice Dictionary](https://github.com/semanticClimate/JEP-article/blob/main/data/words_all_climatejustice.txt) (word list) and [Climate Justice Dictionary Annotated](https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/data/climatejustice_dictionary.html) (word list, Wikipedia enhanced). The enhanced Climate Justice Dictionary shows how existing data can be retrieved and used from the Wikimedia platforms.

The results from the semi-automated literature review demonstrate how data and knowledge hidden in a corpus can be made explicit. Our intention is to show the value of semantic electronic publishing in actually being useful to wider audiences and how if recorded in a database, here a knowledge graph, then the data and knowledge can be given a computable surface of sorts. The frequency of the term ‘climate justice’ in the 6 million open access articles from Europe PMC allowed patterns to be revealed, such as the temporal occurrence from zero mentions of ‘climate justice’ in 2004 to 574 mentions in papers for 2022-24. Revealing the associated terms for ‘climate justice’ Europe PMC enabled the ability to search further for journal names and authors, which contributes to locating the discourse around the topic.

In the context of a more narrow corpus such as the IPCC *AR6* report, chapters could be more easily identified that were relevant to the ‘climate justice’ topic. For our research question about ‘locating climate justice discourse in scholarly literature and the IPCC reports the Climate Justice Dictionary we created concludes the work by being a proof of concept for what can be used from Wikiverse, here this was images and plain language descriptions in the annotated version of the dictionary. The Climate Justice Dictionary is something that is in development and has had to be made by running a number of software programmes on the command line and on Google CoLab Jupyter Notebooks, with researchers evaluating results, but it intended as a proof of concept and as a pointer to the types of results a knowledge graph could generate on the fly in the way that web search engines function.

### Future work

As we have argued in this article, building the Climate Knowledge Graph is important as it is a way to create greater accessibility, for a wider audience, of the IPCC Assessment Reports.

The semi-automated literature review, as well as the literature retrieval and corpus conversion from PDF to HTML, are the building blocks of the ClimateKG. Each step in the workflow from the semi-automated literature review will also be used to add linked open data (LOD) to a Wikibase instance. This is what will be carried out for the ClimateKG when adding the *AR6* corpus and other corpora, with Wikibase being used as the LOD data storage for the knowledge graph project:

* Adding paragraph with IDs of the *AR6* corpus
* Adding IDs of all other significant objects in the *AR6* corpus literature
* Occurrence of a terms in corpus literature
* Occurrences of associated terms in corpus literature
* Term annotations from: Wikidata, Wikipedia, and Wiktionary
* Annotation of the corpus using the #semanticClimate Dictionaries

The combination of the above will be used to allow search and literature retrieval from the knowledge graph that then can be published as multi-format publications that adhere to the Linked Open Document Checklist and to the 5-Star deployment scheme for data. During 2025, the #semanticClimate community in partnership with TIB – Open Science Labs’ Open Science Data Services and Leibniz Joint Lab Future Libraries & Research Data, ORKG, and Lab Knowledge Infrastructures – and NIPGR will be building in these workflows to the ClimateKG.

## References

All citations stored here as open citations on Zotero group tagged as #jep <https://www.zotero.org/groups/2437020/semanticclimate/collections/KD8WW5YA>

Agrawala, Shardul. 1998. ‘Structural and Process History of the Intergovernmental Panel on Climate Change. Climatic Change’. *Climatic Change* 39 (4): 621–42. https://doi.org/10.1023/A:1005312331477.

Berners-Lee, Tim. 2011. ‘5-Star Open Data’. 2011. http://5stardata.info/en/.

Bhadra, Parijat, Renu Kumari, and Peter Murray-Rust. 2024. ‘#semanticClimate Tools Demo - CoLab Notebook’. Zenodo. https://doi.org/10.5281/ZENODO.14176447.

Bush, Vannevar. 1945. ‘As We May Think’. The Atlantic. 1 July 1945. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Butler, Leigh-Ann, Lisa Matthias, Marc-André Simard, Philippe Mongeon, and Stefanie Haustein. 2023. ‘The Oligopoly’s Shift to Open Access: How the Big Five Academic Publishers Profit from Article Processing Charges’. *Quantitative Science Studies* 4 (4): 778–99. https://doi.org/10.1162/qss\_a\_00272.

Calvin, Katherine, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter W. Thorne, Christopher Trisos, José Romero, et al. 2023. ‘IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland.’ First. Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647.

Capadisli, Sarven. 2019. ‘Linked Research on the Decentralised Web’. https://csarven.ca/linked-research-decentralised-web.

Chowdhary, Reema. 2024. ‘Shifting Paradigms of Multilingual Publishing and Scholarship in India’. *The Journal of Electronic Publishing* 27 (1). https://doi.org/10.3998/jep.5592.

Druskat, Stephan. n.d. ‘Citation File Format (CFF)’. Citation File Format (CFF). Accessed 20 November 2024. https://citation-file-format.github.io/.

‘FAIR Principles’. n.d. GO FAIR. Accessed 5 December 2024. https://www.go-fair.org/fair-principles/.

GOFAIR. 2022. ‘FAIR Principles’. GO FAIR. 2022. https://www.go-fair.org/fair-principles/.

‘Google Knowledge Graph’. 2024. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Google\_Knowledge\_Graph&oldid=1250126393.

‘Graphviz’. n.d. Graphviz. Accessed 20 November 2024. https://graphviz.org/.

GreyNet. 2024. ‘Grey Literature - GreyNet, Grey Literature Network Service’. 2024. https://greynet.org/home/aboutgreynet.html.

Hughes, Hannah. 2024. *The IPCC and the Politics of Writing Climate Change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009341554.

ICJ. 2024a. ‘Obligations of States in Respect of Climate Change - Verbatim Record 2024/38’. 3 December 2024. https://www.icj-cij.org/case/187.

———. 2024b. ‘Press Release: Obligations of States in Respect of Climate Change’. 26 November 2024. https://www.icj-cij.org/sites/default/files/case-related/187/187-20241126-pre-01-00-en.pdf.

Intergovernmental Panel On Climate Change. 2022. *Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009157988.

Intergovernmental Panel on Climate Change (IPCC), ed. 2022. ‘Summary for Policymakers’. In *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-Industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*, 1–24. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009157940.001.

Intergovernmental Panel On Climate Change (Ipcc). 2022. *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009157964.

———. 2023a. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009157896.

———. 2023b. *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009325844.

———, ed. 2023c. *Climate Change 2022 - Mitigation of Climate Change: Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009157926.

IPCC. 1988. ‘Reports — IPCC’. 1988. https://www.ipcc.ch/reports/.

———. 1992. ‘Climate Change: The IPCC 1990 and 1992 Assessments — IPCC’. https://www.ipcc.ch/report/climate-change-the-ipcc-1990-and-1992-assessments/.

———. 2013. ‘PROCEDURES FOR THE PREPARATION, REVIEW, ACCEPTANCE, ADOPTION, APPROVAL AND PUBLICATION OF IPCC REPORTS’. https://archive.ipcc.ch/pdf/ipcc-principles/ipcc-principles-appendix-a-final.pdf.

———. 2018. ‘Global Warming of 1.5°C’. 2018. https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15\_Headline-statements.pdf.

———. 2019. ‘Available Data for AR6’. https://ipcc-data.org/ar6landing.html.

Ipcc. 2022. *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-Industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009157940.

IPCC. 2023. ‘Sixth Assessment Report — IPCC’. https://www.ipcc.ch/assessment-report/ar6/.

———. n.d.-a. ‘About the IPCC’. Accessed 4 December 2024. https://www.ipcc.ch/about/.

———. n.d.-b. ‘IPCC Glossary’. Accessed 6 November 2024. https://apps.ipcc.ch/glossary/.

‘IPCC Glossary Search’. n.d. Accessed 30 November 2024. https://apps.ipcc.ch/glossary/.

‘IPCC-WG1’. n.d. GitHub. Accessed 5 December 2024. https://github.com/IPCC-WG1.

Jylhä, Kirsti M., and Kahl Hellmer. 2020. ‘Right-Wing Populism and Climate Change Denial : The Roles of Exclusionary and Anti-Egalitarian Preferences, Conservative Ideology, and Antiestablishment Attitudes’. *Analyses of Social Issues and Public Policy* 20 (1): 315–35.

Kay, Alan C. 1972. ‘A Personal Computer for Children of All Ages’. In *Proceedings of the ACM Annual Conference - Volume 1*. Vol. 1. ACM ’72. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/800193.1971922.

Krajewski, Markus, and Peter Krapp. 2011. *Paper Machines: About Cards & Catalogs, 1548-1929*. History and Foundations of Information Science. Cambridge, Mass: MIT Press.

Kuhn, Tobias, and Michel Dumontier. 2019. ‘FAIR Vocabulary’. 2019. https://peta-pico.github.io/FAIR-nanopubs/principles/index-en.html#https://w3id.org/fair/principles/terms/FAIR-SubPrinciple.

Latour, Bruno. 2018. *Down to Earth: Politics in the New Climatic Regime*. English edition. Cambridge, UK ; Medford, MA: Polity Press. https://hscif.org/wp-content/uploads/2018/04/Latour-Bruno-Downt-to-Earthsmall.pdf.

Licklider, J. C. R. 1965. *Libraries of the Future*. Cambridge, Mass., M.I.T. Press. http://archive.org/details/librariesoffutur00lickuoft.

Milfont, Taciano L., Paul G. Bain, Yoshihisa Kashima, Victor Corral-Verdugo, Carlota Pasquali, Lars-Olof Johansson, Yanjun Guan, et al. 2018. ‘On the Relation Between Social Dominance Orientation and Environmentalism: A 25-Nation Study’. *Social Psychological and Personality Science* 9 (7): 802–14. https://doi.org/10.1177/1948550617722832.

Nelson, Theodor H. 1974. *Computer Lib*. 9. print. South Bend, Ind: The Distributors.

‘Open Access Subset - Europe PMC’. n.d. Accessed 8 November 2024. https://europepmc.org/downloads/openaccess.

‘OpenStreetMap’. 2024. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=OpenStreetMap&oldid=1261023652.

‘Papers with Code - The Latest in Machine Learning’. n.d. Accessed 16 November 2024. https://paperswithcode.com/.

Pearce, D W. 1996. ‘The Social Costs of Climate Change: Greenhouse Damage and the Benefits of Control’. https://www.ipcc.ch/site/assets/uploads/2018/03/ipcc\_sar\_wg\_III\_full\_report.pdf.

Pooley, Jeff. 2022. ‘Surveillance Publishing’. *The Journal of Electronic Publishing* 25 (1). https://doi.org/10.3998/jep.1874.

Raju, Reggie, and Auliya Badrudeen. 2022. ‘Social Justice Driving Open Access Publishing: An African Perspective’. *The Journal of Electronic Publishing* 25 (1). https://doi.org/10.3998/jep.1910.

‘RDF - Semantic Web Standards’. n.d. Accessed 20 December 2024. https://www.w3.org/RDF/.

#semanticClimate. 2019. ‘#semanticClimate Software Repositories on GitHub’. https://github.com/orgs/semanticClimate/repositories?

———. 2023. ‘IPCC Reports and City Climate Change Plans: Proof of Concept Prototype - Open Climate Reader’. 4 August 2023. https://semanticclimate.github.io/city-open-climate-reader/.

———. n.d. ‘IPCC/AR6 as HTML’. GitHub. Accessed 9 November 2024. https://github.com/semanticClimate/ipcc/tree/main/cleaned\_content.

‘#semanticClimate Software Repositories on Zenodo’. n.d. https://zenodo.org/communities/semanticclimate/records?q=&l=list&p=1&s=10&sort=newest.

Sobral, Sonia Rolland. 2020. ‘Mobile Learning in Higher Education: A Bibliometric Review’. *International Journal of Interactive Mobile Technologies (iJIM)* 14 (11): 153. https://doi.org/10.3991/ijim.v14i11.13973.

‘spaCy · Industrial-Strength Natural Language Processing in Python’. n.d. Accessed 6 December 2024. https://spacy.io/.

Stocker, Markus, Lauren Snyder, Matthew Anfuso, Oliver Ludwig, Freya Thießen, Kheir Eddine Farfar, Muhammad Haris, Allard Oelen, and Mohamad Yaser Jaradeh. 2024. ‘Rethinking the Production and Publication of Machine-Reusable Expressions of Research Findings’. arXiv. https://doi.org/10.48550/ARXIV.2405.13129.

Stockhause, Martina, Martin Juckes, Robert Chen, Wilfran Moufouma Okia, Anna Pirani, Tim Waterfield, Xiaoshi Xing, and Rorie Edmunds. 2019. ‘Data Distribution Centre Support for the IPCC Sixth Assessment’. *Data Science Journal* 18 (June):20. https://doi.org/10.5334/dsj-2019-020.

Tamburrini, Guglielmo. 2022. ‘The AI Carbon Footprint and Responsibilities of AI Scientists’. *Philosophies* 7 (1): 4. https://doi.org/10.3390/philosophies7010004.

‘Tim Berners-Lee’. 2024. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Tim\_Berners-Lee&oldid=1253967594.

UN. 2015. ‘The Sustainable Development Goals’. 2015. https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf?ref.

———. 2023. ‘Secretary-General Calls on States to Tackle Climate Change “Time Bomb” through New Solidarity Pact, Acceleration Agenda, at Launch of Intergovernmental Panel Report | Meetings Coverage and Press Releases’. https://press.un.org/en/2023/sgsm21730.doc.htm.

UNDP. 2023. ‘Climate Change Is a Matter of Justice – Here’s Why’. UNDP Climate Promise. June 2023. https://climatepromise.undp.org/news-and-stories/climate-change-matter-justice-heres-why.

UNESCO. 2021. ‘UNESCO Recommendation on Open Science’. UNESCO. https://doi.org/10.54677/MNMH8546.

UNFCCC. 1997. ‘Kyoto Protocol to the United Nations Framework Convention on Climate Change. | UNFCCC’. 10 December 1997. https://unfccc.int/documents/2409.

———. 2016. ‘The Paris Agreement’. 2016. https://unfccc.int/process-and-meetings/the-paris-agreement.

———. 2022. ‘Decision 1/CMA.3. Glasgow Climate Pact’. 2022. https://unfccc.int/sites/default/files/resource/cma2021\_10a01E.pdf#page=1.08.

———. n.d.-a. ‘Conference of the Parties (COP)’. Accessed 1 December 2024. https://unfccc.int/process/bodies/supreme-bodies/conference-of-the-parties-cop.

———. n.d.-b. ‘What Is the Kyoto Protocol? | UNFCCC’. Accessed 1 December 2024. https://unfccc.int/kyoto\_protocol.

Wei, Yi-Ming, Rong Han, Qiao-Mei Liang, Bi-Ying Yu, Yun-Fei Yao, Mei-Mei Xue, Kun Zhang, et al. 2018. ‘An Integrated Assessment of INDCs under Shared Socioeconomic Pathways: An Implementation of C3IAM’. *Natural Hazards* 92 (2): 585–618. https://doi.org/10.1007/s11069-018-3297-9.

Wikipedia. 2023. ‘Open-Notebook Science’. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Open-notebook\_science&oldid=1188104495.

———. 2024a. ‘Graph Database’. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Graph\_database&oldid=1258574566.

———. 2024b. ‘US Withdrawal from Paris Agreement’. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=United\_States\_withdrawal\_from\_the\_Paris\_Agreement&oldid=1257286872.

Willighagen, Egon, and Lars Willighagen. 2023. ‘IPCC Wikibase Queries (Code)’. IPCC-Queries. 2023. https://semanticclimate.github.io/IPCC-Queries/.

Wright, Alex. 2014. *Cataloging the World: Paul Otlet and the Birth of the Information Age*. Oxford New York: Oxford University Press.

Yadav, Gitanjali, Shweata Hegde, Renu Kumari, Neeraj Kumari, Peter Murray-Rust, and Simon Worthington. 2024. ‘The #Semantic Climate Community: Making Open-Source Software for Knowledge Liberation’. *Annals of Library and Information Studies*, December. https://doi.org/10.56042/alis.v71i4.14285.

## Appendices

### Appendix I

### Resources used

* Data processing software tools
* Dictionaries
* Literature corpora and semantified copies
* Annotation and indexing resources

#### Data processing software tools

This is a table of the open-source software for data processing used by #semanticClimate for the literature analysis.

##### Software

A full list of all #semanticClimate software can be found on [GitHub](https://github.com/orgs/semanticClimate/repositories?) (#semanticClimate 2019) and on the #semanticClimate [Zenodo community](https://zenodo.org/communities/semanticclimate/). (‘#semanticClimate Software Repositories on Zenodo’, n.d.)

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Description** | **Code** | **Support** |
| pygetpapers | Searches and downloads articles from repositories. Standalone, but the results may be used by ‘docanalysis’ or possibly ‘pyamiimage’. Can be called from other tools. | <https://github.com/petermr/pygetpapers> | <https://pygetpapers.readthedocs.io/> |
| amilib | It is a Python library designed for document processing, and dictionary creation. Python library of ami software especially NLP, HTML, downloading and related convenience utilities  ‘amilib’ has tools for finding, cleaning, converting, searching, and republishing legacy documents. | <https://github.com/petermr/amilib> | README |
| Py4ami (bundled in amilib) | Translation of ‘ami3(J)’ to Python. Processes CProjects\*  to extract and combine primitives into semantic objects. Some functionality overlaps with ‘docanalysis’ and ‘pyamiimage’. Includes libraries (e.g., for Wikimedia) and includes prototype GUI in Tkinter, and a complex structure of word-dictionaries covering science and related disciplines. (Note the project is called ‘pyami’ locally but there is already a PyAMI project so there it is called ‘py4ami’) | <https://github.com/petermr/pyami> | README |
| pyamihtml (bundled in amilib) | Conversion of documents to styled HTML | <https://github.com/petermr/pyamihtml> | <https://docs.google.com/document/d/1CF68Fw9NytnUH2ZAEecpUeligXczhn4A/edit> |
| docanalysis | Ingests CProjects and carries out text-analysis of documents, including sectioning, NLP/text-mining, vocabulary generation. Uses NLTK and other Python tools for many operations, and spaCy, scispaCy for annotation of entities. Outputs summary data, correlations, word-dictionaries. Links entities to Wikidata. | <https://github.com/petermr/docanalysis> | README |
| pyamiimage | Ingests figures/images, applies many image processing techniques (erode-dilate, colour quantization, skeletons, etc.), extracts words (Tesseract), extracts lines and symbols (uses sknw/NetworkX) and recreates semantic diagrams (not finished) | <https://github.com/petermr/pyamiimage> | README |
| Dictionaries | Collection of Wikidata-based dictionaries for scientific annotation and searching | <https://github.com/petermr/dictionary/> | General: <https://github.com/petermr/dictionary/> and Creating dictionaries: <https://github.com/petermr/tigr2ess/blob/master/dictionaries/TUTORIAL.md> |
| amilib | IPCC Glossary download  Code : <https://github.com/petermr/pyamihtml/blob/main/test/test_headless.py> | <https://github.com/petermr/semanticClimate/tree/main/ipcc/ar6/test/total_glossary/output> | <https://github.com/petermr/amilib/blob/main/test/test_headless.py> |
| amilib | IPCC/AR6 Chapters download  Code: <https://github.com/petermr/pyamihtml/blob/main/test/test_headless.py> | <https://github.com/petermr/amilib/tree/main/test/resources/ipcc/cleaned_content> | <https://github.com/petermr/amilib/blob/main/test/test_headless.py> |

\*CProject is a Corpus Project. This is the name for configuration, initialisation and storage of a literature search and analysis project.

Table 3: Software for data processing used by #semanticClimate

#### Dictionaries

The dictionaries are used for searching and annotating the literature corpora.

* Carbon cycle dictionary

<https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/temp/dictionary/climate/carbon_cycle.html>

* AR6/WG1/Chap03 dictionary

<https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/data/wg1chap03_dictionary.html>

* Various dictionaries created at time of COVID Pandemic: country, disease, drug, npi, organization, test\_trace, virus, Zoonosis  
  <https://github.com/petermr/dictionary/tree/main/openVirus20210120>

#### Literature corpora and semantified copies

Two corpora are used in the research: Europe PMC, and *the IPCC – Sixth Assessment Report.* Firstly, Europe PMC is used as it is a key open academic aggregator and repository. The literature repository holds a significant amount of open access (OA) articles mainly in the bio sciences, 6.5 million with 47.7% of total articles in 2023 being OA as opposed to only 10% being OA when the service started in 2005. (‘Open Access Subset - Europe PMC’, n.d.) Additionally Europe PMC has encouraged preprints since the start of the COVID pandemic which enables early access to research papers and to cutting edge science. Europe PMC provides articles as full text as well supporting a wide range of computational machine access. The #semanticClimate community is able to carry out automated search and document retrieval using its software ‘pygetpapers*’*. Searches can target specific paper parts, e.g., abstracts, findings, or conclusions, etc., and retrieval of papers can be as full text JATS format.

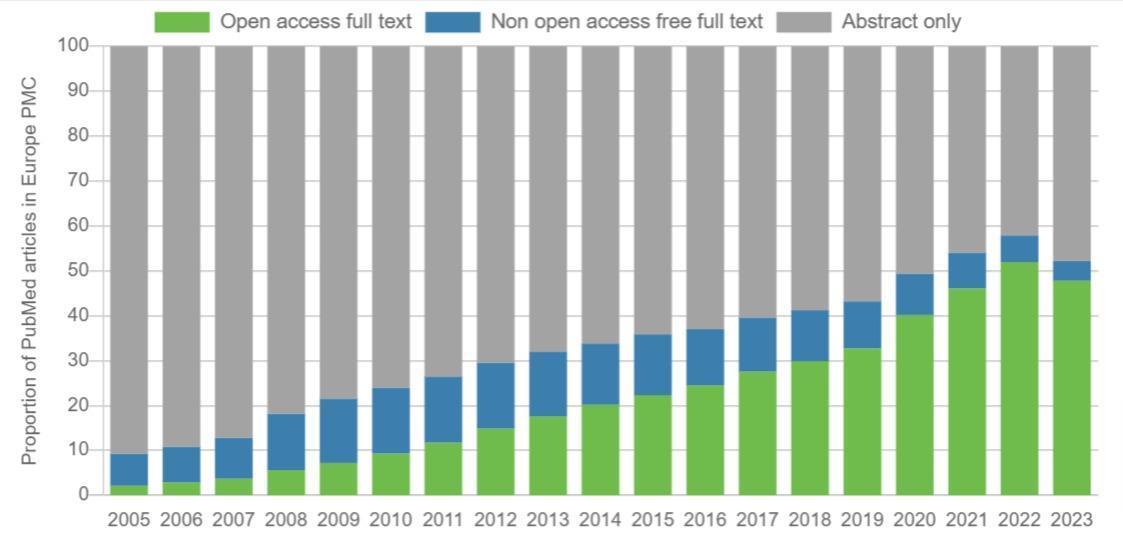


Figure 2: Europe PMC [per cent open access](https://europepmc.org/downloads/openaccess).

##### *IPCC – Sixth Assessment Report*

##### The second resource used was the UN Climate literature corpus for the *IPCC – Sixth Assessment Report (AR6)*, including its associated glossary. ‘The IPCC prepares comprehensive Assessment Reports about knowledge on climate change, its causes, potential impacts and response options.’. (IPCC, n.d.-a) The IPCC Reports are the gold standard for climate science and policy and their importance for understanding and addressing climate change cannot be overemphasized. As UN Secretary-General António Guterres has stated, ‘it is a survival guide for humanity. As it shows, the 1.5°C limit is achievable. But it will take a quantum leap in climate action. This report is a clarion call to massively fast-track climate efforts by every country and every sector and on every timeframe. In short, our world needs climate action on all fronts — everything, everywhere, all at once.’. (UN 2023)

*AR6* is produced over several years, known as the Assessment Cycle, and consists of a number of reports made by working groups, including a synthesis report, working group reports, and special reports.

###### Sixth Assessment Cycle Reports and IPCC Glossary

* ***Synthesis Report (SYR)*** (Standalone end of Assessment Cycle report)
  + [*Climate Change 2023: Synthesis Report*](https://doi.org/10.59327/IPCC/AR6-9789291691647). Contribution of Working Groups I, II and III to the *Sixth Assessment Report* of the Intergovernmental Panel on Climate Change. DOI: 10.59327/IPCC/AR6-9789291691647 (Not open licenced) (Calvin et al. 2023)
    - [*IPCC Glossary*](https://apps.ipcc.ch/glossary/) *(AR6 and Assessment Report 5 (AR5))*: IPCC, 2023: Annex I: Glossary (Calvin et al. 2023) and on the IPCC website. DOI: various (Not open licenced) (IPCC, n.d.-b)
* **Working Group reports** (these feed into the *Synthesis Report*)
  + [*Climate Change 2021: The Physical Science Basis*](https://doi.org/10.1017/9781009157896). Working Group I Contribution to the *IPCC Sixth Assessment Report*. DOI: 10.1017/9781009157896 (CC-BY-NC-ND 4.0) (Intergovernmental Panel On Climate Change (Ipcc) 2023a)
  + [*Climate Change 2022: Impacts, Adaptation and Vulnerability*](https://doi.org/10.1017/9781009325844). Working Group II Contribution to the *IPCC Sixth Assessment Report*. DOI: 10.1017/9781009325844 (CC-BY-NC-ND 4.0) (Intergovernmental Panel On Climate Change (Ipcc) 2023b)
  + [*Climate Change 2022: Mitigation of Climate Change*](https://doi.org/10.1017/9781009157926). Working Group III Contribution to the *IPCC Sixth Assessment Report*. DOI: 10.1017/9781009157926 (CC-BY-NC-ND 4.0) (Intergovernmental Panel On Climate Change (Ipcc) 2023c)
* **Special Reports** (these feed into the *Synthesis Report*)
  + [*Global Warming of 1.5°C*](https://doi.org/10.1017/9781009157940): IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty. DOI: 10.1017/9781009157940 (CC-BY-NC-ND 4.0) (Ipcc 2022)
  + [*Climate Change and Land*](https://doi.org/10.1017/9781009157988): IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. DOI: 10.1017/9781009157988 (CC-BY-NC-ND 4.0) (Intergovernmental Panel On Climate Change 2022)
  + [*The Ocean and Cryosphere in a Changing Climate*](https://doi.org/10.1017/9781009157964): Special Report of the Intergovernmental Panel on Climate Change. DOI: 10.1017/9781009157964 (CC-BY-NC-ND 4.0) (Intergovernmental Panel On Climate Change (Ipcc) 2022)

Unlike the research papers in Europe PMC the *AR6* report is not published in a modern open science and FAIR Principles compliant way. Not all reports are open access, it is not available as full text, it is not held in a single open scientific literature repository, and does not provide computational machine access. It is worth noting that the working group’s data publication is different from the report that has been published, data is being published following modern open science methods, using Git versioning, and FAIR Principles. (Stockhause et al. 2019; ‘IPCC-WG1’, n.d.; IPCC 2019) (IPCC 2019; IPCC WG1 2024; Stockhause et al. 2019). The #semanticClimate community has made semantified copies of *AR6 Climate Change 2023: Synthesis Report* and *IPCC Glossary* using its software tooling, including [70 chapters of *IPCC AR6*](https://github.com/semanticClimate/ipcc/tree/main/cleaned_content) as annotated HTML with IDs per paragraph (work in progress) (#semanticClimate, n.d.) and the [*IPCC Glossary*](https://github.com/petermr/semanticClimate/tree/main/ipcc/ar6/test/total_glossary/output) *AR6* and *AR5* as HTML and as linked open data in Wikibase. (Willighagen and Willighagen 2023)

#### Annotation and indexing resources

The following from the Wikimedia technical ecology and linked open data resources have been used for annotating literature. Annotation in this context is to add additional information for example from Wikipedia and Wiktionary, or as ‘semantic labelling’ with linked open data from Wikidata.

Using these annotations the following features were enabled: Definitions, multilingual definitions, image illustrations, and linked open data information, etc. The #semanticClimate tooling can access these resources via API access, scripting its tools in Jupyter Notebooks, etc.

* Wikipedia: <https://www.wikipedia.org/>,
* Wiktionary: <https://www.wiktionary.org/>, and
* Wikidata: <https://www.wikidata.org/>.

### Appendix II: Work programme in detail

#### A literature review/search for the term ‘climate justice’ using Europe PMC repository: A how-to

Key: Courier font denotes the #semanticClimate software name being used in the process; Bold courier font denotes commands; yellow boxes are commands; blue boxes are outputs.

Link to the [Jupyter Notebook](https://colab.research.google.com/drive/1OFNPR9bQCOWtbOzOCmNEImF0yQQEW1is?usp=sharing) *‘Climate Justice Demo #semanticClimate Tools’* running on Google Colab.

Complete work programme on [GitHub Discussion](https://github.com/semanticClimate/ClimateKG/discussions/1), [GitHub Project (Issues)](https://github.com/orgs/semanticClimate/projects/8), and the [Jupyter Notebook](https://colab.research.google.com/drive/1OFNPR9bQCOWtbOzOCmNEImF0yQQEW1is?usp=sharing) *‘#semanticClimate Tools’*. (Bhadra, Kumari, and Murray-Rust 2024)

The Jupyter Notebook contains an executable version of the code for this section and can be used to re-run the process as described below or be used as a template for performing new term searches.

A literature review on a term is essential before doing any research on the topic. There are many publications being constantly released every day and these are saved on academic repositories. The tools listed below are used to retrieve the articles from academic repositories.

Note: The literature search tooling has to carry out this work retrospectively, either annotating the documents or adding them to a knowledge graph. Under the new publishing model being proposed electronic publishing would be produced in this way from the start.(Stocker et al. 2024)

The following are the descriptions for retrieving articles and extracting information from them. This has been divided into three sections:

1. Searching literature using a query term with pygetpapers
2. Creating tables using [DataTables](https://datatables.net/)(‘DataTables | Javascript Table Library’, n.d.) for the retrieved articles with amilib
3. Extracting entities (e.g.: COUNTRY) from the literature with docanalysis

### A. Literature search for the term climate justice with pygetpapers

pygetpapers is a #semanticClimate tool to search literature from the Europe PMC scholarly literature repository and other repositories. It makes requests to open access scientific text repositories, analyses the hits, and systematically downloads the articles without further interaction.

These are the steps to use for pygetpapers (Figure 3).

|  |
| --- |
|  |

Figure 3**:** Steps to search literature from Europe PMC (Graphvis [source file](https://github.com/semanticClimate/JEP-article/blob/main/graphviz/EPMC_search.dot))

**Step 1:** It can be installed using the following code in the terminal:

|  |
| --- |
| **pip install pygetpapers** |

**Step 2:** The query term is added in the double quotes in the code mentioned below with following considerations. For example:

* The query can be limited to search within the specific time period with addition of startdate and enddate.
* The use of **-n** will give the number of articles without any download of data.

|  |
| --- |
| **pygetpapers --query '"climate justice"' --xml -n --startdate "2010-01-01" --enddate "2024-10-31" --output fin\_climate\_justice --save\_query** |

This code will search the Europe PMC for open access scholarly literature on the term ‘climate justice’ for the specified time frame.

**Output:**

|  |
| --- |
| **INFO: Total number of hits for the query are 870** |

**Step 3:** The limit to download the data for the desired numbers can be added in the code to get the data for the required number of papers. It is done by removing **-n** and adding **--limit** followed by the number in the code.

Here we will call the downloaded papers the corpus.

|  |
| --- |
| **pygetpapers --query '"climate justice"' --xml --limit 100 --startdate "2010-01-01" --enddate "2024-10-31" --output fin\_climate\_justice --save\_query** |

**Output:**

|  |
| --- |
| **INFO: Total Hits are 870**  **WARNING: Could not find more papers**  **100it [00:00, 194541.00it/s]**  **INFO: Saving XML files to /content/fin\_climate\_justice/\*/fulltext.xml**  **100% 100/100 [01:42<00:00, 1.03s/it]** |

### B. Creating tables using JQuery DataTables for the corpus with amilib

The corpus created for scholarly literature on climate justice is summarised in the form of tables using DataTables software, having important metadata of the corpus which are given below.

* pmcid
* doi
* title
* authorString
* journalInfo.journal.title
* pubYear
* abstractText

These are steps to create tables using DataTables software from the scholarly literature corpus using the tool amilib**.**

**Step 1:** It can be installed using the following code in the terminal:

|  |
| --- |
| **pip install amilib** |

**Step 2:** Create JQuery DataTables for the corpus.

|  |
| --- |
| **amilib HTML --operation DataTables --indir fin\_climate\_justice** |

This code will create DataTables for the number of articles received for the term climate justice.

### C. Extracting entity (COUNTRY) from the scholarly literature corpus using docanalysis

docanalysis is a tool that can extract specific entities from the corpus such as the countries appearing in the articles.

The tool can be installed using the code:

|  |
| --- |
| **pip install docanalysis** |

Once the tool is successfully installed, docanalysis can be used to extract entities like COUNTRY from the corpus. The list of countries are extracted from the specific section of the articles using **—-search\_section** (Key: INT = Introduction; RES = Result; CON = Conclusion; DIS = Discussion).

|  |
| --- |
| **docanalysis --project\_name fin\_climate\_justice --make\_section --search\_section INT, RES, CON, DIS --dictionary COUNTRY --output fin\_climatejustice.csv** |

## Results:

### A. The tool pygetpapers has been used to search literature on climate justice per three years period over a twenty-year span (Table 4).

|  |  |  |
| --- | --- | --- |
| **Start Date** | **End Date** | **No. of Articles with term ‘climate justice’** |
| 2004-01-01 | 2006-12-31 | 0 |
| 2007-01-01 | 2009-12-31 | 0 |
| 2010-01-01 | 2012-12-31 | 4 |
| 2013-01-01 | 2015-12-31 | 20 |
| 2016-01-01 | 2018-12-31 | 37 |
| 2019-01-01 | 2021-12-31 | 235 |
| 2022-01-01 | 2024-10-31 | 574 |

Table 4: Yearwise comparison of publications on climate justice from 2004 to 2024. These are not all ‘publications on climate justice’ but papers retrieved in a conventional search for climate justice.

The table for literature search shows that there was no publication on climate justice from 2004 to 2009. From the 2010 onwards, the publication for climate justice articles has increased exponentially.

### B. JQuery DataTables: Summary of the retrieved articles

The data-table has been created for the articles published on climate justice from 2010-2024. It presents the summary of all the retrieved articles for climate justice. This shows the link to pmcid, doi, and gives a brief abstract also (Figure 4). Data-table supports the ‘human-in-the-loop’ semi-automated part of the literature search in which the researcher can get a [condensed view](https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/literature_search2010-2024/datatables2010-2024.html) of search results and for example see all the journal names.

|  |
| --- |
|  |

Figure 4: Data-table for retrieved articles (list of 100 articles from 2010-2024). The [URL link to the data-table](https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/literature_search2010-2024/datatables2010-2024.html).

### C. The list of countries mentioned in the scholarly literature corpus

The countries mentioned in the papers were extracted withdocanalysis and represented in the word cloud where the size of the word depends on their frequency in the paper. (Figure 5) represents the countries occurring in the corpus of 100 articles from 2010-2024. At present in the scoping exercise only data has been retrieved and presented, false positives have been removed, to a degree. The next step would be for a researcher to theorise why certain countries are present.

|  |
| --- |
|  |

Figure 5:The list of countries occurring in the articles for the year 2010-2024.

## The UN IPCC reports mentioning climate justice

The semantic UN IPCC corpus has been searched for the term climate justice with the use of semanticClimate tool pyamihtmlx, which enabled us to compile a list of the chapters that have mentioned climate justice (Figure 6).

|  |
| --- |
|  |

Figure 6:The list of IPCC reports and chapters including the term climate justice. (Graphviz [source file](https://github.com/semanticClimate/JEP-article/blob/main/graphviz/IPCC_reports.dot))

From the IPCC corpus, we found that the many chapters from Sixth Assessment Report– Working Group II (AR6/WG2) and Sixth Assessment Report – Working Group III (AR6/WG3) along with the *SYR* have highlighted climate justice as key principles within mitigation and adaptation strategies. Climate justice concerns have emerged because of loss and damage from climate hazards to the poorer and vulnerable people who have contributed very little to the overall GHG emission. The success of adaptation strategies depends on equitable development and climate justice. See the following section in the paper for more details: ‘Climate justice and IPCC Sixth Assessment Report’.

## Associated terms and dictionary creation

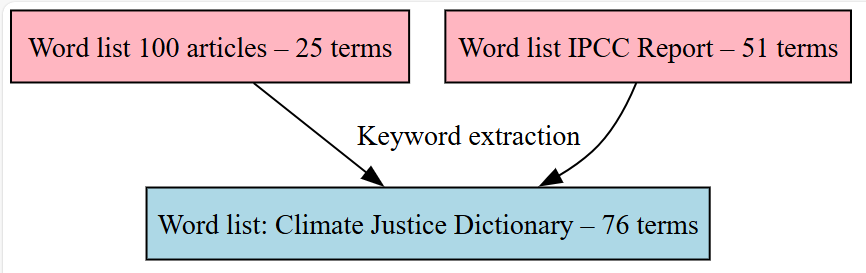
****

Figure 7:Creation of the climate justice wordlist (Graphviz [source file](https://github.com/semanticClimate/JEP-article/blob/main/graphviz/Wordlist_Merge.dot))

### Associated climate justice terms in Europe PMC and IPCC

The list of words/terms that appeared in the literature for climate justice in Europe PMC and the IPCC reports have been extracted withdocanalysis and keyword extractor. The list of words were checked by the research operators to keep the relevant words or phrases to remove false positives.

**Note:** The wordlists have been extracted from IPCC *AR6*/WG2/Chapter 08, IPCC *AR6*/WG2/Chapter 18 and scholarly literature corpus of 100 articles for climate justice.

* ***‘Word list 100 articles*** *–* ***25 terms’***:climate justice related terms in Europe PMC 100 articles from 2010-2024 – List of terms from scholarly literature corpus (Europe PMC).

[words\_from\_lit\_corpus.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_from_lit_corpus.txt)

* ***‘Word list IPCC Report*** *–* ***51 terms’***: Climate justice related terms from the IPCC *AR6*/WG2 Report – Chapter [08](https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/test/resources/ipcc/cleaned_content/wg2/Chapter08/html_with_ids.html) and [18](https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/test/resources/ipcc/cleaned_content/wg2/Chapter18/html_with_ids.html))’

[words\_IPCC\_chao08\_chap18.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_IPCC_chao08_chap18.txt)

Links to #semanticClimate HTML versions of *AR6*/WG2 Report – Chapter 08 and 18

* IPCC/*AR6*/WG2/Chapter 08:

<https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/test/resources/ipcc/cleaned_content/wg2/Chapter08/html_with_ids.html>

* IPCC/*AR6*/WG2/Chapter 18:

<https://html-preview.github.io/?url=https://github.com/petermr/amilib/blob/main/test/resources/ipcc/cleaned_content/wg2/Chapter18/html_with_ids.html>

### Creation of the Climate Justice Dictionary

The dictionary has been created from the list of words from the literature corpus and the IPCC reports. The following is the list of terms used to make the Climate Justice Dictionary.

* *‘****Word list: Climate Justice Dictionary – 76 terms’***: Terms related with climate justice and including climate justice from literature corpus and the IPCC reports’:

[words\_all\_climatejustice.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_all_climatejustice.txt)

## Search IPCC glossary using Climate Justice Dictionary

The list of words used for making the Climate Justice Dictionary has been compared with the *IPCC Glossary* terms and we observed the occurrence of the terms ‘climate justice', and ‘Indigenous Peoples‘ in the glossary.

The dictionary was then further enriched with the information available in Wikipedia, Wikidata, and Wiktionary.

|  |
| --- |
|  |

Figure 8:Flow diagram for creating a dictionary from wordlists. (Graphviz [source file](https://github.com/semanticClimate/JEP-article/blob/main/graphviz/dictionary.dot))

#### Wikipedia enhanced dictionary

As the terms are not easily understandable by lay persons. The information in theClimate Justice Dictionary has been added to for every word/term from Wikipedia. The Wikipedia enhanced dictionary for all the terms was created with the tool amilibusing the following code.

|  |
| --- |
| **amilib DICT --words words\_all\_climatejustice.txt --description wikipedia --figures --dict climatejustice\_dictionary.html --operation create** |

|  |
| --- |
|  |

Figure 9: Wikipedia enhanced dictionary for climate justice associated terms [as web page](https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/data/climatejustice_dictionary.html) from text dictionary: [Words\_all\_climatejustice.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_all_climatejustice.txt)

**The Climate Justice Dictionary:**

Climate Justice Dictionary (Work in Progress): [Words\_all\_climatejustice.txt](https://github.com/semanticClimate/JEP-article/blob/main/data/words_all_climatejustice.txt)

Climate Justice Dictionary (Wikipedia enhanced): [as web page](https://html-preview.github.io/?url=https://github.com/semanticClimate/JEP-article/blob/main/data/climatejustice_dictionary.html)

### Appendix III

### Supplementary material

1. Europe PMC semi-automated literature search [Jupyter Notebook](https://colab.research.google.com/drive/1OFNPR9bQCOWtbOzOCmNEImF0yQQEW1is?usp=sharing) *‘Climate Justice Demo #semanticClimate Tools’*
2. About #semanticClimate Notebook <https://colab.research.google.com/drive/1WUP8IUKvMV14LiOGSvrDMk9k0Oknd9rk>
3. Literature search

<https://github.com/semanticClimate/JEP-article/tree/main/literature_search2010-2024>

1. Figures Graphviz source (all files)

<https://github.com/semanticClimate/JEP-article/tree/main/graphviz>

1. Wordlists for creating the Climate Justice Dictionary

<https://github.com/semanticClimate/JEP-article/tree/main/data>

1. Dictionaries (Drug, diseases, country, virus, etc.)

<https://github.com/petermr/dictionary/tree/main/openVirus20210120>

1. #semanticClimate is an international open research project led by young Indian scientists. Key in the project mission is to make access to scientific climate change knowledge equitable. To achieve its mission #semanticClimate has two areas of activity; firstly, creating software tools for semantic searching of climate change literature, and; secondly, enabling citizen science events, activities, and community building. [↑](#footnote-ref-1)
2. 1.5C Emission Pathway, ‘Emission Pathways and System Transitions Consistent with 1.5°C Global Warming’ are options for ways to limit GHG emissions to reach a 1.5C global warming target. Understanding Global Warming of 1.5°C: Human activities are estimated to have caused approximately 1.0°C of global warming above pre-industrial levels, with a likely range of 0.8°C to 1.2°C. Global warming is likely to reach 1.5°C between 2030 and 2052 if it continues to increase

   at the current rate. (high confidence).(IPCC 2018; Intergovernmental Panel on Climate Change (IPCC) 2022) [↑](#footnote-ref-2)
3. The Wikimedia Foundation produced a large suite of online services and tools. Three of these are Wikipedia, Wikidata, and Wiktionary. Wikipedia is an encyclopedia. Wikidata is a knowledge graph. Wiktionary is a dictionary. [↑](#footnote-ref-3)
4. IPCC pathway frameworks represent likely futures for the earth's climate and the state of climate change, they also represent the effects of mitigating actions. They are represented as surface temperature, e.g., 1.5C, 2C, 4C, etc. [↑](#footnote-ref-4)
5. FAIR Principles are intended for making data human and machine readable. They can also be applied to publications as data. Key: Findability (F), Accessibility (A), Interoperability (I), and Reuse (R). The original 2016 paper can be read here [‘FAIR Guiding Principles for scientific data management and stewardship](https://www.nature.com/articles/sdata201618)’ (Wilkinson et al. 2016) and the GOFAIR page is linked [here](https://www.go-fair.org/fair-principles/) (GOFAIR 2022) for quick reference as well as an [ontology link](https://peta-pico.github.io/FAIR-nanopubs/principles/index-en.html#https://w3id.org/fair/principles/terms/FAIR-SubPrinciple) (Kuhn and Dumontier 2019) [↑](#footnote-ref-5)
6. Linked open data (LOD). Linked open data is data that is openly accessible, usually via a URI and uses open licence, and that makes use of linking to other data. LOD is best expressed via the Resource Description Framework (RDF), see: <https://www.w3.org/RDF/> (‘RDF - Semantic Web Standards’, n.d.) [↑](#footnote-ref-6)
7. The technology used is Wikibase, Jupyter Notebooks, #semanticClimate software, and TIB services: Wikibase4Reseach, Computational Publishing Service, Terminology Services (Antelope), ORKG, Renate (scholarly repository), and PID and Metadata services. [↑](#footnote-ref-7)
8. See our [GitHub](https://github.com/orgs/semanticClimate/repositories) code repositories and Zenodo [#semanticClimate community](https://zenodo.org/communities/semanticclimate/) deposits. [↑](#footnote-ref-8)
9. Persistent Identifiers (PIDs). A long-lasting reference to a digital resource that includes metadata and uses a resolver to connect to the location of the resource. PIDs can be for the following: Publication top-level, and sections or chapters; For data and digital objects – 3D, Images, video, and other media, etc., (DOIs); Other entities including: Persons (ORCID), organisations (ROR), funders (Open Funder Registry), projects (RAID), and events (ConfIDent), etc. [↑](#footnote-ref-9)