

Climate Justice in Electronic Publishing: Supporting Global South Participation in Climate Science Through Semantic Publishing

SIMON WORTHINGTON, GITANJALI YADAV, PETER MURRAY-RUST,
RENU KUMARI, SHWEATA HEGDE, AND PARIJAT BHADRA

Abstract: This article argues that the ways in which scholarly electronic publishing is currently carried out is inherently a climate injustice as it unnecessarily hinders participation by people from the Global South in the climate science discourse, which is further exacerbated by the reliance of publishers on PDF-oriented system architectures. We argue that mega-publishers and societies are responsible for the state of electronic publishing and hence for the resulting climate injustice. A new publishing model for electronic publishing is proposed, informed by the semantic web, hypermedia, and the software system designs of 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 science.

We 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 Intergovernmental Panel on Climate Change (IPCC) climate reports. The IPCC Sixth Assessment Report (AR6) is the most authoritative summation of climate change scientific knowledge and influences the acknowledgment of climate justice globally in policy for addressing climate change. As do other research bodies, the IPCC uses conventional electronic publishing workflows (centered on the PDF), but this holds back its potential wider reach. Our research intends to demonstrate how such reports could be made more accessible.

The research we present in this article includes a semi-automated literature search on the topic of “climate justice” and asks the following questions: What does the open access scholarly corpus know about this 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 have the term and related terms been mentioned? We will also demonstrate the open-source tools that we will use for future work to create the Climate Knowledge Graph (ClimateKG), which aims to make the IPCC report globally accessible.

The research presented in this article covers an experiment carried out by the #semanticClimate team, which focused on searching and computationally retrieving the open access scholarly research corpus from Europe PubMed Central (6.9 million open access papers) as well as all 70 chapters of the AR6 held on GitHub as HTML with IDs. The outcome of this experiment is a first-round scoping exercise to create the Climate Justice Dictionary, which represents terms associated with climate justice collected from this 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: How Scholarly Publishing Exacerbates Climate Injustice

The *Journal of Electronic Publishing* publishes research about contemporary scholarly publishing issues and practices. The argument we put forward in this article for the journal is that scholarly electronic publishing, or the publication process itself as currently practiced, generates climate injustice. This publication process is far from equitable as it is determined by a small number of mega-publishers and some mega-societies who exert significant control and influence over digital publishing infrastructures and access to knowledge (Chen et al. 2019). We suggest that this process is actually fundamentally unjust, particularly for the Global South. As Márton Demeter and Ronina Istratii argue, “Extensive research has demonstrated that the global academy is characterised by a significant Global North–Global South imbalance. Editorial boards, selection committees, funding agencies, publishing houses are predominantly located in Northern high-income societies” (2020, 506).

For the purpose of this article, we have identified three aspects of the digital scholarly publishing process of research articles that contribute most to both epistemic and climate injustice (as we will also outline more in-depth later). The first is fee-paying open access policies. The second aspect is deficiencies in “digital publishing infrastructures” to support open research. Especially problematic are the use of PDF-oriented system architectures and the associated lack of application of technically open methodologies, such as the FAIR principles or the 5-star deployment scheme for data, to

all parts of a publication. Third is the slow progress toward ethical regulatory control of publishing, including open access and open science policy adoption and compliance, to ensure widespread access to knowledge. Our research in this article is concerned predominantly with the second aspect. In other words, we will be looking at “digital publishing infrastructures” and how these can be improved by the adoption of *digital good practice* and the alternative publishing model proposed in this article based on semantic publishing.

Similar to electronic publishing, open access publishing is increasingly controlled by mega-publishers (Butler et al. 2023; Sundell 2021). As Lin Zhang et al. state, “APC [author payment charge] fees collected by DOAJ [Directory of Open Access Journals] journals probably exceeded 1.27 billion US dollars in 2020” (2022, 7654). Here open access 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). Diamond open access is recognized by the open access movement as a more egalitarian open access model, as it charges no APCs and is both “free-to-publish and free-to-read.” The UNESCO Recommendation on Open Science recommends member states support “not-for-profit, academic and scientific community-driven publishing models as a common good” and further encourages “multilingualism in the practice of science, in scientific publications and in academic communications” (UNESCO 2021).

Looking specifically at climate research, what these recommendations reflect is that the Global South currently is not an equal contributor to discussions on climate justice given these fee and language barriers (among others), and they have to accept the agenda and the publishing models, infrastructures, standards, and processes that the Global Northern mega-publishers create and control. This agenda is English-language centric (Chowdhary 2024), which means that many of the countries in the Global South who already suffer disproportionately from the effects of climate change miss out on having their voices heard, both in how knowledge is reported about the climate problems they face and in developing potential mitigations. We believe that this reflects a major aspect of climate injustice in the current publishing system.

This climate injustice is further exacerbated by the fact that the process of publication is not aimed at knowledge transfer but at revenue generation for publishing corporations. This is also reflected in the lack of innovation and investment that publishers have shown in relation to publication technologies, where the predominant technology used in scholarly publishing for the last 25 years has been the portable document format (PDF). PDF is an inefficient way of transmitting knowledge as it is used as a mono-directional form of information transfer, created by publishers and forced upon readers. Readers thus have little say in how they consume climate research, and it is extremely difficult to transform PDFs into any other format, such as a format that does more than

only support reading with human eyes. Converting PDFs into semantic form—that is, making them searchable, translating them into other languages (those who cannot easily read PDFs in English face access barriers to scholarly climate knowledge), and adapting them for different types of readers—is almost impossible.

In general, PDFs are not structured for ease of reading. They do not take account of the fact that in many countries most research consumption is on mobile devices rather than on 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 that need to be rotated by 90 degrees, as well as tables and figures that 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.

To provide some quick practical examples of the issues the PDF format poses, in the open research project #semanticClimate, we address a number of non-academic constituencies both in India and Europe, including high school students (often without English as their first language) and city planners who need climate research to shape regional climate plans (when this research is not available in the kind of form that they can easily consume and poses difficulties for people with limited eyesight, etc.) (Nganji 2015).¹ In relation to the first set of non-academic constituencies, at a visit to Indian schools for the India-UK BBSRC-funded TIGR2ESS program, a project #semanticClimate is involved in, our aim was to meet the young students and encourage them (especially girls) to undertake science education.² The goal of the visit was also to explore possibilities for rural students to undertake short summer internships in Delhi. The students were all engaged in STEAM, including in climate change learning programs, but IPCC reports had not been translated or were not used as part of these programs because they are mostly only available in PDF format (Riley, n.d.). For our second set of non-academic constituencies—city planners developing climate plans who want to use climate literature—their issue mainly concerns locating and sharing derivative data and publications. The PDF format prevents both of these operations. To provide an alternative, we contributed to two prototypes as part of our research: (1) *City - Open Climate Reader* (#semanticClimate 2023), which demonstrated how derivative publications of the IPCC report could be collated and republished, and (2) *Semantic Glossar* (Worthington et al. 2024), which allowed city municipal employees to collaborate on building their own climate glossaries (this prototype was developed

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 more equitable. To achieve its mission, #semanticClimate has two areas of activity: (1) creating software tools for semantic searching of climate change literature and (2) enabling citizen science events, activities, and community building.

2. TIGR2ESS: Transforming India's Green Revolution by Research and Empowerment for Sustainable Food Supplies, accessed April 11, 2025, <https://tigr2ess.globalfood.cam.ac.uk>.

as part of the TH Köln Co-Site project and used #semanticClimate extracted IPCC Glossary and technology).

The research we present in what follows seeks to illustrate the functionalities and possibilities of semantic publishing and is the result of a semi-automated literature search on the topic of “climate justice.” Our research asked the following questions: What does the open access scholarly corpus know about the topic of “climate justice,” 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 been mentioned? The research we present here also functions as 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. We will first discuss the importance of the IPCC reports in the climate science discourse and then outline how the proposed new model allows for improved syntactical and semantic structuring and indexing of publications (including the IPCC reports), which offers various improvements to the Global South’s access to and participation in the climate science discourse. This article should be seen as an indicative scoping exercise and conceptual modeling to show what the #semanticClimate software tooling is capable of. The tooling includes the following modular operations as examples: named entity recognition (NER), corpus creation, text and data mining (TDM), and Wikibase import/export.

The IPCC Reports, Digital Publishing, and Climate Justice

The Intergovernmental Panel on Climate Change (IPCC) is the United Nations body for assessing the science related to climate change. The IPCC has a mandate to “provide governments at all levels with scientific information that can be used in climate policies” (IPCC 1988) and thus 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 as well as the effects of mitigating actions (IPCC, n.d.-b). The actions to achieve viable pathways, such as the 1.5°C emission pathways, are then taken into the United Nations Framework Convention on Climate Change (UNFCCC), a process in which UN member states agree on how to limit climate change.³ The conjoined agencies of the IPCC and UNFCCC operate under a model of Shared

3. 1.5°C emission pathways are options for ways to limit GHG emissions to reach 1.5°C global warming compared to pre-industrial levels. 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, 2022d).

Socioeconomic Pathways (SSPs) (Wei et al. 2018), which inherently focuses on climate justice as a global social justice 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 administration in the United States announced its departure from the Paris Agreement and the decision to take a different approach to climate change (Wikipedia 2024e), which is to abandon the idea of a shared future for humanity, as succinctly hypothesized in Bruno Latour's 2018 book *Down to Earth*. The IPCC reports are thus much more than a report on climate research or the science of climatology and the current status of the planet's climate. Rather, 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 the 32 years since the first Assessment Report was published, the IPCC has evolved its process (IPCC 1992). The open access book *The IPCC and the Politics of Writing Climate Change* by Hannah Hughes (2024) 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 with how the interface of scientific knowledge and society creates political tensions. The study examines the IPCC's editorial workings and sheds light on the IPCC knowledge creation process, providing a useful mapping of the steps involved in the report authoring, the functioning of the report cycles, and how decision-making has changed over time. ClimateKG will incorporate some sample workflow mapping of IPCC report chapters into the knowledge graph creation.

The IPCC as a scientific working group has had to change their authoring process (how they create content) in relationship with and response to climate justice questions. For example, in 1996, the Second Assessment Report, Working Group III (WG3), with an emphasis on economic thinking, lacked a protocol to ensure proper consultation with the UN Observer Organisations and other stakeholders (Hughes 2024). As a result, the following calculation to assess the "social costs" of climate change 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 et al. 1996, 177). The chapter was not approved by the subsequent plenary session. The procedures and rules for subsequent assessment reports were then tightened and codified, such as by assigning chapter lead authors from developing countries (IPCC 2013).

The IPCC reports themselves can be seen as part of a climate justice system of sorts, and the mechanisms of its functioning have evolved over the years. There are also many other agreements and treaties that play significant roles in the process of moving science knowledge into policy and legal frameworks, a few of which merit highlighting here. The initial IPCC Assessment Report in 1992 triggered the UN member states to take action on climate change. The treaty body was formed (the UNFCCC), and

a key task of this body was to establish a framework for how nations would organize themselves to respond to the production of greenhouse gases (GHG), the cause of climate change. This framework was the Kyoto Protocol (UNFCCC 1997, n.d.-b). In 2015, UN countries negotiated the Paris Agreement “to limit the temperature increase to 1.5°C above pre-industrial levels” (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. An example COP treaties clause illustrates how the COP acknowledges climate justice: “*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” (UNFCCC 2022, 2). The IPCC authors, including the chair Jim Skea, have further made a submission to the International Court of Justice (ICJ) on an “advisory opinion” tasked from the court by the UN General Assembly Resolution 77/276 on “Obligations of States in Respect of Climate Change” (ICJ 2024a). It is in these contexts that climate justice has been codified into law: “China submits that developed countries have an obligation to bear their historical responsibilities. IPCC reports reveal that historical emissions from developed countries are the primary cause of the current climate crisis and injustice” (ICJ 2024b).

As we put forward here, the IPCC reports, or the way in which they are made available through digital publishing, are influenced by two key factors: the academic publishing industry and governmental grey literature practices (GreyNet International 2024). First, 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 make up the most authoritative climate science available and are carried out to the highest standards. The data science work of the IPCC has been modernized and applies modern open science methods using FAIR data principles, and data are openly stored on GitHub (GO FAIR, n.d.; IPCC-WG1, n.d.). Second, 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). For the IPCC to achieve this, it has added 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 (AR6 included 195 members).

The IPCC in their Sixth Assessment Report has a specific emphasis on climate justice while addressing climate change. The different volumes of AR6 describe the importance of equitable climate action, integrating local and Indigenous knowledge with scientific approaches for effective climate adaptations and mitigations (IPCC

2023a, 2023b, 2023c).⁴ The IPCC reports have raised the concern of climate justice because of the disproportionate adverse impacts of climatic hazards on vulnerable populations who have contributed less to climate change. Climate vulnerability has also increased due to gender inequalities, so gender considerations are crucial for climate justice (IPCC 2023b, chap. 08). The different findings from the reports have shown the need for climate justice while making adaptation strategies. The Paris Agreement and the 2030 Agenda for Sustainable Development have widened the scope of adaptation governance by linking adaptation to development and climate justice (IPCC 2023b, chap. 18; UNFCCC 2016; United Nations 2015). 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 (WG3, chap. 05). The Civil Society and Social Movements section (IPCC 2023c, chap. 14) focuses on protecting rights, adopting responsibility-based approaches to climate finance, adaptation, and mitigation. IPCC Observer Organizations have also raised questions about the equitable allocations of future climate budgets to developed and developing countries. The *Climate Change 2023: Synthesis Report* (SYR) has mentioned the disproportionate risks from climate hazards such as heat waves, droughts, and flooding to poor and marginalized communities who are less responsible for this adverse climate change (Calvin et al. 2023). The SYR has urged the organization and government to design and structure adaptation and mitigation strategies that should consider equity and justice. Climate justice recommendations mentioned in the volumes of the IPCC report assist in the development of policies that will consider each and every section of society.

A New Publishing Model Based on Semantic Publishing

To further contribute to the climate justice endeavors proposed within the IPCC reports, we propose a new publishing model that could be applied to the publication process of the IPCC reports. This 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 (Wikipedia 2024d; Kay 1972; Nelson 1974; Bush 1945; Knuth 1992; Licklider 1965; Wright 2014; Krajewski 2011). Even with such a long history, semantic

4. The three volumes are *Climate Change 2021—the Physical Science Basis* (WG1), *Climate Change 2022—Impacts, Adaptation and Vulnerability* (WG2), and *Climate Change 2022—Mitigation of Climate Change* (WG3).

publishing is still waiting in the wings and has not reached the mainstream practice of scholarly publishing. The publishing model we put forth therefore represents a new type of publishing that technically enables multilingualism, better global discoverability, more accurate scientific categorization and indexing, and reuse and republishing. In addition, the focus here is on the knowledge being available in academic repositories or corpora as a whole as opposed to prioritizing individual papers or journals. To test this model, we conducted a semi-automated literature review on the topic of climate justice in two corpora of climate scholarship (the Europe PubMed Central [PMC] repository and IPCC's Assessment Round 6 reports) and asked the following questions: 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 an example use case of employing the open-source tools that we then used on a larger scale to create the Climate Knowledge Graph (ClimateKG)—and some of the specifics of knowledge graph capabilities (Worthington 2025). Knowledge graphs are a database that has nodes and properties, and the term “graph” is a mathematical term for “relation” (Wikipedia 2024b). ClimateKG is an online graph database of a literature corpus that can be searched, and the search results can be saved online and collated into a multi-format publication. ClimateKG could be used for frictionless dissemination of the IPCC reports and would include cataloging the literature corpora, word search, content publishing, and reuse.

We searched and retrieved papers and data computationally from the open access repository Europe PMC (2024) and its 6.5 million open access papers, as well as all 70 chapters of the IPCC Sixth Assessment Report held on GitHub as HTML with IDs (*Calvin et al.* 2023). The output of this search showed where the term “climate justice” (in which papers and journals) is being used and creates tables using jQuery DataTables software to browse the collected results that are stored in a mini-corpus.⁵ Data analysis was then carried out to find associated terms, which we used to create the Climate Justice Dictionary, a word list that 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.⁶

Why are knowledge graphs important? And why are they important to improve access to climate justice literature? The scale of scholarly literature and its organization 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 the IPCC reports accessible globally. As argued earlier, the IPCC

5. DataTables, accessed November 12, 2024, <https://datatables.net>.

6. 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, and Wiktionary is a dictionary.

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 such as the web, and as a consequence, their promulgation, reach, and efficacy are limited. The most recent IPCC AR6 report is a substantial corpus itself, and it is the sixth assessment report since the IPCC was created 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 the first decade of the 21st century by the OpenStreetMap (OSM) project (Wikipedia 2024c). This kind of indexing is technically called semantic annotation and uses ontologies, taxonomies, and controlled vocabularies. Concepts within the publication can be tagged, for example, “climatology”⁷ and “climate justice.”⁸ When tagged using Wikidata, a host of trusted source IDs and language translation are available: 97 language translations and 50 identifiers for “climatology” and 40 translations and 8 identifiers for “climate justice.” Many other semantic and syntactic questions can be added as indexed and linked layers to the knowledge graph: IPCC pathways frameworks, use of climate models, data and supporting software used, use of other syntheses papers and their data, and the list goes on.⁹

Currently this syntactical and semantic structuring and indexing has to be done retrospectively until the new publishing model described here and by many others is carried out at the time of authoring and publishing (Stocker et al. 2024; Capadislì 2019; Papers with Code, n.d.).

Digital Publishing Infrastructure: Good Practice

In 2011, Tim Berners-Lee proposed a 5-star deployment scheme for data (see Table 1). Since then, the open science movement has created the infrastructures needed to enable this scheme’s implementation. In data management this has been taken up almost universally with the FAIR principles.¹⁰

7. “Climatology,” accessed April 11, 2025, <https://www.wikidata.org/wiki/Q52139>.

8. “Climate Justice,” accessed April 11, 2025, <https://www.wikidata.org/wiki/Q1291678>.

9. IPCC pathways 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, such as 1.5°C, 2°C, 4°C, etc.

10. FAIR (Findability, Accessibility, Interoperability, and Reuse) principles are intended to make data human and machine readable. They can also be applied to publications as data. The original 2016 paper, “FAIR Guiding Principles for Scientific Data Management and Stewardship” (Wilkinson et al. 2016), and the GO FAIR (n.d.) page (<https://www.go-fair.org/fair-principles/>) are available 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>).

Table 1. 5-star deployment scheme for data (adapted from Berners-Lee 2011)

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

Table 2. Digital publishing infrastructure data parts: dominant scholarly electronic publishing models (based on PDF) compared to the new publishing model and implications for Global South climate science access¹¹

Data part of a publication	Scholarly electronic publishing (PDF)	Semantic publishing (linked open data)	Global South access improvements
1. Metadata	Yes	Yes	
2. Syntactic structure	Poor	Yes	Improved
3. Open citations	No	Yes	Improved
4. Entity tagging	Poor	Yes	Improved
5. Core statements	No	Yes	Improved
6. Open licenses	Yes	Yes	
7. Linked open data	Poor	Yes	Improved
8. Interoperable and computational formats	No	Yes	Improved
9. Open peer review	Yes	Yes	
10. Open bibliometrics	Yes	Yes	
11. Contributor roles and attribution	Yes	Yes	
12. Policy compliance—OA and Open Science	Yes	Yes	

If a publication is understood as being a “digital object,” which is made up of other “digital objects,” then it needs to exist as data and code, and it needs to comply with the FAIR principles to be readable by humans and machines. The difficulty of enabling such “digital objects” that are FAIR compliant is that in many cases adequate systems to publish FAIR-compliant digital objects are at different levels of readiness and maturity or are not yet implemented. Breaking a publication into its data parts remains useful, though, for what might be called the “quality” of the overall publication as a “digital object” (Schöch 2021).

Improving the Data Parts of a Publication to Improve Global South Access and Participation

If we look at the separate data parts of a publication (see Table 2), our research has identified the following improvements that our new publishing model affords:

1. In relation to the publication’s metadata, the new model enables structured and standardized coding of document-related metadata (including bibliographic

11. The data parts of a publication as figured in Table 2 are based on research carried out by the NFDI4Culture—the cultural heritage consortium of the larger German National Research Data Infrastructure Consortium, Data Publications Working Group (Arnold et al. 2023).

information, keywords, license info, persistent identifiers [PIDs] such as Digital Object Identifiers [DOIs]).

2. In relation to the publication's syntactic structure, the new model allows for the explicit coding of text structures (including main text versus notes; introduction, main body, conclusion; if applicable, data, hypotheses, methods, results; author text versus quotations). This offers significant improvements on the existing PDF-based model for Global South actors, as the ability to target document sections makes identifying named entities such as countries or regions more reliable. For example, biographies could be omitted or included as needed, as biographies will likely have a contributor's country location included.
3. With respect to open citations, the new model allows structured coding of bibliographic references (including bibliographic information and PIDs such as DOIs for research literature and, where applicable, primary sources). This affords a variety of improvements for further climate literature analysis if citation data are open, structured, and usable, such as significantly speedier literature reviews and synthesis reports, the location of data and protocols, and display and audit trails of results in search engines. Examples of open citation with bibliometric usage include OpenAlex and Scholia.
4. Entity tagging, the machine-readable labeling of entities (actors, organizations, places, times) and concepts in a research contribution (abstracts, technical terms), is also improved in the new model. This benefits Global South actors who can apply their own terminology services and see how these relate to different regions' terminology use. weADAPT from the Stockholm Environment Institute supports a service to connect climate glossary term use (Barrott et al. 2020).
5. Machine-readable representation of the core statements of a contribution is also improved because with linked open data parts, the identifiers are language independent and therefore support translation by design by stemming terms (in which the term is represented by a language-independent entity ID), which will support non-native English speakers, as it allows them to search English-language publications in their own language.
6. The new model also allows for machine-readable open licensing of all parts of a publication, such as text, images, media, code, data, software, metadata, standards, and bibliometric data.
7. Linked open data are further improved in the new model, allowing for machine-readable PIDs, graph and semantic data that are making use of the supporting systems of classification systems, controlled vocabularies, authority records, schemas, taxonomies, and ontologies. With the use of linked open data, research can utilize computational analysis—such as semi-automate literature reviews—and improve findability in general as well as enabling improvements in research reproducibility.

8. With respect to interoperable and computational formats, document, file, and data need a versioned source interoperability for machine readability. This is improved in the new model and allows for more accurate machine translation.
9. In relation to open peer review, the design of open, transparent review processes will result in the further improvement of knowledge production in the humanities.
10. Similarly open bibliometrics are necessary for transparency in research evaluation and ranking and for scientometrics and bibliometrics research.
11. Having clear contributor roles and attribution mechanisms are also important, and the new model allows for the encoding of roles and attribution for contributions to parts of a publication, such as for software, micro-publications, methods, protocols, data gathering, etc. Example schemas are CRediT for expanding scholarly contribution roles and Software Citation standards to give citations for software contributions.
12. Further improvements include clear indications of policy compliance with open access and open science agreements (e.g., Jisc Open policy finder, Platform for Responsible Editorial Policies [PREP], Open Access Journals Toolkit).

However, in scholarly publishing, publishers have been the implementation bottleneck where it concerns making sure that the research publications realize their *benefit through the use value of the research covered*. Even if researchers want to provide linked open data (LOD)¹² formats, interoperable formats, embedded LOD, or computational forms with literate programming, most academic publishers across all fields have not implemented any infrastructure to accommodate this (Stocker et al. 2024; Capadisli 2017; Adema 2023).

One thing that would be helpful in this context is a Linked Open Document Checklist (see Table 3), 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, such as in complex, data-driven, fast-moving knowledge domains.

The work we are doing with the ClimateKG has been produced as part of the open research community that organizes itself under the banner of #semanticClimate (Bhadra et al. 2024; Yadav et al. 2024). In this article, we are conceptualizing the ClimateKG, and in 2025 the software development phase will start. The #semanticClimate

12. Linked open data (LOD) are data that are openly accessible, usually via a URI and open licenses, and that make use of the possibilities to link to other data. LOD are best expressed via the Resource Description Framework (RDF). See <https://www.w3.org/TR/rdf10-xml/>.

Table 3. Linked Open Document Checklist (preprint-with-code)

No.	Item	Check
1.	Add to Wikidata to connect to the Linked Open Data Cloud, including metadata; bibliographic information; keywords; license; persistent identifiers such as DOIs; text structures; and labels for entities, representation of the core statements (Schöch 2021).	•
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	•

community is creating software, organizing events, and contributing to creating the ClimateKG, which, as explained earlier, will be a web service for searching and publishing to enable global dissemination of the knowledge contained in the UN IPCC reports. Future work will take place during 2025 with partners Wikibase4Research, Open Research Knowledge Graph (ORKG), and Lab Knowledge Infrastructures at TIB—Leibniz Information Centre for Science and Technology (TIB), alongside the #semanticClimate community.¹³

The Climate Knowledge Graph (Conceptual Modeling 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 itself has a design fault—it has no cataloging or indexing system. The web’s creator, Tim Berners-Lee, proposed a solution to this, which he coined the Semantic Web. But his intervention came too late, after the web had already developed in another direction. 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 of web search and social media (Meta, Google, Microsoft, Meta, and Twitter [now X]) exploited the lack of a central cataloging or indexing system for the web and created much of their wealth by making their own private knowledge graphs. An example of the Semantic Web is Google Knowledge Graph (Wikipedia 2024a), which has 500 billion facts (in 2020) and generates the info boxes on search results.

13. TIB, “Open Research Knowledge Graph,” accessed April 11, 2025, <https://orkg.org>.

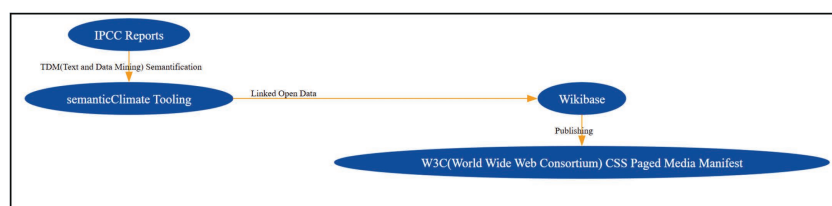


Figure 1. ClimateKG tech stack schematic (Graphviz source file) (Graphviz 2021)

Our work on the ClimateKG involves building an open knowledge graph for the open access parts of the IPCC Assessment Report AR6. The knowledge graph would be used for dissemination of the reports and would include various functionalities—from cataloging the literature corpora using ontologies, taxonomies, and controlled vocabularies to creating word search interfaces as used on academic repositories—and opportunities for content publishing and reuse with the ability to save the search results collated into new full text editions and output them as multi-format (e.g., web, ePUB, PDF).¹⁴

As outlined earlier, the IPCC reports are the gold standard of climate science, but they are rudimentarily published 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, and so forth. The ClimateKG goal would be to have the AR6 available online with the added functionality afforded by semantic publishing. The created knowledge graph then allows for queries that can be collated and exported as multi-format publications.

The #semanticClimate community has been developing software tooling since 2019 for searching and semantifying climate literature. In 2023 and 2024, numerous hackathons and further developments have resulted in several forms of semantification of the IPCC Sixth Assessment Report, including text and data mining conversion to HTML of all 70 chapters,¹⁵ 2,000 climate terms aligned with Wikidata,¹⁶ and the extraction of a glossary of 600 terms.¹⁷ In 2025, #semanticClimate will be working with TIB and the National Institute of Plant Genome Research (NIPGR) to build the ClimateKG as

14. The technology used to do so is Wikibase, Jupyter Notebook, #semanticClimate software, and TIB services: Wikibase4Research, Computational Publishing Service, Terminology Service (Antelope), ORKG, Renate (scholarly repository), and PID and metadata services.

15. “amilib/test/resources/ipcc/cleaned_content,” at “pmr_aug petermr/amilib,” GitHub, accessed April 11, 2025, https://github.com/petermr/amilib/tree/pmr_aug/test/resources/ipcc/cleaned_content.

16. “semanticClimate/ipcc/ar6/test/total_glossary/glossaries/total/acronyms_wiki.csv,” at “main · petermr/semanticClimate,” GitHub, accessed April 11, 2025, https://github.com/petermr/semanticClimate/blob/main/ipcc/ar6/test/total_glossary/glossaries/total/acronyms_wiki.csv.

17. “IPCC Glossary,” accessed April 11, 2025, <https://vivliostyle.vercel.app/#src=https://raw.githubusercontent.com/semanticClimate/glossary-demo/main/ipccglossary.jsonld>.

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 that there is support for the different public and user groups.

The ClimateKG will be a modern catalog and index of climate science literature with a subject-focus on UN Climate and UN IPCC corpora. It will be open source and based on open science principles. The ClimateKG technology allows for search, semantification (adding meaning), data analysis, and machine learning and AI functions of all types (scientifically validated, open science compliant, and ethically based in the SDGs, which is necessary for climate justice), as well as for republishing. Republishing here means supporting frictionless and authoritative reuse of content; as explained above, our first 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. An example is “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 a 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”¹⁸ (e.g., the IPCC). The IPCC has expertise in the following, so has the *property* of “knowsAbout,”¹⁹ “climatology,”²⁰ and “climate justice,”²¹ which reflects its relationships.

With only these few bits of data (the organization IPCC and its expertise), we can tap into Wikidata linked open data on “climatology” and retrieve 97 language translations and 50 identifiers and on “climate justice” 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), and so on.

Methodology underlying the semi-automated literature search

With our research, we want to explore and demonstrate, with the open-source software tools used by #semanticClimate, the use value of “knowledge graph” technology to

18. “Organization - Schema.Org Type,” accessed April 11, 2025, <https://schema.org/Organization>.

19. “knowsAbout - Schema.Org Property,” accessed April 11, 2025, <https://schema.org/knowsAbout>.

20. “Climatology,” accessed April 11, 2025, <https://www.wikidata.org/wiki/Q52139>.

21. “Climate Justice,” accessed April 11, 2025, <https://www.wikidata.org/wiki/Q1291678>.

establish the reach and impact of a given topic in a scholarly literature corpus. The ability to have data on where and when a term or concept occurs, either in a single paper, over corpora, or in a temporal aspect, allows data analysis to reveal patterns. A temporal word frequency N-Gram is a good example of this.

As explained earlier, we used “climate justice” as a term to search scholarly literature corpus repositories. The focus was on open access research articles that are slanted toward 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 to provide a representative view of contemporary, up-to-the-minute research.

Our main research question was how to efficiently and usefully locate discourse on climate justice in the open access scholarly literature corpus and the AR6 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 temporally—for example, three-year periods—and geographically? Word frequency is determined using the natural language processing software library called spaCy (2024). Using word frequency helps reveal patterns and gain insights into the literature.
- What associated terms do authors use for climate justice? The terminology is important as it encapsulates the discourse around 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 examines word frequency analysis, generating vector embeds, machine learning, NLP, and other aspects.

- How are “climate justice” and its associated terms used in UN Climate literature—both in AR6 (IPCC 2023d) and the IPCC Glossary (IPCC 2011)? The searches are semi-automated, which means researchers review the data on term use, which can either aid in finding relevant research or help in understanding or interpreting patterns.

- How do “climate justice” and its associated terms appear in Wikipedia, Wikidata, and Wiktionary? These resources were selected because they are important open public resources for climate information, but they are also important for knowledge graph creation, annotation, indexing, cataloging, and multilingualism, among others.

Methods

Our work on the ClimateKG until now has been based on various methods. First, it has been organized as an open research community. The #semanticClimate community, who as mentioned make open-source software, is the larger context for the research we discuss here. Community refers to “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 conducting hackathons, prototyping and demoing, running intern programs, learning programs, and experts and partners organization contributions. Second, we have employed open-source coding when creating software. The code made by the #semanticClimate community is literally the “why” and “what” of the project.²² Open-source methodologies are inherently social and facilitate community building, including 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 both novices and experts. In addition, we have made use of open channels for communication that support synchronous and asynchronous working (e.g., our weekday meetings are held 13:30 IST on Slack). We have also shared all our research in development employing open-notebook science, an open science practice that means making all research activities openly available as it is recorded to ensure there is no insider knowledge (Wikipedia 2023). We have also followed UNESCO Recommendation on Open Science values and principles and FAIR principles (UNESCO 2021). A further significant component is that our research outputs were versioned using storage on Git and the deposits minted DOIs on the CERN-hosted online service Zenodo and other PIDs systems with a guidance implementation checklist for researchers, including review.²³ Finally, we have employed the Linked Open Document Checklist (see Table 3) to follow guidelines and submit a preprint-with-code and have used

22. See our GitHub code repositories and Zenodo #semanticClimate community deposits.

23. Persistent identifiers (PIDs) are 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 used for publications and sections or chapters therein and for data and digital objects, such as 3D, images, video, and other media (DOIs), and for other entities, including persons (ORCID), organizations (Research Organization Registry), funders (Open Funder Registry), projects (RAID), and events (ConflDent).

software citation in GitHub using Citation File Format (CFF) (Druskat 2023) with Zenodo.

Work program

The work program we devised outlines the research carried out, the steps taken, and results as data and documentation of findings. The program 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 informatic searches can perform on such textual corpora. As mentioned before, we are not presenting final results here; this article is not a representative analysis of the state of the art of the frequency of research on climate justice. Instead, this article reflects on an indicative scoping exercise and conceptual modeling to show what the tooling is capable of, the steps involved, its scale, and terms. In other words, the #semanticClimate project is currently a “development project” and not yet a “production project.” The difference between these statuses in software development is that the development aspects of the project have not been optimized and made usable by the community. For example, technologies have not been optimized yet to lower their carbon footprint when being used at scale.

An open-notebook science log has been made for the complete work program on GitHub Discussions, GitHub Project (Issues), and the Jupyter Notebook “#semanticClimate Tools” (Bhadra et al. 2024). All data was stored on GitHub and made available according to FAIR principles. The work program covered the following steps to address the four sub-questions mentioned earlier. First, we created a Google Colab Jupyter Notebook to search Europe PMC for the term “climate justice” and retrieve literature in three-year periods from 2004 to 2024. We retained the literature and data and reported on the findings from the literature search already completed. Second, we carried out data analysis on this mini-corpus (the papers downloaded from Europe PMC) to find associated terms. Third, we searched semantified versions of IPCC reports held by the #semanticClimate team on GitHub (AR6 and the IPCC Glossary) and again reported on findings below. Fourth, 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 that 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 of what we have done as well as associated software and supporting data can be found in Appendix II. The

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

Initial findings

What were the outcomes of the scoping exercise we carried out of a semi-automated literature review on climate justice (see Table 4), and why is a new model so important to publish and 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 should instead be seen as a modeling exercise and indicative of the capabilities of the process.

As explained, the semi-automated literature review can be used on any topic and in different domains (see Appendix II for how the results were gained). For the search we conducted for the term “climate justice,” the data showed that in the open access scholarly literature corpus from Europe PMC, there was no publication on climate justice from 2004 to 2009. From 2010 onward, the publication of climate justice articles increased exponentially. In the most recent three-year period (2022–24), a significant increase occurred with 574 papers mentioning the term “climate justice” (see Appendix II, Table A2).

Carrying out data analysis using word frequency analysis, the researchers further looked for associated terms for climate justice on Europe PMC. DataTables are used to display a list of retrieved articles that use the term “climate justice” in addition to

Table 4. Data parts of a publication and their usage in a semi-automated literature review on climate justice (in a corpus where the PDF has been removed from the workflow)

Data part of a publication	Semi-automated literature
Syntactic structure	Document sections were pinpointed to enable searching for countries.
Open citations	Europe PMC's 6.5 million open access papers could be searched.
Entity tagging	“Climate justice” and associated terms were located using word frequency analysis.
Core statements	A set of building blocks were put in place to later carry out this process. Climate justice-related terms from Europe PMC and IPCC were retrieved from Wikipedia with definitions that would be used to create semantic representation of core statements.
Linked open data	Wikimedia: Wikidata, Wikipedia, and Wiktionary terms identified and marked up in IPCC and Europe PMC papers.
Interoperable and computational formats	IPCC reports were converted to HTML with IDs, enabling the creation of a DataTable that a researcher could review and use for their research or make improvements to the process. Contents, process, and outputs were mainly carried out with Jupyter Notebook.

associated terms (see Appendix II, Figure A3).²⁴ This allows the researcher to have an overview of journal names, authors, and other parts of a paper (e.g., discussion, methods, or even tables or figures) that have used the term “climate justice,” which helps the researcher evaluate the results in a convenient way as they can scan read across the retrieved data, similar to how a chart or data visualization can help summarize information or data. Additionally, the DataTables include a list of countries mentioned in the articles for the years 2010 to 2024 (see Appendix II, Figure A4).

To create the Climate Justice Dictionary, a word list of 25 words was generated from the Europe PMC search of climate justice–related terms in 100 Europe PMC articles from 2010 to 2024.²⁵ This word list of associated terms helps improve our understanding and overview of the wider discourse around climate justice. Then, searching the IPCC AR6 semantified version held by #semanticClimate on GitHub and the IPCC Glossary GitHub version identified the chapters that include the term “climate justice” (see Appendix II, Figure A5). From this a word list was generated of 51 climate justice–related terms in chapters 08 and 18 of AR6 WG2 (see “*Word list IPCC report – 51 terms*”: *Climate justice*).²⁶

The Climate Justice Dictionary was made by combining the two word lists to create a list of 76 terms. These terms were used to search Wikipedia, Wikidata, and Wiktionary for all associated terms and then to store the annotations as linked open data. The annotations provide additional descriptive information such as dictionary definitions, illustrative images, or encyclopedia articles (see Appendix II, Figures A7 and A8). This resulted in the creation of The Climate Justice Dictionary (word list) and Climate Justice Dictionary Annotated (word list, Wikipedia enhanced). The annotated 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 much more explicit. Our intention is to show the value of semantic electronic publishing by presenting use case prototypes highlighting how this is actually useful to wider audiences and how, if recorded in a database (here a knowledge graph), the data and knowledge can be given a computable surface of sorts. The frequency of the term “climate justice” in the six million open access articles from Europe PMC revealed patterns such as zero mentions of “climate justice” in 2004 to mentions in 574 papers in 2022 to 2024. Revealing the associated

24. The list of 100 articles from 2010 to 2024 is available at https://github.com/semanticClimate/JEP-article/tree/main/literature_search2010-2024.

25. See “words_from_lit_corpus.txt,” available at https://github.com/semanticClimate/JEP-article/blob/main/data/words_from_lit_corpus.txt. See Appendix III (#1) for a step-by-step reproducible process of the word list generation.

26. “JEP-article/data/words_IPCC_chao08_chap18.txt,” in “semanticClimate/JEP-article,” GitHub, accessed April 11, 2025, https://github.com/semanticClimate/JEP-article/blob/main/data/words_IPCC_chao08_chap18.txt.

terms for climate justice in Europe PMC enabled the ability to search further for journal names and authors, which contributes to locating the discourse around the topic.

In a narrower corpus such as the IPCC AR6 report, specific chapters relevant to climate justice could be more easily identified. Our research question focused on locating the climate justice discourse in both scholarly literature and the IPCC reports, and the Climate Justice Dictionary (see Appendix II) concludes this work by being a proof of concept for what can be retrieved about this discourse from the Wikiverse, which is made up of multilingual resources such as Wikidata, Wikipedia, and Wiktionary. This included images and plain language descriptions that have been included in the annotated version of the dictionary. The result is that each term is annotated with definitions, descriptions, and illustrative images. The Climate Justice Dictionary is still in development and was made by running a number of software programs on the command line and on Google Colab Jupyter Notebooks, with researchers evaluating results. However, it is predominantly intended as a proof of concept and as an example of the types of results a knowledge graph could generate on the fly similar to the way that web search engines function.

Future work

Building the Climate Knowledge Graph (ClimateKG) has the potential to create greater accessibility, for a wider audience, of the IPCC Assessment Reports. The semi-automated literature review and 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. When adding the AR6 corpus and other corpora to the ClimateKG (with Wikibase being used as the LOD data storage for the knowledge graph project), the following will be carried out:

- Adding paragraph with IDs of the AR6 corpus
- Adding IDs of all other significant objects in the AR6 corpus literature
- Occurrence of a term 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’ Wikibase4Research service, as well as from other TIB projects ORKG and Lab Knowledge Infrastructures—and with NIPGR will be building these workflows into the ClimateKG.

Author Biographies

Simon Worthington is a publishing technology researcher at the Open Science Lab, TIB—Leibniz Information Centre for Science and Technology, Hanover, Germany. He works creating software for open source publishing pipelines, for multi-format and computational publishing. Current research is on the NFDI4Culture (German National Research Infrastructure Consortium for Culture) Task Area 4—Data Publication. He is on the board of directors of FORCE11, a LIBER Citizen Science Working Group member, editor-in-chief of the Citizen Science Guide for Research Libraries, and a volunteer with #semanticClimate.

Gitanjali Yadav is a scientist at the National Institute of Plant Genome Research (NIPGR) New Delhi, India, and a data science professor at IISER Bhopal. She serves as a trustee at St. Edmund’s College, Cambridge. Her research specializes in genomics and structural informatics for applications in climate, conservation, and food security.

Peter Murray-Rust is Professor of Chemistry at the University of Cambridge. He is a pioneer in text mining through ContentMine, an advocacy and implementation project consisting of a community, framework, and toolset that facilitates the freeing of scientific data and building of further tools for improved open science, technology, and medicine communication. His is co-author of Chemical Markup Language and co-founder of #semanticClimate.

Renu Kumari is a Project Scientist III at the National Institute of Plant Genome Research (NIPGR) New Delhi, India. She has a PhD in Biotechnology. Her research area includes abiotic stress, phytochemistry, climate science, and epigenetics. She is also associated with #semanticClimate as a program manager.

Shweeta Hegde is an open source developer and a science communicator. She was the project manager of #semanticClimate and is currently a volunteer. She has co-authored a Python package called docanalysis, a text data mining tool.

Parijat Bhadra is a third-year undergraduate student at Garden City University, Bengaluru. Her core subjects include biotechnology, biochemistry, and genetics. Currently, she is interning at the National Institute of Plant Genome Research (NIPGR) and #semanticClimate.

References

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

- Adema, Janneke. 2023. “Developing a Publishing Workflow for Computational Books.” Copim, April 28, 2023. <https://doi.org/10.21428/785a6451.30c8c105>.
- Arnold, Matthias, Alexandra Büttner, Jörg Heseler, and Simon Worthington. 2023. “Digital Publications in Culture: Examples and Key Features—Survey Results from the NFDI4Culture Community.” Zenodo, March 29, 2023. <https://doi.org/10.5281/ZENODO.7107214>.
- Barrott, Julia, Denise Recheis, Sukaina Bharwani, Christina Daszkiewicz, and Ruth Butterfield. 2020. “The PLACARD Taxonomies for CCA & DRR: Development Note and Future Work.” weADAPT, July 13, 2020. <https://weadapt.org/knowledge-base/climate-knowledge-brokers/taxonomies-developed-for-cca-drr/>.
- 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, November 17, 2024. <https://doi.org/10.5281/ZENODO.14176447>.
- Bush, Vannevar. 1945. “As We May Think.” *The Atlantic*, July 1, 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, et al. 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*. Edited by H. Lee and J. Romero. Intergovernmental Panel on Climate Change (IPCC). <https://doi.org/10.59327/IPCC/AR6-9789291691647>.
- Capadisli, Sarven. 2017. “Calls for Linked Research.” <https://linkedresearch.org/calls>.
- Capadisli, Sarven. 2019. “Linked Research on the Decentralised Web.” <https://csarven.ca/linked-research-decentralised-web>.
- Chen, George, Alejandro Posada, and Leslie Chan. 2019. “Vertical Integration in Academic Publishing: Implications for Knowledge Inequality.” In *Connecting the Knowledge Commons—from Projects to Sustainable Infrastructure*, edited by Leslie Chan and Pierre Mounier. OpenEdition Press. <https://doi.org/10.4000/books.oep.9068>.
- Chowdhary, Reema. 2024. “Shifting Paradigms of Multilingual Publishing and Scholarship in India.” *Journal of Electronic Publishing* 27 (1). <https://doi.org/10.3998/jep.5592>.
- Demeter, Márton, and Ronina Istrate. 2020. “Scrutinising What Open Access Journals Mean for Global Inequalities.” *Publishing Research Quarterly* 36 (4): 505–22. <https://doi.org/10.1007/s12109-020-09771-9>.
- Druskat, Stephan. 2023. “What Is a CITATION.cff File?” Citation File Format (CFF). 2023. <https://citation-file-format.github.io/>.

- Europe PubMed Central (PMC). 2024. "Open Access Subset." Accessed November 8, 2024. <https://europepmc.org/downloads/openaccess>.
- GO FAIR. n.d. "FAIR Principles." Accessed December 5, 2024. <https://www.go-fair.org/fair-principles/>.
- Graphviz. 2021. "Graphviz." <https://graphviz.org>.
- GreyNet International. 2024. "Grey Literature Network Service." 2024. <https://greynet.org/home/aboutgreynet.html>.
- Hughes, Hannah. 2024. *The IPCC and the Politics of Writing Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009341554>.
- ICJ (International Court of Justice). 2024a. "Obligations of States in Respect of Climate Change." Press release, November 26, 2024. <https://www.icj-cij.org/sites/default/files/case-related/187/187-20241126-pre-01-00-en.pdf>.
- ICJ (International Court of Justice). 2024b. "Obligations of States in Respect of Climate Change—Verbatim Record 2024/38." December 3, 2024. <https://www.icj-cij.org/case/187>.
- IPCC. 1988. *Report of the First Session of the WMO/UNEP Intergovernmental Panel on Climate Change (IPCC)*. World Climate Programme Publications Series. <https://www.ipcc.ch/meeting-doc/1st-session-of-the-ipcc-geneva-9-11-november-1988/>.
- IPCC. 1992. *Climate Change: The IPCC 1990 and 1992 Assessments*. <https://www.ipcc.ch/report/climate-change-the-ipcc-1990-and-1992-assessments/>.
- IPCC. 2011. "IPCC Glossary." <https://apps.ipcc.ch/glossary/>.
- IPCC. 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>.
- IPCC. 2018. "Global Warming of 1.5°C: Headline Statements from the Summary for Policymakers." https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15_Hheadline-statements.pdf.
- IPCC. 2019. "Available Data for AR6." <https://ipcc-data.org/ar6landing.html>.
- IPCC. 2022a. *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*. Cambridge University Press. <https://doi.org/10.1017/9781009157988>.
- IPCC. 2022b. *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*. Cambridge University Press. <https://doi.org/10.1017/9781009157940>.
- IPCC. 2022c. *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009157964>.
- IPCC. 2022d. "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 University Press. <https://doi.org/10.1017/9781009157940.001>.
- IPCC. 2023a. *Climate Change 2021—the Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009157896>.
- IPCC. 2023b. *Climate Change 2022—Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009325844>.
- IPCC. 2023c. *Climate Change 2022—Mitigation of Climate Change: Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009157926>.
- IPCC. 2023d. "Sixth Assessment Report." <https://www.ipcc.ch/assessment-report/ar6/>.

- IPCC. n.d.-a. "About the IPCC." Accessed December 4, 2024. <https://www.ipcc.ch/about/>.
- IPCC. n.d.-b. "IPCC Glossary Search." Accessed November 30, 2024. <https://apps.ipcc.ch/glossary/>.
- IPCC-WG1. n.d. GitHub repository. Accessed December 5, 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. <https://doi.org/10.1111/asap.12203>.
- Kay, Alan C. 1972. "A Personal Computer for Children of All Ages." In *ACM '72: Proceedings of the ACM Annual Conference—Volume 1*. Association for Computing Machinery. <https://dl.acm.org/doi/10.1145/800193.1971922>.
- Knuth, Donald E. 1992. "Literate Programming." CSLI Lecture Notes 27. Center for the Study of Language and Information. <https://www-cs-faculty.stanford.edu/~knuth/lp.html>.
- Krajewski, Markus. 2011. *Paper Machines: About Cards & Catalogs, 1548–1929*. Translated by Peter Krapp. MIT Press.
- Latour, Bruno. 2018. *Down to Earth: Politics in the New Climatic Regime*. Translated by Catherine Porter. Polity Press.
- Licklider, J. C. R. 1965. *Libraries of the Future*. MIT Press. <http://archive.org/details/librariesoffutur00lickuoft>.
- Milfont, Taciano L., Paul G. Bain, Yoshihisa Kashima, 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*. Published by the author.
- Nganji, Julius T. 2015. "The Portable Document Format (PDF) Accessibility Practice of Four Journal Publishers." *Library & Information Science Research* 37 (3): 254–62. <https://doi.org/10.1016/j.lisr.2015.02.002>.
- Papers with Code. n.d. Accessed November 16, 2024. <https://paperswithcode.com>.
- Pearce, D. W., W. R. Cline, A. N. Achanta, et al. 1996. "The Social Costs of Climate Change: Greenhouse Damage and the Benefits of Control." In *Climate Change 1995—Economic and Social Dimensions of Climate Change: Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change*. https://www.ipcc.ch/site/assets/uploads/2018/03/ipcc_sar_wg_III_full_report.pdf.
- Raju, Reggie, and Auliya Badrudeen. 2022. "Social Justice Driving Open Access Publishing: An African Perspective." *Journal of Electronic Publishing* 25 (1). <https://doi.org/10.3998/jep.1910>.
- Riley, Susan. n.d. "What Is STEAM Education?" The Institute for Arts Integration and STEAM. Accessed April 11, 2025. <https://artsintegration.com/what-is-steam-education-in-k-12-schools/>.
- Schöch, Christof. 2021. "Open Access Für Die Maschinen." In *Die Zukunft des kunsthistorischen Publizierens*, edited by Maria Effinger and Hurbertus Kohle. arthistoricum.net. <https://doi.org/10.11588/ARTHISTORICUM.663.C9210>.
- #semanticClimate. 2019a. "Records." Zenodo. <https://zenodo.org/communities/semanticclimate/records?q=&l=list&p=1&s=10&sort=newest>.
- #semanticClimate. 2019b. "Repositories." GitHub. <https://github.com/orgs/semanticClimate/repositories?>
- #semanticClimate. 2023. "IPCC Reports and City Climate Change Plans: Proof of Concept Prototype—Open Climate Reader." FSCI Hackathon Team, August 4, 2023. <https://semanticclimate.github.io/city-open-climate-reader/>.
- #semanticClimate. n.d. "ipcc/cleaned_content." GitHub. Accessed November 9, 2024. https://github.com/semanticClimate/ipcc/tree/main/cleaned_content.
- Semantic Web Standards. n.d. "RDF." Accessed December 20, 2024. <https://www.w3.org/RDF/>.

- Sobral, Sonia Rolland. 2020. "Mobile Learning in Higher Education: A Bibliometric Review." *International Journal of Interactive Mobile Technologies (ijIM)* 14 (11): 153–70. <https://doi.org/10.3991/ijim.v14i11.13973>.
- spaCy. 2024. "Industrial-Strength Natural Language Processing in Python." Explosion, 2024. <https://spacy.io/>.
- Stocker, Markus, Lauren Snyder, Matthew Anuso, et al. 2024. "Rethinking the Production and Publication of Machine-Reusable Expressions of Research Findings." Preprint, arXiv, May 21, 2024. <https://doi.org/10.48550/ARXIV.2405.13129>.
- Stockhause, Martina, Martin Juckes, Robert Chen, et al. 2019. "Data Distribution Centre Support for the IPCC Sixth Assessment." *Data Science Journal* 18 (1): 20. <https://doi.org/10.5334/dsj-2019-020>.
- Sundell, Taavi. 2021. "Political Economy of Plan S: A Post-Foundational Perspective on Open Access." *Political Research Exchange* 3 (1): 1934049. <https://www.tandfonline.com/doi/abs/10.1080/2474736X.2021.1934049>.
- UNDP (United Nations Development Programme). 2023. "Climate Change Is a Matter of Justice—Here's Why." UNDP Climate Promise, June 30, 2023. <https://climatepromise.undp.org/news-and-stories/climate-change-matter-justice-heres-why>.
- UNESCO. 2021. *UNESCO Recommendation on Open Science*. UNESDOC Digital Library. <https://doi.org/10.54677/MNMH8546>.
- UNFCCC. 1997. "Kyoto Protocol to the United Nations Framework Convention on Climate Change." December 10, 1997. <https://unfccc.int/documents/2409>.
- UNFCCC. 2016. "The Paris Agreement." <https://unfccc.int/process-and-meetings/the-paris-agreement>.
- UNFCCC. 2022. "Decision 1/CMA.3. Glasgow Climate Pact." March 8, 2022. https://unfccc.int/sites/default/files/resource/cma2021_10a01E.pdf#page=1.08.
- UNFCCC. n.d.-a. "Conference of the Parties (COP)." Accessed December 1, 2024. <https://unfccc.int/process/bodies/supreme-bodies/conference-of-the-parties-cop>.
- UNFCCC. n.d.-b. "What Is the Kyoto Protocol?" Accessed December 1, 2024. https://unfccc.int/kyoto_protocol.
- United Nations. 2015. "The Sustainable Development Goals." In *Transforming Our World: The 2030 Agenda for Sustainable Development*. <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf?ref>.
- United Nations. 2023. "Secretary-General Calls on States to Tackle Climate Change 'Time Bomb' Through New Solidarity Pact, Acceleration Agenda, at Launch of Intergovernmental Panel Report." Press release, March 20, 2023. <https://press.un.org/en/2023/sgsm21730.doc.htm>.
- Wei, Yi-Ming, Rong Han, Qiao-Mei Liang, 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." Updated December 3, 2023. https://en.wikipedia.org/w/index.php?title=Open-notebook_science&oldid=1188104495.
- Wikipedia. 2024a. "Google Knowledge Graph." Updated October 8, 2024. https://en.wikipedia.org/w/index.php?title=Google_Knowledge_Graph&oldid=1250126393.
- Wikipedia. 2024b. "Graph Database." Updated November 20, 2024. https://en.wikipedia.org/w/index.php?title=Graph_database&oldid=1258574566.
- Wikipedia. 2024c. "OpenStreetMap." Updated December 3, 2024. <https://en.wikipedia.org/w/index.php?title=OpenStreetMap&oldid=1261023652>.
- Wikipedia. 2024d. "Tim Berners-Lee." Updated October 28, 2024. https://en.wikipedia.org/w/index.php?title=Tim_Berners-Lee&oldid=1253967594.

- Wikipedia. 2024e. "US Withdrawal from Paris Agreement." Updated November 14, 2024. https://en.wikipedia.org/w/index.php?title=United_States_withdrawal_from_the_Paris_Agreement&oldid=1257286872.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (2016): Article 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Willighagen, Egon, and Lars Willighagen. 2023. "IPCC Wikibase Queries." #semanticClimate, IPCC-Queries. <https://semanticclimate.github.io/IPCC-Queries/>.
- Worthington, Simon. 2025. "Climate Knowledge Graph." April 2025. <https://github.com/TIBHannover/climate-knowledge-graph>.
- Worthington, Simon, Kolja Bailly, and Anna Rahr. 2024. "TIBHannover/semantic glossar." Jupyter Notebook. <https://github.com/TIBHannover/semantic-glossar>.
- Wright, Alex. 2014. *Cataloging the World: Paul Otlet and the Birth of the Information Age*. Oxford University Press.
- Yadav, Gitanjali, Shweeta 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* 71:480–95. <https://doi.org/10.56042/alis.v71i4.14285>.
- Zhang, Lin, Yahui Wei, Ying Huang, and Gunnar Sivertsen. 2022. "Should Open Access Lead to Closed Research? The Trends Towards Paying to Perform Research." *Scientometrics* 127 (12): 7653–79. <https://doi.org/10.1007/s11192-022-04407-5>.

Appendix I: Resources used

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

Data processing software tools

Below is a table of the open-source software for data processing used by #semanticClimate for the literature analysis (Table A1).

Software

A full list of all #semanticClimate software can be found on GitHub (#semanticClimate 2019b) and on the #semanticClimate Zenodo community (#semanticClimate 2019a).

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 the COVID-19 pandemic: country, disease, drug, np, organization, test_trace, virus, Zoonosis
- <https://github.com/petermr/dictionary/tree/main/openVirus20210120>

Table A1. Software for data processing used by #semanticClimate

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/1CF68Fw9NymUH2ZAEepUeligXczhn4A/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, color 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/ Creating dictionaries: https://github.com/petermr/tigr2ess/blob/master/dictionaries/TUTORIAL.md https://github.com/petermr/amilib/blob/main/test/test_headless.py
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	
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, initialization, and storage of a literature search and analysis project.

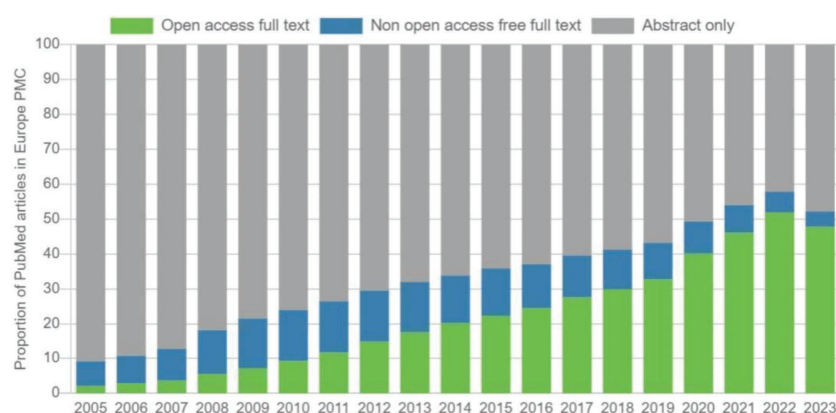


Figure A1. Europe PMC percentage open access

Literature corpora and semantified copies

Two corpora are used in the research: Europe PMC and the IPCC Sixth Assessment Report (AR6). First, 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 (Europe PMC 2024; see Figure A1). Additionally, Europe PMC has encouraged preprints since the start of the COVID-19 pandemic, which enables early access to research papers and to cutting-edge science. Europe PMC provides articles as full text as well supports a wide range of computational machine access. The #semantic-Climates community is able to carry out automated search and document retrieval using its software “pygetpapers.” Searches can target specific paper parts, such as abstracts, findings, or conclusions, and retrieval of papers can be as full text JATS format.

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 time-frame. In short, our world needs climate action on all fronts—everything, everywhere, all at once” (United Nations 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*. 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 licensed) (Calvin et al. 2023)
 - IPCC 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 licensed) (IPCC 2011)
- **Working Group reports** (these feed into the Synthesis Report)
 - *Climate Change 2021—the Physical Science Basis*. Working Group I Contribution to the IPCC Sixth Assessment Report. DOI: 10.1017/9781009157896 (CC-BY-NC-ND 4.0) (IPCC 2023a)
 - *Climate Change 2022—Impacts, Adaptation and Vulnerability*. Working Group II Contribution to the IPCC Sixth Assessment Report. DOI: 10.1017/9781009325844 (CC-BY-NC-ND 4.0) (IPCC 2023b)
 - *Climate Change 2022—Mitigation of Climate Change*. Working Group III Contribution to the IPCC Sixth Assessment Report. DOI: 10.1017/9781009157926 (CC-BY-NC-ND 4.0) (IPCC 2023c)
- **Special Reports** (these feed into the Synthesis Report)
 - *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. DOI: 10.1017/9781009157940 (CC-BY-NC-ND 4.0) (IPCC 2022b)

- *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. DOI: 10.1017/9781009157988 (CC-BY-NC-ND 4.0) (IPCC 2022a)
- *The Ocean and Cryosphere in a Changing Climate*. Special Report of the Intergovernmental Panel on Climate Change. DOI: 10.1017/9781009157964 (CC-BY-NC-ND 4.0) (IPCC 2022c)

Unlike the research papers in Europe PMC, the AR6 is not published in compliance with modern open science and FAIR principles. 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 it 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 are being published following modern open science methods, using Git versioning and FAIR principles. The #semanticClimate community has made semantified copies of the AR6 *Climate Change 2023: Synthesis Report* and IPCC Glossary using its software tooling, including 70 chapters of AR6 as annotated HTML with IDs per paragraph (work in progress) (#semanticClimate, n.d.) and the IPCC Glossary 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 has been used for annotating literature. Annotation in this context refers to adding additional information, for example, from Wikipedia and Wiktionary, or as “semantic labeling” with linked open data from Wikidata. Using these annotations the following features were enabled: definitions, multilingual definitions, image illustrations, and linked open data information, among others. The #semanticClimate tooling can access these resources via API access, scripting its tools in Jupyter Notebook, and so forth.

- Wikipedia: <https://www.wikipedia.org>
- Wiktionary: <https://www.wiktionary.org>
- Wikidata: <https://www.wikidata.org>.

Appendix II: Work program in detail

A literature review/search for the term “climate justice” using the 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 “Climate Justice Demo `#semanticClimate` Tools” running on Google Colab.

Complete work program on GitHub Discussion, GitHub Project (Issues), and the Jupyter Notebook “`#semanticClimate` Tools” (Bhadra et al. 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 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 released every day and 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 publications 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:

- A. Searching literature using a query term with `pygetpapers`
- B. Creating tables using `DataTables` (<https://datatables.net>) for the retrieved articles with `amilib`
- C. 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

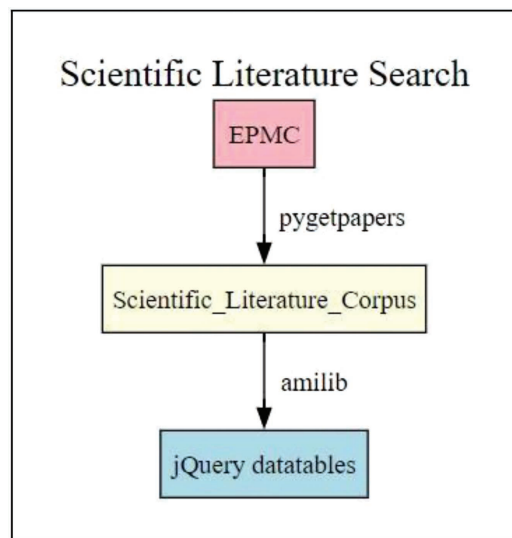


Figure A2. Steps to search literature from Europe PMC (Graphviz source file)

scientific text repositories, analyzes the hits, and systematically downloads the articles without further interaction.

Figure A2 illustrates the steps to use for `pygetpapers`.

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 a start date and end date.
- The use of `-n` will give the number of articles without any download of data.

```
pygetpapers --query' "climate justice" '--xml -n --start-date "2010-01-01" --enddate "2024-10-31" --output fin_climate_justice --save_query
```

This code will search 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 summarized in the form of tables using DataTables software. The important metadata of the corpus 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.

```
amilibHTML--operationDataTables--indirfin_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 such as 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

pygetpapers has been used to search literature on climate justice per three-year period over a 20-year span (Table A2). The results are not all “publications on climate justice” but papers retrieved in a conventional search for the term “climate justice.”

Table A2. Comparison of publications on climate justice from 2004 to 2024, by three-year span

Start date	End date	No. of papers 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

The table shows that there was no publication mentioning climate justice published from 2004 to 2009. From 2010 onward, the publication of articles mentioning the term “climate justice” has increased exponentially.

B. jQuery DataTables: Summary of the retrieved articles

A data table has been created for the articles published from 2010 to 2024 that mention climate justice (Figure A3). It presents a summary of all the retrieved articles, gives links to PMCID and DOIs, and also provides a brief abstract. DataTable supports the “human-in-the-loop” semi-automated part of the literature search in which the researcher can get a condensed view of search results and, for example, see all the journal names.

10 entries per page

Search:

query: "climate justice"; start: 2010-01-01; end: 2024-10-31

pmcid	doi	title	authorString	journalInfo:journal.title	pubYear	abstractText
PMCID10011356	10.1371/journal.pgrh.0001356	Achieving climate justice, safeguarding planetary health: Diagnosis and demands from next generation leaders for COP27 and beyond.	• Guinto RR • Devanayagam TA • Chai PT • Singh A • Taenua BKE	PLOS global public health	2022	
PMCID10942636	10.1515/mmpg-2021-0071	Climate Justice and the Duty of Restitution.	• Truccone-Borgogno S.	Moral philosophy and politics	2023	Much of the climate justice discussion revolves around how the remaining carbon budget should be globally allocated. Some authors defend the unjust enrichment interpretation of the beneficiary pays pr
PMCID10063200	10.1016/s0140-6736(23)00276-3	No climate change justice in lies of global authorship equity.	• Raz MMA • Wangari MC • Mugambi JK	Lancet (London, England)	2023	
PMCID10122113	10.1016/j.jama.2023.100494	Climate change: we must act now to secure a sustainable, healthy future for all.	• The Lancet Regional Health- Americas	Lancet regional health- Americas	2023	

Figure A3. Data table for retrieved articles (list of 100 sample articles from 2010–24; URL link to the data table)

The names of countries mentioned in the papers were extracted with `docanalysis` and represented in a word cloud (Figure A4), in which the size of the word depends on how frequently the word is mentioned in the corpus of 100 articles published from 2010 to 2024. At present in the scoping exercise, only cleaned-up data have been retrieved and presented; false positives have been removed, to a degree. The next step would be for a researcher to theorize why certain countries are mentioned more frequently than others in relation to the topic of climate justice.

Figure A4. The list of countries mentioned in the corpus of climate justice articles for the years 2010 to 2024

The semantic UN IPCC corpus has been searched for the term “climate justice” with the use of the `#semanticClimate` tool `pyamihtmlx`, which enabled us to compile a list of the chapters that have mentioned climate justice (Figure A5).

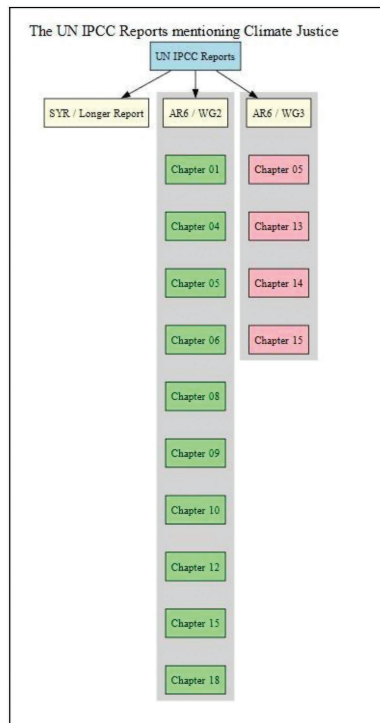


Figure A5. The list of IPCC reports and chapters that include the term “climate justice” (Graphviz source file)

emissions. The success of adaptation strategies therefore depends on equitable development and climate justice. See the section “The IPCC Reports, Digital Publishing, and Climate Justice” in the article for more details.

Associated terms and dictionary creation

Associated climate justice terms in Europe PMC and IPCC

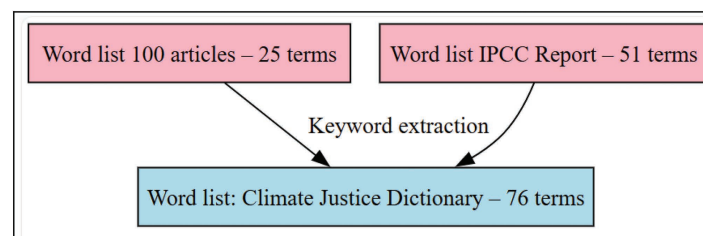


Figure A6. Creation of the climate justice word list (Graphviz source file)

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

Note: The word lists have been extracted from IPCC AR6/WG2/Chapter 08, IPCC AR6/WG2/Chapter 18, and the scholarly literature corpus of 100 articles that mention climate justice.

- **“Word list 100 articles – 25 terms”:** Climate justice–related terms in Europe PMC 100 articles from 2010 to 2024—list of terms from scholarly literature corpus (Europe PMC):
words_from_lit_corpus.txt
- **“Word list IPCC report – 51 terms”:** Climate justice–related terms from the IPCC AR6/WG2 Report—Chapters 08 and 18:
words_IPCC_chao08_chap18.txt
Links to #semanticClimate HTML versions of AR6/WG2 Report—Chapters 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 (see Figure A6). The following is the list of terms used to make the Climate Justice Dictionary.

- **“Word list: Climate Justice Dictionary – 76 terms”:** Terms related to climate justice and including “climate justice” from literature corpus and the IPCC reports:
words_all_climatejustice.txt

Search IPCC Glossary using Climate Justice Dictionary

The list of words used to make the Climate Justice Dictionary has been compared with the IPCC Glossary terms, and we observed the occurrence of the terms “climate justice”

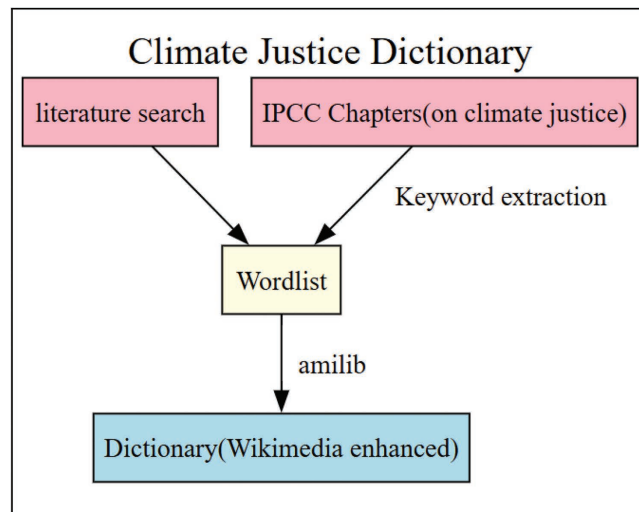


Figure A7. Flow diagram for creating a dictionary from word lists (Graphviz source file)

and “Indigenous Peoples” in the glossary. These were the only two terms to appear in the IPCC Glossary from our Climate Justice Dictionary. Such a result is important in a number of ways; for example, it can highlight the possible inclusion of other terms in a glossary but also shows how related terms can be used to indicate the meaning of a section of text.

The dictionary was then further enriched with the information available in Wikipedia, Wikidata, and Wiktionary (see Figure A7).

Wikipedia-enhanced dictionary

As climate science–related terms are not always easily understandable by laypersons, the information in the Climate Justice Dictionary has been added to every word/term from Wikipedia (see Figure A8). The Wikipedia-enhanced dictionary for all the terms was created with the tool `amilib` using the following code.

```
amilib  DICT  --words  words_all_climatejustice.txt
--description wikipedia --figures --dict climatejustice_
dictionary.html --operation create
```

search term: Climate justice Wikipedia Page

Climate justice is a type of environmental justice^[1] that focuses on the unequal impacts of climate change on marginalized or otherwise vulnerable populations.^[2] Climate justice seeks to achieve an equitable distribution of both the burdens of climate change and the efforts to mitigate climate change.^[3] The economic burden of climate change mitigation is estimated by some at around 1% to 2% of GDP^{[4][5]} Climate justice examines concerns such as equality, human rights, collective rights, justice and the historical responsibilities for climate change.^[6]



Fridays for Future demonstration in Berlin in September 2021 with the slogan “fight for climate justice”.

search term: social justice Wikipedia Page

This is an accepted version of this page



An artist's rendering of what Plato might have looked like. From Raphael's early 16th century painting “*Scuola di Atene*”.

search term: urban migration Wikipedia Page

Urbanization (or **urbanisation** in British English) is the population shift from rural to urban areas, the corresponding decrease in the proportion of people living in rural areas, and the ways in which societies adapt to this change. It can also mean population growth in urban areas instead of rural ones.^[1] It is predominantly the process by which towns and cities are formed and become larger as more people begin living and working in central areas.^[2]

Figure A8. Wikipedia-enhanced dictionary for climate justice–associated terms as web page from text dictionary: Words_all_climatejustice.txt

The Climate Justice Dictionary:

Climate Justice Dictionary (work in progress): Words_all_climatejustice.txt

Climate Justice Dictionary (Wikipedia enhanced): as web page

Appendix III: Supplementary material

1. Europe PMC semi-automated literature search Jupyter Notebook “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_search_2010-2024
4. Figures Graphviz source (all files)
<https://github.com/semanticClimate/JEP-article/tree/main/graphviz>
5. Word lists for creating the Climate Justice Dictionary
<https://github.com/semanticClimate/JEP-article/tree/main/data>
6. Dictionaries (drug, diseases, country, virus, etc.)
<https://github.com/petermr/dictionary/tree/main/openVirus20210120>