



## The #Semantic Climate Community: Making Open-Source Software for Knowledge Liberation

Gitanjali Yadav<sup>a</sup>, Shweata Hegde<sup>b</sup>, Renu Kumari<sup>c</sup>, Neeraj Kumari<sup>d</sup>, Peter Murray-Rust<sup>e</sup> and Simon Worthington<sup>f\*</sup>

<sup>a</sup>Scientist, NIPGR, New Delhi, India ORCID: 0000-0001-6591-9964

<sup>b</sup>Independent, developer at #semantic Climate. ORCID: 0000-0002-8258-274X

<sup>c</sup>Project Manager, NIPGR, India ORCID: 0000-0002-9451-7814

<sup>d</sup>PhD. Scholar, NIPGR, India. ORCID: 0009-0008-4758-2200

<sup>e</sup>Reader Emeritus, University of Cambridge, UK. ORCID: 0000-0003-3386-3972

<sup>f</sup>Open Science Lab - German National Library of Science and Technology (TIB), Hanover, Germany. ORCID: 0000-0002-8579-9717

Received: 09 August 2024; Accepted: 20 August 2024

#semantic Climate is an international open research community led by young Indian scientists who use Open Notebook Science to transform information into structured filtered and actionable knowledge. The key project mission is to liberate scientific climate data, making it equitable and freely accessible to everyone. The #semantic Climate community achieves this through two central activities, namely collaborative open notebook science, and citizen engagement. The first activity is research oriented and involves creation of a proof-of-concept software toolkit that uses AI over NLP to transform locked literature (such as PDF documents) into semantic, hypermedia form. This is a non-trivial task, that has haunted developers for over three decades, and the #semantic Toolkit makes complex climate reports not just easily accessible, but also processable by machines, embedded in the Global Knowledge Graph and thereby connected to multilingual resources. The second activity is where the #semantic Climate community engages citizens in climate action and awareness through interactive hackathons, open and transparent working practices, and using Git versioning. From a citizen science perspective, this includes designing community outreach activities (games), giving attribution to all participants, and engaging the wider public in the culture and practices of science (verifiable knowledge, review, data science, modern infrastructure use, etc). This article is an overview of the #semantic Climate community building efforts, and how the project employs strategies, techniques, and ideas from the fields of Open Notebook Science. The open-source software culture and projects follow UNESCO Open Science values, and knowledge justice for the Global South, towards addressing knowledge neo-colonisation.

**Keywords:** Climate change, Semantic web, Free and open source software, Text and data mining, UN Climate, IPCC, UNFCCC, Open science, Citizen science, Semantification, Wikimedia, Wikidata, Linked open data, Knowledge justice, Global south

### Introduction

The paper is intended to offer an overview of the #semantic Climate movement, its community building work and to perform an assessment of these efforts against the project's values and goals, with the aim of learning from work done as a community based open research project. The focus of this community is primarily Climate change, a complex problem with dimensions, not only in the physical, natural and social sciences, but equally so in humanities, arts and policy. Thus, Climate change is an urgent interdisciplinary issue, and addressing the problem requires worldwide collaboration to overcome barriers of cost, culture, and policy that prevent knowledge

from flowing between scientists and society. Scholars around the globe are dissecting its causes, mechanisms and impacts for generating adaptation and mitigation solutions, while the #semantic Climate community believes in making knowledge freely accessible to all, under the premise that every citizen should be able to read Climate related information, and that the full breadth of knowledge about Climate must be made immediately openly available — to everyone, because, openness is a powerful accelerant for finding solutions to our most pressing challenges,

This is a project that has put principles of Open Science and Citizen Science (UNESCO, 2021) at its ethical foundation as an open research project — as operationalised along the lines of using Open-Notebook Science methodologies<sup>1</sup>.

\*Corresponding Author  
simon.worthington@tib.eu

The project has two key values as pillars of its mission — firstly, global citizen access and engagement with scientific climate change literature and secondly, community building. The project goals are creating open-source knowledge liberation software tools to support the first value, to run hackathons for community building, and more recently to work on enabling greater access to UN Climate corpora — as this is the gold standard for climate change scientific knowledge.

#semantic Climate makes open-source software. Community is the ‘how’ in how we make software. And we choose to work as an open and public community to develop this software. The software is a suite of tools for enabling scientific literature access — it supports searching and downloading scientific articles, it can then semantically enrich articles, and uses Wikidata and the Wikimedia tools and communities to create knowledge graphs, translations, data analysis, and language translations, etc. Making software requires development cycles, #semantic Climate engages youth in these development cycles by hosting an internship programme which provides, ideation, alpha testing, documentation, and outreach; engaging peers and supporting institution and communities, at both regional and global scales, and finally; by prototyping, putting ideas into action in a ‘design research’ methodologies of learning by doing.

#semantic Climate is currently (2024) in ‘transition’ from *alpha* to *beta*. This has been marked by two recent events; firstly, engaging with UN Climate corpora the UN IPCC (Intergovernmental Panel on Climate Change) and UNFCCC (United Nations Framework Convention on Climate Change) as of 2022. Secondly, the scaling of community engagement through online and offline live interactive datathons, beginning in 2022, culminating with a pan-India 2024 #semantic Climate Tour. In both cases how we do community work has to be adapted and changed to meet these new challenges<sup>2,3</sup>.

Outlined in the paper are two sections that cover ‘what type of community activities’ are carried out, followed by a detailed ‘history and chronology’ of the #semantic Climate movement. Our aim is to highlight how the project evolved from a handful of people into a dynamic global community composed of undergraduate, postgraduate and Ph.D. scholars who came together to work across rural/urban geographies, nationalities and language boundaries! Together, the sevoluteers and climate enthusiasts have built:

\* A dynamic global community based on collaboration and communal values

- \* An Open platform-independent client-side technology
- \* Semantic resources for text and data mining (TDM), streamlined for Climate
- \* A dictionary system based on Wikidata & Multilinguality

Learning from being a ‘community based’ open research project

The framing question for this paper is: ‘What can be learned from being a community based open research project and how to take these lessons forward to strengthen the project and its community building efforts?’

The question has been broken down into four parts to cover our *learning from community building*: 1. Climate crisis and scaling; 2. Success in terms of what has been produced; 3. Volunteering and questions of equity and access, and; 4. What works for community building and cohesion.

### 1. Climate crises and scaling solutions

Not only is there the alarm of the climate being in crisis, there is also a crisis of climate denial as succinctly hypothesised in Bruno Latour's 2018 book *Down to earth*<sup>4</sup>, but also a crisis climate science digital knowledge infrastructures<sup>5</sup>. The digital knowledge infrastructures needed for climate knowledge simply have not been built at the way that is needed or at scale. Policy makers and scientists have been diligently studying climate change to very high standards, but they are let down by the digital tools available, which results in a barrier to the use and promulgation of their work. It is in the context of ‘digital knowledge infrastructures’ that the #semantic Climate project applies itself. The project makes works intended for scaling solutions: software, using open-source, and outreach events and activities.

### Lessons learned

- Scaling is ironically a ‘waiting game’, waiting for the hockey stick curve *takeoff* to arrive. The waiting game requires a tremendous amount of work where volunteers need to be encouraged that their contributions are helping to create the right circumstances and conditions for success.

- Promoting the learning of professional software development practices — code testing, documentation, organising hackathons, and prototyping, etc. This acts as a strong incentive for participants.

- Connecting the ‘stand-outwork’ of exceptional individual volunteers to the ongoing project is

important. If volunteers make some special software or documentation, additional support work has to be put in place to make sure that others are made aware of the contribution and that new volunteers are able to build on top of the work.

## 2. Success in terms of materials produced

All of what has been produced by the #semantic Climate project has come from volunteers, student researchers and short term interns who at one time or other have been sponsored by the paid internship programmes coordinated by NIPGR (National Institute of Plant Genome Research). Here are some examples. Ayush Garg rewrote the TDM software *pygetpapers* in Python as a teenager, while creating professional standard documentation on the Read-the-Docs platform<sup>6</sup>. Shweeta Hegde and a team of volunteers worked on the UN Climate corpus at the UN Datathon and other events in 2022 and 2023<sup>7</sup>. To date (June 2024) 75 chapters of UN IPCC AR6, semantic identifiers for 100,000+ paragraphs, 2000 climate terms have been aligned with Wikidata by a team of short term interns. Additionally the UN IPCC Glossary of 600 terms has been semantified. Software from the #semantic Climate tech stack has targeted software researchers' needs for 'Automated Literature Reviews' presenting the software in easy to use Jupyter Notebooks on the Google Colab platform and supported with user documentation.

### Lessons learned

- Many improvements come from the intern contributions and this is especially the case in data science where the tooling has been fast catching up with the ambitions of researchers.
- As #semantic Climate software is emerging from an *alpha* state, meaning it's not ready for general use, to *beta* where it can be publicly used — it means we have to focus more on user support as opposed to only engineering and new skills and competences are needed.
- If the number of volunteers at any one time can be maintained then the number of development cycles and the value that comes out of them in terms of improving software is likely to be on par with a directly funded software research project.

## 3. Volunteering and Issues of access, equality, equity

As a community driven project questions of access and equality come into play as we rely on volunteers and interns. Unless measures to support people from

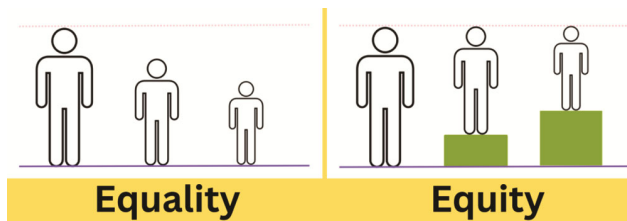


Fig. 1 — Equity vs. Equality: What's the Difference?. Knowledge equity is a solution for addressing imbalanced social systems<sup>10</sup>. Knowledge justice can take equity one step further by fixing the systems in a way that leads to long-term, sustainable, equitable access for generations to come. Source: Wiki Commons [https://commons.wikimedia.org/wiki/File:Equality\\_vs\\_Equity.png](https://commons.wikimedia.org/wiki/File:Equality_vs_Equity.png) | Creative Commons Attribution-Share Alike 4.0 International

background experiencing inequality then volunteers and interns will represent those with the privileged backgrounds, those who have free time. To address this issue the internships are paid and efforts are made to support volunteers, where possible.

In addition to these systemic inequalities a key value of the project is 'knowledge equity and justice' (fig. 1) and here steps need to be taken to enable a more level playing field<sup>8-10</sup>. In terms of systemic inequalities direct action is taken to encourage girls and women on the #semantic Climate programmes, as well as supporting participation from rural areas in India.

To date, funding for this program has been sourced from various national and international bodies and projects including NIPGR, TIGR<sup>2</sup>ESS, Indian Science Academies summer research program, KARYA DST Rajasthan, INYAS young scientists, ICLEI South Asia and others.

### Lessons learned

- Funding an internship programme with paid internships is a direct way to enable access and to prioritise certain groups. Also stating clearly our values and acting on them in terms of equity as advocates has an amplifying effect.
- Carrying out outreach with schools in rural areas and profiling contributions: Code, video, and game prototyping, etc.
- Early Career Research networks play an important role for the climate change community as many young people are drawn to the groups to develop their own way of contributing and learning.

Fig. 1: Equity vs. Equality: What's the Difference?. Knowledge equity is a solution for addressing imbalanced social systems<sup>10</sup>. Knowledge justice can take

equity one step further by fixing the systems in a way that leads to long-term, sustainable, equitable access for generations to come. Source: Wiki Commons [https://commons.wikimedia.org/wiki/File:Equality\\_vs\\_Equity.png](https://commons.wikimedia.org/wiki/File:Equality_vs_Equity.png) | Creative Commons Attribution-Share Alike 4.0 International

#### 4. Community building and cohesion strategies

Open-Notebook Science and open science practice, methods, and values are core to our community building. As with most capacity building or growth for organisations the effort involved is high and requires constant attention. The excitement of events are what provide the drive to community building and core are the hackathons – learning new skills drives the participants and having a goal to present at the end of a hackathon helps in testing out new ideas.

The challenge is to have participation on an ongoing basis by individuals or to have events that others start to pick up and run themselves. This has started with the hackathon “The Climate Knowledge Hunt Hackathon!” with Manav Rachna International Institute of Research and Studies (MRIIRS) which ran as a partnership event with over 50 participants in February 2024, (#semanticClimate, 2024) and other partners are now organising other hackathons in Pondicherry and Kerala over 2024<sup>11-17</sup>. Although as yet the hackathons are not autonomous, there is an effort underway to run these as course curricula.

#### Lessons learned

- Hackathons: It is still a challenge to fix or stabilise the format and technicalities of easy

reproduction, but this has been given a boost by the India 2024 #semantic Climate Tour when so many events were undertaken.

- Flat hierarchy, Open-Notebook Science, open science, and citizen science practices make the project open to the public and school students.
- Spreading the load among volunteers and supporting organisations has to be put in place otherwise individuals and partners are overburdened and fall away.

#### Creating software

The #semantic Climate project develops software for tools to liberate knowledge from locked PDFs and host events where everybody gets a chance to explore the content from world’s open scientific literature and reports through using our tools. (fig. 2)

As of 2022 we have focused on climate literature from the UN, for example the IPCC and UNFCCC, but also other UN climate corpora as well. A number of prototypes have been produced in this context.

Fig. 2: Pygetpapers architecture: <https://pygetpapers.readthedocs.io/en/latest/index.html#architecture> © Copyright 2021, Ayush Garg, Apache licence: Apache License: <https://github.com/petermr/pygetpapers/blob/main/LICENSE>

Figure 3 (fig. 3) is a representative list of our software and prototypes; a full list can be found on GitHub and on a Zenodo community. Figure 4 (fig. 4) depicts some of our current prototypes and demos. A link to all our resources is here: <https://semanticclimate>.

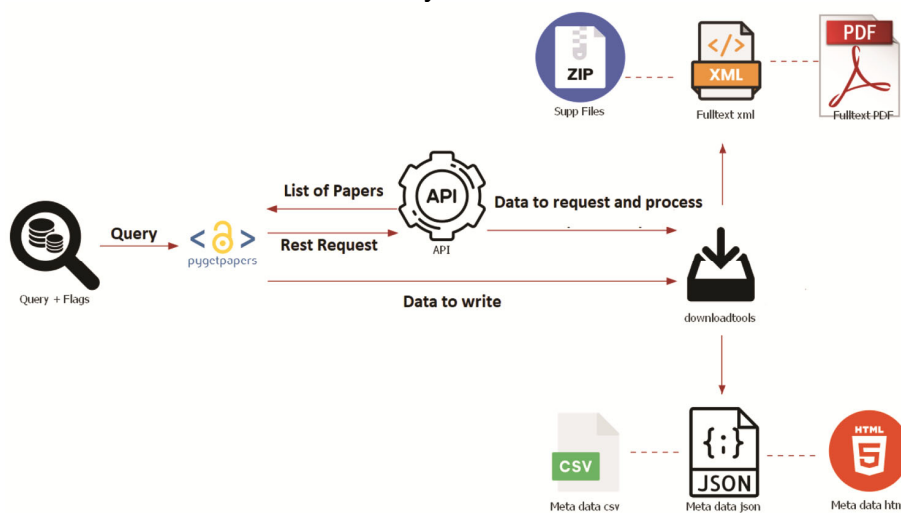


Fig. 2 — Pygetpapers architecture: <https://pygetpapers.readthedocs.io/en/latest/index.html#architecture> © Copyright 2021, Ayush Garg, Apache licence: Apache License: <https://github.com/petermr/pygetpapers/blob/main/LICENSE>

Software

Name	Description	Code	Support
pygetpapers	Searches and downloads articles from repositories. Can be called from other tools.	<a href="https://github.com/petermr/pygetpapers">https://github.com/petermr/pygetpapers</a>	<a href="https://pygetpapers.readthedocs.io/">https://pygetpapers.readthedocs.io/</a>
docanalysis	Ingests <a href="#">CProjects</a> and carries out text-analysis of documents, including sectioning, NLP/text-mining, vocabulary generation. Uses NLTK and other Python tools and <a href="#">spaCy</a> , <a href="#">scispaCy</a> for annotation of <a href="#">entities</a> . Links entities to <a href="#">Wikidata</a> .	<a href="https://github.com/petermr/docanalysis">https://github.com/petermr/docanalysis</a>	README
<a href="#">pyamihtml</a>	Conversion of documents to styled HTML	<a href="https://github.com/petermr/pyamihtml">https://github.com/petermr/pyamihtml</a>	<a href="https://docs.google.com/document/d/1CF68Fw9NytnUH2ZAEecpUeliqXczhn4A/edit">https://docs.google.com/document/d/1CF68Fw9NytnUH2ZAEecpUeliqXczhn4A/edit</a>
<a href="#">pyamiimage</a>	Ingests Figures/images, applies many image processing techniques (erode-dilate, colour quantization, skeletons, etc.), extracts words (Tesseract), extracts lines and symbols (uses <a href="#">sknw/NetworkX</a> ) and recreates semantic diagrams (not finished)	<a href="https://github.com/petermr/pyamiimage">https://github.com/petermr/pyamiimage</a>	<a href="#">READMe</a>
py4ami	Translation of <a href="#">ami3(J)</a> to Python. Processes <a href="#">CProjects</a> to extract and combine primitives into semantic objects. (Note the project is called <a href="#">pyami</a> locally but there is already a <a href="#">PyAMI</a> project so there it is called py4ami)	<a href="https://github.com/petermr/pyami">https://github.com/petermr/pyami</a>	README
Dictionaries	Collection of <a href="#">Wikidata</a> -based dictionaries for scientific annotation and searching	<a href="https://github.com/petermr/dictionary/">https://github.com/petermr/dictionary/</a>	General: <a href="https://github.com/petermr/dictionary/">https://github.com/petermr/dictionary/</a> and Creating dictionaries: <a href="https://github.com/petermr/tigr2ess/blob/master/dictionaries/TUTORIAL.md">https://github.com/petermr/tigr2ess/blob/master/dictionaries/TUTORIAL.md</a>

Fig. 3 — Software used by #semantic Climate

Prototypes and demos

See all resources: <https://semanticclimate.github.io/p/en/posts/resources/>

Name	Description	Code	Support
Content of IPCC reports in this <a href="#">Wikibase</a>	A brief sample and work in progress	<a href="https://kg-ipclimate-reports.wikibase.cloud/">https://kg-ipclimate-reports.wikibase.cloud/</a> and <a href="https://github.com/semanticClimate/IPCC-Queries">https://github.com/semanticClimate/IPCC-Queries</a>	<a href="https://semanticclimate.github.io/IPCC-Queries/">https://semanticclimate.github.io/IPCC-Queries/</a>
IPCC Glossary CSS Paged Media		<a href="https://vivliostyle.vercel.app/#src=https://raw.githubusercontent.com/semanticClimate/glossary-demo/main/ipccglossary.jsonld">https://vivliostyle.vercel.app/#src=https://raw.githubusercontent.com/semanticClimate/glossary-demo/main/ipccglossary.jsonld</a>	<a href="https://github.com/semanticClimate/glossary-demo/blob/main/README.md">https://github.com/semanticClimate/glossary-demo/blob/main/README.md</a>
IPCC Glossary in <a href="#">Wikibase</a>		<a href="https://github.com/semanticClimate/IPCC-Glossary">https://github.com/semanticClimate/IPCC-Glossary</a>	<a href="https://github.com/orgs/semanticClimate/discussions/4">https://github.com/orgs/semanticClimate/discussions/4</a>
IPCC Glossary knowledge graph		<a href="https://semanticclimate.github.io/p/en/ipcc_glossary/">https://semanticclimate.github.io/p/en/ipcc_glossary/</a>	
UNFCCC as CSS Paged Media		<a href="#">Code</a>	
IPCC as CSS Paged Media		<a href="#">Code</a>	
Climate Knowledge Graph	A climate change knowledge graph	<a href="https://github.com/semanticClimate/ClimateKG">https://github.com/semanticClimate/ClimateKG</a>	
70 chapters of <a href="#">ipcc/ar6</a>   cleaned content		<a href="https://github.com/semanticClimate/ipcc/tree/main/cleaned_content">https://github.com/semanticClimate/ipcc/tree/main/cleaned_content</a>	

Fig. 4 — #semantic Climate prototypes and demos

Activity	Use value	Examples
<b>Hackathons</b>	Alpha testing; user priorities. Code has to be ready and working for a deadline.	<a href="#">AI for Climate   A Data Mining Workshop</a> at 'CEPT Research and Development Foundation' (CRDF), Ahmedabad, Gujarat.
<b>Internship programme</b>	Writing software; writing documentation; unit tests; alpha testing; debugging.	<a href="#">Pygetpapers software documentation</a> : <sup>6</sup>
<b>Publication sprints and documentation</b>	Support documentation is much more than just providing a manual. The documentation writing is a form of coding itself as each step is give detailed scrutiny: Alpha testing; writing software; user adoption	<a href="#">Automated Literature Review (ALS)</a> using <a href="#">pyami</a> software with <a href="#">Jupyter Notebooks</a> <sup>25</sup> .
<b>Open-source coding</b>	Writing software; community ideation; debugging; reuse; global collaboration.	At the Climate Knowledge Hunt Hackathon Summer Edition 2023 - Delhi <sup>26</sup> . ( <i>Climate Knowledge Hunt Hackathon Summer Edition 2023, 2023</i> ) Egon and Lars <a href="#">Willighagen</a> were able to code IPCC reports in <a href="#">Wikibase: IPCC Wikibase Queries</a> <sup>27</sup> .
<b>GitHub</b>	Debugging; writing software; user adoption	GitHub allows for collaborative working on the <a href="#">semantification of IPCC Reports</a> ( <i>Conversion of IPCC Documents into Semantic Form, 2024</i> ) <sup>11</sup> .

Fig. 5 — Types of community activities

### The #semantic Climate organisation

#semanticClimate is an open research project which is scientific in all of its practices. Participation in the group is determined by individuals agreeing to our Code of Conduct.

As an organisation #semanticClimate functions along similar lines as a research group or working group. Commonly these are not constituted as organisations in their own right but have a set of agreements among the individuals or partner organisations. This choice of not being constituted is because #semantic Climate identifies as a ‘movement’ — something that others will pick up and emulate. Even if not a legal entity of its own as an open research group there are certain requirements, standards, obligations for participants and members — if not always explicitly stated. #semanticClimate strives to find the right balance between being structured and horizontal. It has been established in the practice of community development that creating inclusive and horizontal communities can easily become undemocratic, dysfunctional, and perpetuate societies imbalances unless checked. As is noted in the famous text “The Tyranny of Structurelessness” by Jo Freeman from the American 1960s feminist movement which was based largely on cooperative

action<sup>18</sup>.

- About: <https://semanticclimate.github.io/p/en/about/>
  - Team and partners: <https://semanticclimate.github.io/p/en/team/>
  - Code of Conduct: [https://www.contributor-covenant.org/version/2/0/code\\_of\\_conduct/](https://www.contributor-covenant.org/version/2/0/code_of_conduct/)
- Community activities: ‘Learning By Doing’

The community activities are where the project values are acted out or at least create a surface to make contact with reality. An important example is helping create greater engagement with climate science knowledge. The approach we have is to create self-replicating engagement strategies analogous to the idea of the Meme (“Meme” 2024)<sup>19,61</sup> — but instead of melodies, catchphrases, fashion our examples are software, hackathons, and games.

What cuts across the community activities and community building is creating environments for learning that are welcoming and supportive. Part of that is about breaking down barriers and hierarchies to recognise and value contributions from all of those taking part. **Figure 5** (fig. 5) is a list of the types of activities we carry out. The list is followed by some in depth descriptions of how our core community activities work and why they are important.

More recently (2024) we have been able to quantify the size and impact of community building



Fig. 6 — semanticClimate Alumni Reunion 2024. CC BY SA 4.0. More images.

efforts. For example the India 2024 #semanticClimate Tour held 13 events over 14 days which involved more than 300 participants. More about each of these activities, including event pages, descriptions, participants lists and photographs, is available on the community events page.

### Community Structure

#semanticClimate is not constituted as a formal organisation and NGO or as an academic working group or research project. This means it doesn't have a secretariat or formal positions or committees and chairs that would usually characterise a working group. The intention is not to have such overheads and to be horizontal in the organisational structure. And the result of the configuration means we rely solely on the community for volunteers and participation. (fig. 6)

To address the imbalance that being volunteer only creates by unfairly affording individuals of privilege to participate — stipends, fee waivers, and outreach are often used to redress these imbalances.

The following list is organised in terms of the current effort:

- **Organisers and volunteers:** At any one time between 5 and 10 organisers are active. The volunteer number varies depending on the events cycles and this will rise to 10 to 30 members or more when running events. Activities will be from (slack) discussions, coding, event organising and presenting, to documentation and teaching, etc.
- **Academic supporters – Organisations, research groups, and individuals:** NIPGR, TIB (TIB –

Leibniz Information Centre for Science and Technology), and the University of Cambridge are the primary institutional supporters. Dr Gitanjali Yadav of NIPGR is the #semanticClimate's *de facto* director by agreement of the organisers, and her affiliations with various national and global academic or non-academic partners help support the extensive internship programme. Other organisations such as CODATA (Committee on Data for Science and Technology) or SDG Solutions Space at Geneva University often have ongoing engagements with the community in their own event programming. Other supporters engage periodically.

- **Partner hackathon organisers:** These members are important for #semanticClimate's operations as they carry out the hard work of setting up and running the events and contribute resources and funds, such as for example, faculty members and investigators from NIPGR, DRIIV, NII, MRIISS, the IIC, Jamia Hamdard University, Amrita AVV University, CEPT Gujarat, University of Delhi and many others.
- **Interns:** NIPGR and the Indian Science Academies' Summer research fellowship program provides the backbone of the internship programme. These are paid internships that run for 8 to 12 weeks and are run in up to two cohorts a year. The interns learn about data science tooling, software development, participating in international events, and more. Efforts are made to ensure women students are recruited and students from rural low-income areas join the programme.
- **Hackathon participants:** Participants numbers can be from less than 15 to more than 50 members per event. Participation in hackathons tend to be short lived for most people, but for others it can bring about long term engagement.
- **Early Career Researchers (ECR):** Due to our need to create new and better software tools for searching and using academic literature — this open science data science tooling is appealing to ECR academics needing such services. Especially popular is TMD and search for open repositories.
- **School student participants:** School students are an important constituency for the project and a big effort is made to engage with a wide variety of schools and NGOs supporting young people's education.
- **Government supporters:** Many government

agencies are willing to support the project or look for ways to engage their stakeholder communities in #semanticClimates work.

- **Climate activists:** This is an area that is not yet developed to a large degree but we anticipate this as being a space we will move into when the tooling is more developed. Climate NGOs are not so aware of open science issues as it would appear their science and data use is further downstream in the knowledge process — not always being scientific, e.g., not attributed or transparent.
- **Citizen science participants:** Although #semanticClimate identifies as a citizen science project — dedicated citizen science (community science) projects is not something we have been able to run as yet. But it is a space the project wants to work in.

### Community Policy & Code of Conduct

The list is a catalogue component we combine for community building and to discover and refine the self-replicating engagement strategies mentioned above.

#### *Policies and methods*

- **Code of Conduct:** Codes of Conduct have come from the open science movement and are there to ensure people act respectfully and with care.
- **Open-Notebook Science:** Open-Notebook Science is an open science practice and means making all research activities open and made available as they are created to ensure there is no insider knowledge<sup>1</sup>.

#### *Activities*

The list has been ordered alphabetically.

- **Attribution:** Attribution is important for rewarding participants for their efforts. Improvements are being made in two ways. Firstly, GitHub updates help as ORCID is now supported, as is software citations, and secondly, Zenodo academic deposits now import GitHub personal data.
- **Being global:** #semanticClimate is led by young Indian scientists. Participation is actively sought in Global South events: CODATA ‘International Training Workshop on Open Science and SDGs’, Beijing, (‘International Training Workshop on Open Science and SDGs, Beijing’ 2023) or the ‘UN Datathon’ organised from Uruguay<sup>20</sup>.
- **Citizen science (Community science):** Public access to climate science is part of the project's

core mission. Events are designed for non-experts to engage in. A citizen science project is being planned around regional climate plans. The City - Open Climate Reader<sup>22</sup>.

- **Open channels:** Synchronous and asynchronous working is supported. Weekday meetings are held 13:30 IST on Slack. And at other times asynchronous messaging also uses Slack and GitHub.
- **Open-source coding and creating software:** The code made by the #semanticClimate community is literally ‘why’ and ‘what’ of the project. The open-source methodologies are inherently social and facilitate community building. Participants are immersed in a continuous learning environment that has room for novices and experts.
- **Distributing software:** Published software is distinct from the coding as users will autonomously engage with the tools made by the team. As the current software is in *alpha* it entails higher degrees of technical knowledge than desired for widespread adoption.
- **Games:** Games are important for providing a way to break down complex climate issues, as well as being a way of socialising at events and hackathons. Currently card games, Snakes and Ladders — originally invented in India<sup>22</sup>, and jigsaws are being prototyped.
- **GitHub:** GitHub facilitates collaboration with a full tooling for group working: versioning, discussion, task management, publishing, and VM provision, etc. GitHub is where the project published its software and website from, and enables global asynchronous working and learning.
- **Hackathons:** These are the main event types used by the project. Usually these happen over two days and are in person events. Teams are set a challenge and then have to create and present a solution at the end of the second day. For participants the benefits are learning and developing their ideas as a prototype. For the software they help motivate the team to develop and document features and release.
- **Internships:** The internship programme has been in place for many years now with the support of NIPGR. The programme acts as the heartbeat of the community — with a constant coming and going of new cohorts. Many students stay on as long term contributors and have created important parts of the software and designed

event formats, etc.

- **Partnerships and funders:** These are important connections to different communities and stakeholders and help pose new questions, or help add additional knowledge not available from the other stakeholders.
- **Publication sprints and documentation:** Creating support material is essential to the software development process as it acts as a kind of alpha testing — where issues and features gets worked out.
- **Schools and young people:** Efforts are made to visit schools and create contexts for young students to engage in the scientific process and in exploring climate change topics.
- **Wikimedia Foundation tools:** The Wikimedia tools and especially Wikidata and Wikibase are some of the most important knowledge infrastructures globally as well as being a platform for public engagement plugging the ‘citizen science’ gap that academia has woefully ignored. An Achilles’ heel of academia and its digital infrastructures is its lack of care to cater for the public who are the ones who foot the bill for its very existence. It is still the case that the majority of climate science is closed source held behind paywalls<sup>5</sup>. In addition, much of academia and research libraries are ‘Global South-blind’ in that they do not mention or acknowledge that the Global South is financially excluded from science knowledge systems<sup>23</sup>.

### Benefits & Use Value of the Community

The project’s mission is an amalgam of two values ‘global citizen access and engagement with climate science literature’ and ‘community’. The reason we need to create software for this purpose is that there are ‘blind spots’ in the current infrastructure provisions that neither the private or public sector have covered. The area we focus on is making literature semantically machine readable so that scientists and the public can search and use it<sup>24</sup>.

The process of making software is where community comes into play. Software is made with rounds of development, as an iterative process. For these development rounds you need to consult with user groups, code, and make technical tests — then repeat the process. In a reciprocal and open way of working with ‘outreach’ and ‘professional engagement’ with a ‘community’ that we nurture — provides the

environment we need to create the software.

Outreach involves working with schools and community groups such as wildlife conservation or climate activists, and professional engagement falls more into the realm of hackathons. Although the two are not mutually exclusive.

**Developing software:** Community building is a way to drive the software development process as it provides a constant stream of ways to test and develop code. Without the community coding would be time consuming and expensive.

**Outreach:** At a visit to schools included in the India-UK BBSRC funded TIGR2ESS program our aim was to meet the young students, especially girls and encourage them to undertake science education. The goal of the visit was also to explore possibilities of rural students being able to undertake short summer internships in Delhi.

The visits were a two way learning exchange — team #semanticClimate had their climate card games and with the students demoing end of term IoT projects and making climate presentations. The experience showed again how immersed in climate issues school students are — as examples the students introduced us to the ‘Climate Clock’, a maker-project for a countdown to 1.5C temperature increase<sup>28</sup>. (*Climate Clock*, n.d.) (fig.7) and reminded us of the importance of expanding STEM to STEAM — and the value of creativity<sup>29</sup>.

School visits: Rural Outreach in Haryana Villages (Climate Education)<sup>14</sup>, February, 2024. Schools included: Govt.Girls Sr. Sec. School, Dahina; Divine Dale International School, Sherpur, and Shivam International School. (fig. 8)

**Professional engagement:** At a weekend workshop with urban planning professionals were able to work with #semanticClimate tools for searching open literature repositories. Using the software tools with planning, climate, and social programme professionals is how we get to understand their needs and contexts.

For example, the importance of running the software on mobiles in a browser and using the software when out on field projects was important to some groups who are working with underprivileged groups, who themselves have no access to devices and the Internet. If the software ran on a mobile, which is possible, then they could use it to explore questions being discussed by the groups they are visiting.

Workshop: AI for Climate | A Data Mining Workshop at ‘CEPT Research and Development



Fig. 7 — Climate Clock: <https://climateclock.world/> – accessed 2024-06-03T16:59:15Z



Fig. 8 — Divine Dale International School, Sherpur, 2024. CC BY SA 4.0. More images.

Foundation' (CRDF), Ahmedabad, Gujarat<sup>30</sup>. [https://semanticclimate.github.io/p/en/past-events/datathon\\_CRDF/](https://semanticclimate.github.io/p/en/past-events/datathon_CRDF/) (*AI for Climate | A Data Mining Workshop*, 2024) (fig. 9)

### Project history and chronology of community activities

The chronology covers the community organisation and activities of the #semantic Climate project.

The timeline reaches back earlier before climate change was explicitly the subject of the project or before the name #semantic Climate was taken on board. The moniker #semantic Climate was adopted in June 2022. The current project stems from, is pervaded by, and is directly based on the earlier work:

1. The basis in plant sciences and the India and UK research collaboration, and the insistence of community building and being open research

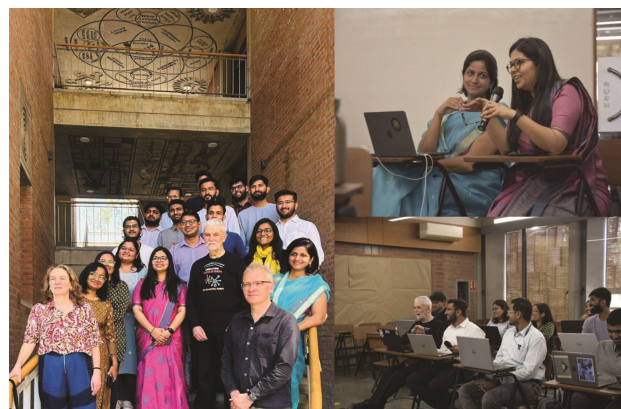


Fig. 9 — A Data Mining Workshop at CRDF, Gujarat. CC BY SA 4.0, 2024. More images.



Fig. 10 — Climate Knowledge Hunt Hackathon Summer Edition 2023 organised by #semanticClimate. CC BY SA 4.0. Link: <https://semanticclimate.github.io/p/en/past-events/climate-knowledge-hunt-may/>

projects, and the open-source software used. The project instantiations contributing to #semantic Climate goes back more than ten years starting from 2013 to this paper's authoring in 2024.

The details listed in the chronology of community activities are a sample of key types of community activity: Hackathons, Barcamps, schools outreach – including games, internship programmes, volunteer engagement from academia, community and activist workshops, engagement with cultural communities, training and teaching and conventional academic events, etc. (fig. 10, fig. 11)

There are two threshold events in the #semantic Climate community building activities which have shaped and consolidated the nature of who the community being engaged has become and how we carry out the community work. Firstly, in the engagement with UN Climate literature corpora from 2022 and attending the SDG Solution Space Hackathon,

and secondly, is the India 2024 #semanticClimate Tour.

The projects listed have all significantly contributed to the accumulated knowledge for community building.

The project is built on open-source software and the foundation for the software comes from the work of Dr Peter Murray-Rust which was consolidated in the enterprise ContentMine. ContentMine was a commercial enterprise that combined text and data mining with semantic technologies to produce knowledge tools and services from scientific literature. Content Mine was engaged in research with Indian colleagues in 2014, see [Figure 12](#) (fig. 12) and one day workshop at SciDataCon 2014 New Delhi, International Conference on Data Sharing and Integration for Global Sustainability<sup>31</sup>.

The software ecology of Content Mine was then used in two research projects TIGR2ESS and CEV Open Plant Oils. The two research projects both had a focus on plant sciences and enabled the software to go through numerous development cycles which added to the tooling's ability to be applied to a dedicated subject domain. The TIGR2ESS project was a UK and India collaboration establishing this aspect as a central pillar of the project going forwards and brought on board the drive and expertise of Dr Gitanjali Yadav from Cambridge (UK) and NIPGR, Delhi. CEV Open Plant Oils had significant contributions from Emanuel Faria an entrepreneur and phytomedicine detective, looking at using the tooling to help identify plants and their essential oil extracts that could be used in products by the company Verriclear. CEV Open Plant Oils had other research activities as well, but the demands of a commercial application as well as combining commercial activity in an Open-Notebook Science project stand as valuable field testing for the overall endeavour of the software and #semanticClimate.

The project Open Climate Knowledge (OCK) was the initial instantiation of a direct climate change mission and its focus was on how to advocate for '100% Open Access for Climate Change research' as the bibliometric understanding of the time (2019) was that only less than 30% of climate science was open access<sup>5</sup>. OCK was a dedicated community building project and formed a working group with the open scholarly communications organisations FORCE11.

OCK's work was quickly curtailed by the COVID pandemic, its activities were put on hold and the team efforts were directed towards helping with the pandemic. The project openVirus was formed and numerous community engagements undertaken. It is worth noting that the pandemic opened the door to community building and breaking down hierarchical



Fig. 11 — #semantic Climate makes its first game paper-prototype of 'Climate Snakes and Ladders'. Youth Interaction about climate action at Garden of Five Senses, Delhi. 2024. CC BY SA 4.0. Link: [https://semanticclimate.github.io/en/posts/garden\\_senses](https://semanticclimate.github.io/en/posts/garden_senses)



Gita Yadav + colleagues

Delhi 2014

Fig. 12 — Content Mine one day workshop at Sci Data Con New Delhi on 2 Nov 2014, <https://github.com/ContentMine/SciDataCon2014> CC0 1.0 Universal

barriers but at the same time interactions were fluid and short lived and this presented new challenges and learning for community building.

#semanticClimate formed as a project in 2022 ([Table 1](#)). The hackathon as part of the The Geneva Trialogue in March 2022 brought the issue of applying #semanticClimate software to help create greater access to UN Climate Change corpora.

### Chronology of community activities

Below is an indicative sample of community

## activities carried out by #semanticClimate (Table 2).

Table 1 — Various projects with #semanticClimate

Project name	Period	Status	Description
Content Mine	2013–Jun 28, 2022	Closed	<a href="https://github.com/orgs/ContentMine/repositories">https://github.com/orgs/ContentMine/repositories</a> The ContentMine is extracting 100 million facts from the academic literature. ( <i>CONTENTMINE LIMITED Company Information - GOV.UK, n.d.</i> ) <sup>31,32</sup>
TIGR <sup>2</sup> ESS	Jul 21, 2017–Jun 21, 2022	Closed	<a href="https://github.com/petermr/tigr2ess">https://github.com/petermr/tigr2ess</a> Joint UK(Cambridge) - India project on Food Security. TIGR2ESS develops and strengthens alliances across a wide ranging network of UK and Indian experts in crop science, hydrology, social science and policy, creating a two-way knowledge exchange partnership <sup>33,34</sup> .
CEVOpen (CEVOpen Plant Oils)	Aug 26, 2019–Jan 6, 2023	Closed	<a href="https://github.com/petermr/CEVOpen/">https://github.com/petermr/CEVOpen/</a> An OpenNotebook project to analyse the composition and activity of oils derived from plants as reported in Open Access literature. <sup>35</sup>
Open Climate Knowledge (OCK)	Oct 5, 2019–May 9, 2020	Ongoing	<a href="https://github.com/OCKProject/climate">https://github.com/OCKProject/climate</a> #OCK – 100% open research for climate change. Open Climate Knowledge (OCK) is an open research project for data mining Open Access (OA) papers related to Climate Change <sup>5,35-40</sup> .
openVirus	Mar 11, 2020–Aug 13, 2021	Closed	<a href="https://github.com/petermr/openVirus">https://github.com/petermr/openVirus</a> Aggregation of scholarly publications and extracted knowledge on viruses and epidemics <sup>41,42</sup>
#semanticClimate	Jun 3, 2022–ongoing as of June 1, 2024	Ongoing	<a href="https://github.com/semanticClimate/semanticClimate">https://github.com/semanticClimate/semanticClimate</a> We develop tools to liberate knowledge from locked PDFs and host events where everybody gets a chance to explore the content in these reports through our tools <sup>43</sup>

Table 2 — Different activities organized and participated by #semantic Climate

Year	Community Activities
2019	<ul style="list-style-type: none"> <li>- Open Climate Knowledge—Open research group. <a href="https://doi.org/10.25815/EB9F-AD4837">https://doi.org/10.25815/EB9F-AD48<sup>37</sup></a></li> <li>- Open Climate Knowledge: 100% OA for Climate Change □ / □ – Generation R. <a href="https://doi.org/10.25815/eqne-xc8238">https://doi.org/10.25815/eqne-xc82<sup>38</sup></a></li> <li>- Sidcot School Darwen Lecture, 2019. Presentation: ‘Climate change (“The Keeling Curve”)<sup>44</sup>.</li> <li>- Twenty-First International Conference on Grey Literature, Hannover, 22-23 October 2019. Presentation: ‘100% open collaborative research for climate change knowledge<sup>36</sup> <a href="https://zenodo.org/records/3517114">https://zenodo.org/records/3517114</a></li> </ul>
2020	<ul style="list-style-type: none"> <li>- Open Energy Modelling Workshop - Berlin 2020, Berlin, 15–17 January 2020.<sup>39</sup> <a href="https://doi.org/10.5281/zenodo.3608216">https://doi.org/10.5281/zenodo.3608216</a></li> <li>- FORCE11 – Open Climate Knowledge Working Group, <a href="https://force11.org/groups/ock/">https://force11.org/groups/ock/</a> (<i>FORCE11 Working Group: Open Climate Knowledge<sup>45</sup></i>).</li> <li>- Open video conference OCK FORCE11 Working Group, 11 Feb '20 (OCK, 2020)</li> <li>- Summer School. “X-Sprint: Climate and Migration” June, Malta.<sup>46</sup></li> <li>- EUvsVirus: Simon Worthington (Director). (2020, April 26). #EUvsVirus: Contentmine open Virus. contentmine openvirus<sup>47</sup>.</li> <li>- Conference. Barcamp Open Science@GeNeMe 2020 (oscibar), Dresden, 07.10.2020. Hamadani, Shweata, Murray-Rust, &amp; Worthington. Open Science: From the crisis of science to science for times of crisis?. <a href="https://doi.org/10.5281/zenodo.4071623">https://doi.org/10.5281/zenodo.4071623</a></li> <li>- Cambridge-India OpenVirus.<sup>48</sup></li> <li>- Open Science: From the crisis of science to science for times of crisis?<sup>48</sup>. <a href="https://doi.org/10.5281/zenodo.4071623">https://doi.org/10.5281/zenodo.4071623</a></li> </ul>
2021	<ul style="list-style-type: none"> <li>- Presentations OMDI2021: Scientific Ontologies in the Digital Age. <a href="https://github.com/semanticClimate/presentations/blob/main/OMDI2021.pptx">https://github.com/semanticClimate/presentations/blob/main/OMDI2021.pptx</a></li> </ul>
2022	<ul style="list-style-type: none"> <li>- Hackathon Piaget: Programme, <a href="https://hackathon-piaget.sparkboard.com/">https://hackathon-piaget.sparkboard.com/</a></li> <li>- Challenge #6: Hacking UN’s Climate Reports with Wikidata<sup>17</sup></li> <li>- Peter Murray-Rust, Ayush Garg, Anuv Chakraborty, &amp; Shweata N Hegde. (2022, July 2). Presentations: MADICES2022. The Language of Materials. <a href="https://github.com/semanticClimate/presentations/blob/main/MADICES2022.pptx">https://github.com/semanticClimate/presentations/blob/main/MADICES2022.pptx</a></li> </ul>

(Contd.)

Table 2 — Different activities organized and participated by #semantic Climate

Year	Community Activities
2023	<ul style="list-style-type: none"> <li>- Hackathon: FSCI 2023   Enhancing Global Impact of Open Scholarship   #semanticClimate.<sup>49</sup></li> <li>- Hannover Open Science Community Meeting: Talking about Open Science: First Open Science Community Hannover Meetup. (2023, April 27)<sup>50</sup>.</li> <li>- Beijing CODATA workshop: International Training Workshop on Open Science and SDGs, Beijing, China   #semanticClimate. (2023, September 1)<sup>51</sup>.</li> <li>- UN Geneva 2023 Hackathons: #semanticClimate. (2023, November 3). UN DATATHON 2023, Montevideo, Uruguay   #semanticClimate<sup>20</sup></li> <li>- UN DATATHON 2023, Montevideo, Uruguay: #semanticClimate. (2023). <i>UN Datathon 2023, #semanticClimate hackathon submission</i> [Jupyter Notebook]. semanticClimate<sup>52</sup>. <a href="https://github.com/semanticClimate/un-datathon-2023">https://github.com/semanticClimate/un-datathon-2023</a> (Original work published 2023) (#semanticClimate, 2023/2023)</li> <li>- UN Library Hackathon: #semanticClimate. (2023). <i>Hack the shifts: UN Library hackathon—#semanticClimate submission</i> [HTML]. semanticClimate. <a href="https://github.com/semanticClimate/hack-the-shifts">https://github.com/semanticClimate/hack-the-shifts</a> (Original work published 2023) (#semanticClimate, 2023/2023)<sup>53</sup></li> <li>- Hack the Hack Hackathon 2023: SDG Solutions Space and CERN - <a href="https://github.com/semanticClimate/un-datathon-2023">https://github.com/semanticClimate/un-datathon-2023</a> (#semanticClimate, 2023/2023)<sup>52</sup></li> <li>- #semanticClimate Hackathon - UNFCCC and IPCC Glossary: #semanticClimate. (2024). <i>Glossary demo: IPCC Glossary: Geneva 2023 Hackathon project</i> [Computer software]. <a href="https://github.com/semanticClimate/glossary-demo">https://github.com/semanticClimate/glossary-demo</a> (#semanticClimate, 2024)<sup>17</sup></li> </ul>
2024	<p>India 2024 #semantic Climate Tour: Full event details can be found here. The topics of the different seminars and events are:</p> <ul style="list-style-type: none"> <li>- Artificial Intelligence in Education” at Jamia Hamdard University<sup>54</sup></li> <li>- “AI in Biological Sciences” at Atma Ram Sanatan Dharma College<sup>55</sup>,</li> <li>- “Unlocking Innovative Solutions: Open Access and Democratisation of Knowledge for Climate, Biodiversity, and Sustainability” at Hansraj College<sup>56</sup>,</li> <li>- Hands-On Workshop on “Data Mining and Semantification”, NIPGR<sup>57</sup>,</li> <li>- The Climate Knowledge Hunt Hackathon! at MRIIRS<sup>12</sup>,</li> <li>- AI for Climate   A Data Mining Workshop: Center for Applied Geomatics, CRDF, Ahmedabad, Gujarat, India<sup>30</sup>,</li> <li>- “AI over NLP” at NIPGR<sup>60</sup>,</li> <li>- Rural Outreach in Haryana Villages (Climate Education), India-UK BBSRC funded TIGR2ESS program<sup>14</sup>,</li> <li>- Youth Interaction about climate action. CBN foundation. The Garden of Five Senses, Said-ul-Ajaib, New Delhi<sup>58</sup>,</li> <li>- Biodiversity Field Trip   Dheerpur Wetland park. Centre for Urban Ecology and Sustainability(CUES) and Cambridge India Research Foundation (CIRF)<sup>59</sup></li> <li>- Swedish-India Center of Excellence for Sustainable Society/City, NIPGR, New Delhi<sup>1</sup></li> </ul>

This is a sample of activities that directly relate to the climate change theme. More events can be found tagged as #community on the #semanticClimate Zotero Group. Associated resources can also be found on Zenodo #semanticClimate community. Where possible presentation Powerpoint files are available here on GitHub.

## Conclusion

Events through the end of 2023 and into 2024 have helped with community developments and which gave a chance to overhaul many parts of the project.

Starting with the Open Science Barcamp hosted at Wikimedia Deutschland in Berlin, followed with the UN Datathon and three other hackathons operating out of SDG Solutions Space at Geneva University and culminating in the India 2024 #semanticClimate Tour which put on thirteen events in fourteen days.

Firstly, was making much greater contact with the Wikidata community and NFDI and TIB/NFDI colleagues working with Wikibase and Wikidata.

Secondly, the Geneva hackathons allowed the team to talk directly with members of the UN IPCC, UN FCCC, and the UN Library. And, thirdly, the India tour hosted by Dr Gitanjali Yadav of NIPGR road tested the hackathon formats as well as connecting a host of institutional supporters.

Assessing the five months of work the immediate community improvements included these four priorities and renewed energy into development tracks.

## Community action priorities

### 1. Attribution and PID use

The number of people, organisations and people involved as well as research outputs means that organising all of these parts needs processes, automation, and easy ways of doing things. Attribution and rewards are important for community building and this is where the use of PID, connecting repositories, and using open science infrastructures comes into play. #semanticClimate is a scientific

research group and not an activists group therefore the maintenance of scientific quality standards must also be checked with policies on review and publishing.

The practice of using the following needs to be embedded in participants' day-to-day working so that retrospective archiving does not need to take place and instead making such records is moved to the front of the queue.

It is worth noting that research projects and organisations need their own linked open data knowledge graphs, for example in Wikibase, but these considerations are beyond the scope of this paper or the #semanticClimate community actions at present.

PID systems proposed to use:

- ORCID
- ORCID and GitHub integration
- Software citation in GitHub using CFF
- Zenodo Community use: <https://zenodo.org/communities/semanticclimate/>
- Zenodo DOI minting
- Zenodo Google CoLab Jupyter Notebook deposits Zenodo and GitHub integration for code releases
- Event DOI deposits using ConfIDent: <https://www.confident-conference.org/>
- Video DOI use and repository deposits using TIB AV Portal: <https://av.tib.eu/>

## 2 Software development: Code, products, and documentation

Community driven software development is run through: Conducting hackathons, prototyping and demoing, running intern programmes, learning programmes, and experts and partners organisation contributions.

- Making hackathons more easily reproducible and potentially able to run autonomously. The #semanticClimate early career researchers have been vital in making this happen. The next step is to create guidance material for these to be run more easily. The #semanticClimate India 2024 #semanticClimate Tour pushed this forwards, for example:
  - Google Colab Jupyter Notebooks have started to be catalogued with PIDs
  - Google Colab Notebooks have new code rationalisation
  - A guide if being written for Automated Literature Review Colab Notebook use
  - Software packages have been refactored
  - Development tracks and community engagement activities. These are prototypes and demos that

the community can contribute to or engage with:

- Software products and games: Automated literature review, and; Snakes and ladders UN Climate corpora: IPCC and UNFCCC corpora; Climate Knowledge Graph, and; UN Climate IPCC Glossary
- Citizen Science: Open Climate Readers - City climate plan pilots

## 2. Autonomous hackathons: Consolidate documentation

- To run hackathons autonomously will take at least two more rounds of hackathons to be possible. A documentation package along the lines of the Carpentries will need to be created and supported on an ongoing basis.

- Tests using GitHub Codespace have been run. Google Codespace allows for a virtual machine to run with a custom environment and a VSCode IDE. Again this is another level of automation that takes the headache of local custom configurations without vendor lockin.

## 3. Starting a Wikimedia Group

- #semanticClimate wants to engage with the global community of Wikimedians, and Wikidata and Wikibase. This has been identified as a priority but is beyond current capacity as it would take a small team to nurture such a community. Only preparatory work can be done for this at present.

## References

- References as collection in #semanticClimate Zotero Group collection “ALIS”: <https://www.zotero.org/groups/2437020/semanticclimate/collections/94RK3L7Y>
- 1 Wikipedia. (2023). Open-notebook science. In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Open-notebook\\_science&oldid=1188104495](https://en.wikipedia.org/w/index.php?title=Open-notebook_science&oldid=1188104495)
  - 2 IPCC. (n.d.). *IPCC — Intergovernmental Panel on Climate Change*. Retrieved June 16, 2024, from <https://www.ipcc.ch/>
  - 3 UN. (n.d.). *UNFCCC*. Retrieved June 16, 2024, from <https://unfccc.int/>
  - 4 Latour B, *Down to earth: Politics in the new climatic regime* (English edition), Polity Press, (2018). <https://hscif.org/wp-content/uploads/2018/04/Latour-Bruno-Down-to-Earthsmall.pdf>
  - 5 Tai T C and Robinson J P W, Enhancing Climate Change Research With Open Science, *Frontiers in Environmental Science*, 6 (2018) 115. <https://doi.org/10.3389/fenvs.2018.00115>
  - 6 Garg A, Smith-Unna R D and Murray-Rust P, (pygetpapers: A Python library for automated retrieval of scientific literature, *Journal of Open Source Software*, 7 (75) (2022) 4451. <https://doi.org/10.21105/joss.04451>
  - 7 Hegde S N, *Hacking UN's Climate Reports with Wikidata*, (2022, June 17). <http://hackathon-piaget.sparkboard.com/project/62a327036c6ee3003da01352>

- 8 Worthington S, Blümel I, Hülk L, Sael M, Ilgaz K, Nguyen Thi M L, Al Nasouh M, Reschner E, Görzen L, Günes Ü, Matern J, Ulrike Ahlborn F and Gaab S, *The Open Science Guide of Guides*. OSL, (2021). <https://doi.org/10.5281/ZENODO.4740163>
- 9 Juda E, Equity vs. Equality: What's the Difference? *GW-UMT*, (2020, November 5). <https://onlinepublichealth.gwu.edu/resources/equity-vs-equality/>
- 10 Maeda J, Design in Tech Report 2019 | Section 6 | Addressing Imbalance. *John Maeda | Design in Tech Report*, (2019, March 11). <https://designintech.report/2019/03/11/%f0%9f%93%b1design-in-tech-report-2019-section-6-addressing-imbalance/>
- 11 #semanticClimate. (2024a). *Conversion of IPCC documents into semantic form*. <https://github.com/semanticClimate/semanticClimate>
- 12 #semanticClimate. (2024b, February 2). *The Climate Knowledge Hunt Hackathon! MRIIRS*. [https://semanticclimate.github.io/p/en/past-events/climate\\_knowledge\\_hunt\\_Feb24/](https://semanticclimate.github.io/p/en/past-events/climate_knowledge_hunt_Feb24/)
- 13 #semanticClimate. (2024c, February 16). *Swedish-India Center of Excellence for Sustainable Society/City*. <https://semanticclimate.github.io/p/en/past-events/Indo-Swedish/>
- 14 #semanticClimate. (2024d, February 21). *Rural Outreach in Haryana Villages (Climate Education)*. [https://semanticclimate.github.io/p/en/past-events/rural\\_outreach/](https://semanticclimate.github.io/p/en/past-events/rural_outreach/)
- 15 #semanticClimate. (2024e, February 22). *Discussion Meeting with UKRI, British Council*. [https://semanticclimate.github.io/p/en/past-events/cirf\\_BHC/](https://semanticclimate.github.io/p/en/past-events/cirf_BHC/)
- 16 #semanticClimate. (2024f, February 25). *semanticClimate Alumni Reunion*. [https://semanticclimate.github.io/p/en/past-events/alumni\\_reunion/](https://semanticclimate.github.io/p/en/past-events/alumni_reunion/)
- 17 #semanticClimate. (2024g). *Glossary demo: IPCC Glossary: Geneva 2023 Hackathon project* [Computer software]. <https://github.com/semanticClimate/glossary-demo>
- 18 Jo Freeman, The Tyranny of Structurelessness, *The Second Wave*, 2 (1) (1972). <https://www.jofreeman.com/joreen/tyranny.htm>
- 19 Richard Dawkins, The Selfish Gene. In *Wikipedia*, (1976). [https://en.wikipedia.org/w/index.php?title=The\\_Selfish\\_Gen\\_e&oldid=1194952061](https://en.wikipedia.org/w/index.php?title=The_Selfish_Gen_e&oldid=1194952061)
- 20 #semanticClimate. (2023, November 3). *UN DATATHON 2023, Montevideo, Uruguay*. [https://semanticclimate.github.io/p/en/past-events/datathon\\_UN/](https://semanticclimate.github.io/p/en/past-events/datathon_UN/)
- 21 Worthington S, *IPCC Reports and City Climate Change Plans: Proof of concept prototype - Open Climate Reader*, (2023b, August 4). <https://semanticclimate.github.io/city-open-climate-reader/>
- 22 Wikipedia. (2024b). Snakes and ladders. In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Snakes\\_and\\_ladders&oldid=1224684955](https://en.wikipedia.org/w/index.php?title=Snakes_and_ladders&oldid=1224684955)
- 23 Karaganis J, *Shadow Libraries*, (2018). <https://mitpress.mit.edu/9780262535014/shadow-libraries/>
- 24 UNESCO. (2021). *UNESCO Recommendation on Open Science*. UNESCO. <https://doi.org/10.54677/MNMMH8546>
- 25 Worthington S, *Automated Literature Review (ALS) using pyami software with Jupyter Notebooks*, (2024). <https://github.com/semanticClimate/automated-literature-review>
- 26 #semantiClimate & NIPGR. (2023, May 18). *Climate Knowledge Hunt Hackathon Summer Edition 2023*. <https://semanticclimate.github.io/p/en/past-events/climate-knowledge-hunt-may/>
- 27 Willighagen E and Willighagen L, *IPCC Wikibase Queries (code)*. IPCC-Queries, (2023). <https://semanticclimate.org/IPCC-Queries/>
- 28 *Climate Clock*. (n.d.). Climate Clock. Retrieved June 3, 2024, from <https://climateclock.world>
- 29 STEM Education Guide. (2024, February 12). *STEM vs. STEAM - What's the difference?* <https://stemeducationguide.com/stem-v-steam/>
- 30 Center for Applied Geomatics, CRDF. (2024, February 24). *AI for Climate—A Data Mining Workshop*. [https://semanticclimate.github.io/p/en/past-events/datathon\\_CRDF/](https://semanticclimate.github.io/p/en/past-events/datathon_CRDF/)
- 31 Murray-Rust P, TDM Workshop, SciDataCon2014, Delhi. *TDM Workshop*. International Conference on Data Sharing and Integration for Global Sustainability, New Delhi, (2014, November 2). <https://github.com/ContentMine/SciDataCon2014>
- 32 *CONTENTMINE LIMITED overview—Find and update company information—GOV.UK*. (n.d.). Retrieved June 12, 2024, from <https://find-and-update.company-information.service.gov.uk/company/10172863>
- 33 Garget J, *Funding for TIGR2ESS Announced* [Text], (2017, December 20) <https://tigr2ess.globalfood.cam.ac.uk/news/funding>
- 34 Youngman A B, *TIGR2ESS 2022 Final Symposium* [Text], (2022, June 13). <https://tigr2ess.globalfood.cam.ac.uk/final-outcomes/tigr2ess-2022-final-symposium>
- 35 Murray-Rust P, *CEVOpen GitHub repository*, (2019c, August 26). <https://github.com/petermr/CEVOpen/commit/9df7046891b1cedc4049fd3eb031f72c759dd898>
- 36 Murray-Rust P and Worthington S, *100% open collaborative research for climate change knowledge (article)*, (2019a). <https://doi.org/10.5281/ZENODO.3517114>
- 37 Murray-Rust P and Worthington S, *Open Climate Knowledge—Open research group*, (2019b, September). <https://doi.org/10.25815/EB9F-AD48>
- 38 Murray-Rust P, *Open Climate Knowledge: 100% OA for Climate Change □ / □ – Generation R*, (2019d, October 1). <https://web.archive.org/web/20191001130053/https://genr.eu/wp/open-climate-knowledge-100-oa-for-climate-change/>
- 39 Murray-Rust P and Worthington S, *OCK – Open Climate Knowledge (poster)—Open Energy Modelling Workshop—Berlin 2020*, (2020, January 14). <https://doi.org/10.5281/ZENODO.3608216>
- 40 Worthington S, *Open Climate knowledge project close date*, (2020b, May 9). <https://github.com/OCKProject/climate/commit/db22d9aa8a475e597adb87ee6f6c11ee9626d43a>
- 41 Murray-Rust P, *openVirus project start date*, (2020, March 11). <https://github.com/petermr/openVirus/commit/7a00c752bb7449770c4c1f5d897d7f46b5ea2e86>
- 42 Hegde S N, *openVirus project end date*, (2021, August 13). <https://github.com/petermr/openVirus/commit/60bf0ada8860b4c629bc6cf8ffec7590bf872c64>
- 43 Murray-Rust P, *semanticClimate project start date*, (2022, June 3). <https://github.com/petermr/semanticClimate/commit/7010ffc4850f2ee980b591ebf05dc9086c612d74>
- 44 Murray-Rust P, *Climate change (“The Keeling Curve”)*. Sidcot School Darwin Lecture 2019, (2019a).

- <https://ockproject.github.io/climate/schools/SidcotSchool201909.html>
- 45 FORCE11 Working Group: *Open Climate Knowledge (OCK)*. (2020). <https://force11.org/groups/ock/>
  - 46 Worthington S, *X-Sprint: Climate and Migration - 100% Open Research for Climate Change*, (2020a). <https://doi.org/10.5281/ZENODO.3892627>
  - 47 Simon Worthington (Director). (2020, April 26). *#EUvsVirus: Contentmine openVirus*. contentmine openvirus. <https://www.youtube.com/watch?v=gBFpiOs7wZI>
  - 48 Hamadani A, *Project Cambridge-India OpenVirus*, (2020). [https://github.com/semanticClimate/presentations/blob/main/Hackathon\\_OpenVirus\\_1.ppt](https://github.com/semanticClimate/presentations/blob/main/Hackathon_OpenVirus_1.ppt)
  - 49 Team #semanticClimate. (2023, July 31). *Hackathon: FSCI 2023 | Enhancing Global Impact of Open Scholarship*. [https://semanticclimate.github.io/p/en/past-events/fsci2023\\_event/](https://semanticclimate.github.io/p/en/past-events/fsci2023_event/)
  - 50 Worthington S, *Talking about Open Science: First Open Science Community Hannover Meetup*, (2023a, April 27). <https://www.tib.eu/en/tib/news-and-events/events/details/talking-about-open-science-first-open-science-community-hannover-meetup>
  - 51 CODATA. (2023, August 28). *International Training Workshop on Open Science and SDGs, Beijing*. [https://semanticclimate.github.io/p/en/past-events/codata23\\_event/](https://semanticclimate.github.io/p/en/past-events/codata23_event/)
  - 52 #semanticClimate. (2023a). *UN Datathon 2023, #semanticClimate hackathon submission* [Jupyter Notebook]. semanticClimate. <https://github.com/semanticClimate/un-datathon-2023> (Original work published 2023)
  - 53 #semanticClimate. (2023b). *Hack the shifts: UN Library hackathon—#semanticClimate submission* [HTML]. semanticClimate. <https://github.com/semanticClimate/hack-the-shifts> (Original work published 2023)
  - 54 India International Centre, *National Conclave on “Artificial Intelligence in Education”* (2024, February 20). [https://semanticclimate.github.io/p/en/past-events/jamia\\_AI/](https://semanticclimate.github.io/p/en/past-events/jamia_AI/)
  - 55 ARSD College. (2024, February 28). *National Seminar on “AI in Biological Sciences.”* [https://semanticclimate.github.io/p/en/past-events/DU\\_seminar/](https://semanticclimate.github.io/p/en/past-events/DU_seminar/)
  - 56 Hansraj College, *Unlocking Innovative Solutions*, (2024, February 28). [https://semanticclimate.github.io/p/en/past-events/cirf\\_event/](https://semanticclimate.github.io/p/en/past-events/cirf_event/)
  - 57 Worthington S, Murray-Rust P, Yadav G and Perez de Eulate R, *Hands-On Workshop on Data Mining and Semantification*, (2024, February 26). [https://semanticclimate.github.io/p/en/past-events/NIPGR\\_workshop/](https://semanticclimate.github.io/p/en/past-events/NIPGR_workshop/)
  - 58 CBN foundation. (2024, February 19). *Youth Interaction about climate action*. [https://semanticclimate.github.io/p/en/past-events/cbn\\_foundation/](https://semanticclimate.github.io/p/en/past-events/cbn_foundation/)
  - 59 Centre for Urban Ecology and Sustainability(CUES). (2024, February 17). *Biodiversity Field Trip—Dheerpur Wetland park*. <https://semanticclimate.github.io/p/en/past-events/biodiversity/>
  - 60 Perez de Eulate R, Murray-Rust P and Worthington S, *NIPGR seminar—AI over NLP*, (2024, February 22). [https://semanticclimate.github.io/p/en/past-events/nipgr\\_symposia/](https://semanticclimate.github.io/p/en/past-events/nipgr_symposia/)
  - 61 Wikipedia. (2024a). Meme. In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Meme&oldid=1224874050>