data.org

Climateverse: Making the Case for Climate Data Analytics

Content

Executive Summary

In response to the escalating climate crisis and its profound impacts on global health, particularly in low- and middle-income countries (LMICs), data.org launched an initiative in May 2023 designed to harness the potential of data-driven climate modeling and analysis to catalyze the creation of actionable digital public goods (DPGs) under the banner of the Climateverse program. This report presents a detailed landscaping of the current state of climate data usage in health decision-making, underscoring the immense potential and significant challenges that define this emerging field, with a focus on LMICs. It addresses the following key questions:

- How can climate data be effectively integrated into health decision-making processes to mitigate climate-related health risks in LMICs?
- What are the critical barriers and gaps in current climate data infrastructure and capacity that hinder effective climate risk management?
- What strategies and tools are necessary to build local capacity and enhance the accessibility and usability of climate data in LMICs?

The Climateverse initiative was conceived through a combination of in-depth desk research and active engagement in the 17 Rooms initiative, a collaborative effort led by the Brookings Institution and The Rockefeller Foundation to propel action towards the Sustainable Development Goals (SDGs). Our methodology involved crafting a detailed problem statement, highlighting the necessity for a coordinated effort to catalyze DPGs, and conducting a series of virtual and hybrid meetings with global experts to define actionable steps for advancing the climate data landscape.

This report details the groundwork laid through these collaborations and then proposes the next steps toward realizing an integrated, scalable platform for climate data analytics, which we call Climateverse. Climateverse aims to empower decision-makers with the tools, data sets, services, and insights necessary to mitigate climate-related health risks.

The findings from our extensive landscaping efforts, which include examining initiatives such as SilverLining's Safe Climate Research Initiative and the World Resources Institute's analytical work, underscore the critical role of data in deepening our understanding of climate dynamics and formulating actionable strategies. Key outcomes include the development of a robust consortium comprised of data.org, Amazon Web Services (AWS), New York University, Radiant Earth, SilverLining, and the University of Cape Town, focusing on local capacity-building and actionable use cases. We identified significant gaps in data literacy, interoperability, and institutional coordination, pointing to the need for structured approaches to integrate climate and health data. Additionally, the importance of democratizing local access to tools, data sets, and infrastructure by investing in open, cloud-based solutions to enable local modeling was emphasized.

Based on these insights, we recommend strengthening data integration and accessibility, building local capacity, promoting interdisciplinary collaboration, enhancing data quality and reliability, and scaling proven solutions. These actions will help establish a scalable, accessible, and interoperable platform for climate data analytics, fostering innovation and enabling informed decision-making to mitigate climate-related health risks in LMICs.

Through this landscaping effort, data.org reaffirms its commitment to leveraging data for social impact, setting the stage for transformative action that aligns with a shared global vision to safeguard public health against the backdrop of a changing climate.

Background

In May 2023, data.org initiated a detailed exploration of the complex interplay between climate data and health outcomes within LMICs. This initiative, designed to create actionable digital public goods (DPGs), marked a significant step towards developing the Climateverse program. Central to this initiative is data.org's commitment to leveraging data's transformative power to tackle global challenges and enhance the field of data for social impact.

Our research, through the 17 Rooms initiative and collaborations with entities like the World Resource Institute, showed that the sector requires greater collaboration between scientists across fields and significant capacity building due to a lack of translational skills. This document presents the findings from our extensive landscaping efforts, which included examining initiatives such as SilverLining's Safe Climate Research Initiative and the World Resources Institute's analytical work. These efforts underscored the critical role of data in deepening our understanding of climate dynamics and formulating actionable strategies. Additionally, we recognized the need to democratize local access to tools, data sets, and infrastructure, which is currently concentrated in the Global North, by investing in open, cloud-based infrastructure to enable local modeling.

By harnessing advanced technologies, data science, and AI, this initiative seeks to develop essential open-source resources and tools that are vital for cultivating a sustainable and localized community of practitioners. This strategy intends to magnify the impact of climate interventions, ensuring that innovations are accessible and pertinent to those on the front lines of the climate crisis.

Through collaborative efforts and strategic insights, this landscaping document strives to propel a substantial leap forward in our collective endeavor to protect our planet and improve public health. It is designed to map the current landscape, identify critical gaps, and propose effective strategies to bolster global climate resilience, focusing on integrating climate data into health decision-making frameworks within LMICs.

Methodology

Climateverse was conceived through a strategic combination of in-depth desk research and active engagement in the 17 Rooms initiative, a collaborative effort led by the Brookings Institution and The Rockefeller Foundation to propel action towards the Sustainable Development Goals (SDGs). This approach facilitated a detailed landscaping of existing capabilities and needs, setting the stage for addressing climate challenges with innovative data-driven solutions.

Design Process Steps:

1. **Initial Research and Engagement**
    - Conducted extensive reviews of existing literature, climate data initiatives, and technological advancements to understand the current state of climate data usage.
    - Collaborated with global experts to discuss and identify key challenges and opportunities in climate data integration, focusing on SDG 9—aiming to build resilient infrastructure, promote sustainable industrialization, and foster innovation through unlocking climate data for impactful use.
2. **Problem Identification**
    - Identified critical barriers in utilizing climate data effectively, including gaps in data literacy, interoperability, and institutional coordination. This process aimed to bridge the gap between "data supply" from technology providers and "data demand" from Africa-based practitioners and academics.
    - Recognized the significant capacity-building gap due to the lack of translational skills, indicating the potential emergence of a new interdisciplinary field.
3. **Strategic Collaborations and Meetings**
    - Collaborated with key partners like The Rockefeller Foundation, Brookings Institution, and Amazon Web Services to facilitate a series of virtual and hybrid meetings with global experts over the summer of 2023.
    - Focused on defining concrete steps to propel the climate data landscape forward, with discussions transcending theoretical debate and focusing on practical implementation.

4. **Blueprint Development**
    - Developed a comprehensive plan for creating an integrated, scalable platform for climate data analytics, influenced by the 17 Rooms process. This blueprint outlined strategic goals for integrating climate data into health decision-making frameworks within LMICs.
5. **Consortium Formation**
    - Formed a group comprised of data.org, Amazon Web Services (AWS), New York University, Radiant Earth, SilverLining, and the University of Cape Town. This consortium is now at the forefront of the Climateverse program, championing actionable use cases and local capacity-building efforts.

Through these steps, Climateverse aims to transform climate modeling and data analysis in LMICs, establishing a solid foundation for impactful, sustainable, and scalable solutions. The development journey of Climateverse underscores a deep commitment to leveraging data, technology, and cross-sector collaboration to tackle climate challenges, ensuring that innovations are accessible and pertinent to those on the front lines of the climate crisis. This holistic approach fosters local expertise and enhances data governance, providing a clear roadmap for the future trajectory of Climateverse.

Outcomes and Key Insights

An Emerging Climate Data Consortium and Solution Package

Building on this foundation, the Climateverse consortium is advancing towards an integrated initiative that unites the strengths of each participating organization. This collaborative effort is crucial for ensuring interoperability across projects, enhancing coordination, and significantly expanding the scalability of climate data solutions tailored for LMICs. Currently, the focus is on the assessment of feasibility, robust stakeholder engagement, and strategic alignment of implementation tactics to ensure the interventions are both actionable and sustainable.

The consortium has identified critical challenges in climate data management within LMICs, which include computational demands, data scarcity, barriers to data sharing, and the need for enhanced data capacity and application in real-world scenarios.

Cloud for Climate - Global South Computing Hub and Spoke Initiative (Lead: SilverLining)

Leveraging the power of artificial intelligence (AI) and machine learning (ML) is fundamental to the analysis of large datasets and predictive modeling for effective climate risk management. Climate data is computationally intensive and requires costly infrastructure for effective analysis. Significant disparities

exist in accessing crucial geophysical models and supercomputing resources, especially in LMICs, resulting in poor-quality local climate information and reduced influence on global climate governance.

To address these challenges, the creation of a Global South Climate Computing Hub is proposed. This hub would leverage AI/ML to eliminate cyberinfrastructure barriers and strengthen climate risk adaptation. It would serve as a center for capacity building, organizing workshops and tutorials to enhance AI/ML competencies and ensure sustainable, locally-driven climate solutions. This initiative aims to improve climate capabilities in the least developed countries through scalable, high-performance computing and build foundational workforce skills in data science.

Community-based Capacity Building to Create a Social & Environmental Data Resource (Lead: NYU)

Addressing climate challenges requires inclusive data that encompasses both social and environmental factors. Often, this data is unstructured or collected inappropriately from local communities, particularly in tribal and Indigenous contexts. This proposal outlines a systematic approach to capture, structure, and integrate unstructured data by combining local knowledge with global data science best practices, including AI/ML techniques. The development of a new template for responsible data collection in tribal and indigenous communities is proposed, ensuring greater accuracy and applicability. It emphasizes capacity building in modern data analytics, empowering data scientists to leverage AI/ML for downscaling global climate data to more precisely address local climate change. The goal is to identify social factors directly relevant to climate and health decision-making, thereby bridging the gap between complex, multidimensional data and effective climate modeling.

Source Cooperative (Lead: Radiant Earth)

To address climate change impacts in Africa, diverse and well-organized data products are essential for data-driven interventions. Launched and currently in beta, Source Cooperative will provide a non-profit, neutral data publishing utility, offering a common interface for trusted organizations and individuals to share data products. It emphasizes neutrality, scalability, and non-proprietary data sharing, ensuring broad accessibility and support for various data types. Leveraging generic public cloud object storage services, Source Cooperative will facilitate data accessibility and large-scale analysis, with a particular focus on AI and ML based approaches. The aim is to make public cloud object storage widely accessible, expanding access to vast datasets and enabling large-scale ML training workloads in public cloud environments. This approach opens doors for a wide array of organizations to train ML models tailored to their specific community needs.

A Climate Change Research Hub for Africa (Lead: University of Cape Town)

The urgency of addressing global warming's impact in Africa is undeniable, yet infrastructure and funding challenges hinder climate research on the continent. To tackle this, an African Climate Change Research Hub is proposed at the Climate System Analysis Group at the University of Cape Town. The hub will

nurture a multidisciplinary climate science community and drive essential research in core climate science and modeling. It will place a strong emphasis on comprehensive data science training, encompassing data analysis, ML/AI methods, data visualization, and cloud computing. Collaborating with the "Cloud for Climate - Global South Computing Hub," it will empower researchers to conduct cloud-based modeling and data analysis, bridging the gap between climate science and policy. The group will foster partnerships and position African researchers as leaders in climate resilience, equipped with advanced AI and ML skills to address climate change effectively.

Coordinating a Unified Strategy for Global Climate Data Services (Lead: data.org)

As the urgency of climate change continues to press upon global health, livelihoods, food security, and the built environment, comprehensive mitigation, adaptation, transition, and resilience strategies are critical. Governments, businesses, and social impact organizations are actively seeking effective interventions but often lack the actionable insights needed to implement them. At the same time, the consortium is building a robust climate data infrastructure, an essential layer is required to translate this complex information into practical solutions and strategies.

This part of the initiative, led by data.org, will develop and coordinate a suite of use cases aimed at solving key problems identified by stakeholders in areas such as public health, urban policy, and global climate science. These use cases focus on adapting to climate change impacts, transitioning to resilient practices, and building long-term resilience within vulnerable communities, especially in LMICs.

Beyond the development of these use cases, data.org will facilitate several pilot projects that apply these insights in real-world scenarios, demonstrating how varied challenges can be addressed through strategic data application. The outcomes include a range of documented strategies, such as case studies and playbooks, which illustrate successful interventions and guide replication across different regions and contexts.

Strategic Coordination for Robust Implementation

Recognizing the need for substantial investment in interoperability and coordination, the consortium has prioritized these elements as the cornerstone of our strategy. Enhanced interoperability is crucial to ensuring that the suite of solutions developed can be robust, harmonized, and integrated with other tools and systems, avoiding the pitfalls of isolated datasets and duplicative solutions witnessed during the COVID-19 crisis.

To achieve this, data.org will lead the consortium's coordination efforts, focusing on:

- Technical Coordination: Ensuring integration, interoperability, and economies of scale within and among the various components of Climateverse. This involves streamlining technical aspects such as data sharing, model integration, and tool compatibility.
- Partnership Coordination: Serving as a platform for partnerships, harmonizing proposals, supporting grant acquisition and funder coordination, and ensuring alignment of interests among consortium members. Ensuring that each stakeholder has the requisite resources and collaborative support to successfully implement their projects.
- Knowledge Network (Ecosystem) Coordination: Coordinating a community of local and international practitioners, including identifying and recruiting collaborators; convening meetings, co-working sessions, and roundtables; and advancing the global outreach of the initiative. Amplifying the work of ecosystem partners to reach new audiences and promote new collaborations.

Learning from our experience in developing Epiverse, we understand the importance of modularity. For Climateverse to be robust and sustainable, it needs to be built gradually with modular components that do not have too many dependencies. This approach ensures that the system is adaptable and scalable, able to grow organically as new challenges and opportunities arise.

These strategic coordination efforts are vital for realizing a unified and effective suite of solutions that serve as a global platform for locally informed climate modeling, analysis, and decision-making.

Leveraging the Collective and Building Interoperability via Climateverse

The Climateverse program will build on the foundational investments in climate and health data capacity through the Capacity Accelerator Network (CAN), which trains a new generation of purpose-driven data practitioners. Funded by Wellcome, CAN is active in the US, Africa, India, Latin America, and Asia Pacific, fostering an enabling environment for a new type of data practitioner with expertise in both climate and health domains. Resources developed under CAN are adaptable and are being leveraged to support Climateverse, promoting active engagement across its network.

Additionally, Climateverse will build upon the successes of Epiverse, a program initiated in August 2021 with support from Wellcome and The Rockefeller Foundation, which is designed to develop open software tools for outbreak analytics. Epiverse's approach to bridging significant data utilization gaps during health crises has informed the strategies now being adapted for climate data challenges. This initiative's success has showcased our capacity to foster robust partnerships and maintain tool sustainability and adaptation across health and now, climate sectors.

Aiming to replicate these successes, data.org is now focusing on using institutional knowledge from Epiverse to develop tools that facilitate climate change adaptation and resilience. These tools will also be used to model the impact of climate on health, supported by existing partnerships with entities like the World Health Organization and several national ministries of health. This approach addresses a critical gap, as fewer than a quarter of all countries currently use climate data to inform public health strategies, a missed opportunity that Climateverse is poised to rectify.

Recommendations

Based on the detailed landscaping of the current state of climate data usage in health decision-making and the identified key insights, we propose the following recommendations to drive transformative actions:

- **Enhancing Data Quality and Reliability:** Address data scarcity and quality issues by implementing robust data collection, validation, and standardization processes. This includes developing new templates for responsible data collection in tribal and indigenous communities to ensure greater accuracy and applicability.
- **Strengthening Data Integration and Accessibility:** Enhance the integration of climate and health data by developing standardized data protocols and promoting open, cloud-based solutions. This will ensure that diverse data sources are accessible and usable for decision-makers in LMICs.
- **Building Local Capacity:** Invest in targeted capacity-building programs to empower local stakeholders, including public health professionals, data scientists, policymakers, and other interdisciplinary leaders. Establish training hubs in Africa and India to foster local expertise and support the development of sustainable, locally-driven climate solutions.
- **Promoting Interdisciplinary Collaboration:** Foster collaboration across scientific disciplines and sectors to bridge gaps in data literacy, interoperability, and institutional coordination. Encourage partnerships that leverage the collective strengths of various organizations to create a unified approach to climate data analysis.
- **Scaling Proven Solutions:** Pilot projects that demonstrate the practical application of climate data in addressing specific challenges should be scaled and adapted to other regions. Document successful interventions through case studies and playbooks to provide a roadmap for replication, enabling LMICs to benefit from proven strategies and solutions.
- **Infrastructure and Access:** Address the need for investment in open, cloud-based infrastructure to enable local modeling, rather than relying solely on globally produced models. This includes investing in scalable, high-performance computing resources to support climate data analysis.

These recommendations are designed to establish a scalable, interoperable platform for climate data analytics, fostering innovation and enabling informed decision-making to mitigate climate-related health risks in LMICs. Through these actions, data.org reaffirms its commitment to leveraging data for social impact, setting the stage for a sustainable and equitable future in the face of a changing climate.

Climate Datasets and Tools

Name	Description	Links
Global Temperature Data (Monthly Averages)	Berkeley Earth combines its land data with a modified version of the HadSST ocean temperature data set. The result is a global average temperature data set.	Dataset 1 Dataset 2
Global Temperatures (Land Only)	Berkeley Earth's primary product is an analysis of summary air temperatures over land. The following files and links provide a time series that summarizes those results for the global average.	Dataset 1 Dataset 2 Dataset 3
Global Gridded data	Datasets are also provided in a gridded NetCDF format. Two types of grids are provided, a grid based on dividing the Earth into 15984 equal-area grid cells and a latitude-longitude grid. The equal area grid is the primary data format used in most of our analyses and provides generally smaller files; however, that format may be less convenient for many users.	Dataset
Global temperatures (High-resolution Time series)	Berkeley Earth is preparing an updated high-resolution temperature data set. Some materials related to this update are being released as a beta version to allow for additional feedback before final publication. Results are preliminary and subject to change without notice.	Dataset
Climate Change Data	Data from World Development Indicators and Climate Change Knowledge Portal on climate systems, exposure to climate impacts, resilience, greenhouse gas emissions, and energy use.	Dataset

Temperature Time Series for India	Monthly, Seasonal, and Annual Minimum Temperature in India from 1901 onwards	Dataset
Land Surface Temperature at Night	MODIS (or Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths (see MODIS Technical Specifications). This map shows the temperature of Earth's lands during the nighttime.	Dataset
Global 1-km Sea Surface Temperature (G1SST)	JPL OurOcean Portal: A daily, global Sea Surface Temperature (SST) data set is produced at 1 km (also known as ultra-high resolution) by the JPL ROMS (Regional Ocean Modeling System) group.	Dataset 1 Dataset 2
Dataset Card for UK Nimrod 1km Rainfall Radar Dataset	This dataset contains UK Nimrod rainfall radar data for 2016-2019 as used in the Skillful Precipitation Nowcasting Using Deep Generative Model of Radar paper by DeepMind. This dataset is an unofficial mirror of the open-sourced dataset available here: gs://dm-nowcasting/datasets/nowcasting_open_source_osgb/nimrod_osgb_1000m_yearly_splits/radar/20200718	Dataset
NCAR Climate Data Guide	NCAR's *Climate Data Guide* provides concise and reliable information on the strengths and limitations of the key observational data sets, tools, and methods used to evaluate (or initialize or force) Earth system models and to understand the climate system.	Tools

Climate Analytics	Explore our open-access tools for policymakers and researchers working on climate impacts and action.	Tools
OS Climate	The Linux Foundation's OS-Climate provides open-source data and software for assessing climate risk including Physical Risk & Resilience (led by BNP Paribas), Portfolio Alignment (led by Allianz), and Transition Analysis (led by Airbus) tools. These tools allow financial institutions to assess and manage climate risks, as well as make investment decisions that work towards their net zero ambitions.	Dataset and Tools
Climate Data Tools	Climate Data Tools CDT is a set of utility functions for meteorological data quality control, homogenization, and merging station data with satellite and other proxies such as reanalysis, all functions are available in the GUI mode.	Tools