

Harmonizing climate, environment and health data: an introduction to the HARMONIZE 4health toolkit

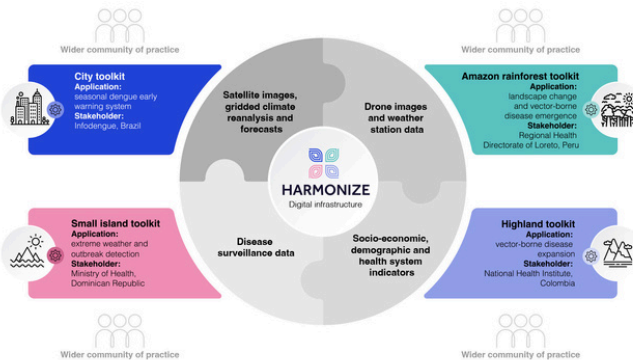
Daniela Lührsen; Anna B. Kawiecki; Carles Milà; Raquel Lana; Emily Ball; Alba Liabrés Brustenga; Raúl Capellán; Emma Roberts; Erick Lantigua; Diego Irreño Torres; Ingrid Vanessa Mora Sanchez; Juan Montenegro Torres; Yomali Ferreyra; Edgar Joao Manrique Valverde; Antony Barja; Marcos Rodrigues; Adeline Marinho Maciel; Yuri Domaradzki Moreira Nunes; Miguel Monteiro; Miriam Matos; Gabriel Carrasco-Escobar; Mauricio Santos-Vega; Rachel Lowe
Barcelona Supercomputing Center, Universidad de los Andes, Instituto Nacional de Pesquisas Espaciais, Universidad Peruana Cayetano Heredia, Instituto Dominicano De Meteorología

INTRODUÇÃO

Extreme weather events, environmental degradation, and social inequalities are increasing the vulnerability of communities in climate change hotspots to infectious disease outbreaks. However, the data required to assess and respond to these risks are often scattered across sources, stored in incompatible formats, and difficult to integrate. In many low-resource settings, limited digital infrastructure and data science capacity further constrain analysis, with valuable time spent wrangling data instead of generating insights.

OBJETIVOS

The HARMONIZE project aims to streamline access to, cleaning of, and integration of key datasets relevant to climate-sensitive disease surveillance. The resulting tools are designed to enable reproducible, cost-effective workflows that support timely, actionable outputs such as risk maps, early warnings, and informed public health responses in vulnerable regions.



METODOLOGIA

We began by mapping the data landscape for climate-sensitive disease. Through expert consultation, we built a data inventory and identified the main data types: disease case data, climate indicators, land use patterns, and socioeconomic variables. We analyzed the steps required to standardize and integrate each data type and, due to the difference in data type, decided to create one tool for each data type. Given the heterogeneity of sources and formats, we developed one tool per data type. To ensure usability, we also evaluated the technological context of intended users in low-resource settings, including hardware access, internet connectivity, and data science experience.

RESULTADOS



data4health cleans, filters, and aggregates line-list disease data. It supports key pre-processing steps such as variable standardization, quality checks, and spatiotemporal aggregation.



clim4health obtains, transforms and exports climate data. Transformations includes calibration, spatial and temporal harmonization, index calculation, and visualisation of climate data.



socio4health is an extraction, transformation and loading (ETL) classification tool designed to collect and merge data from multiple sources, focusing on sociodemographic and census datasets



land4health calculates and extracts remote sensing metrics for spatial health analysis. It obtains areal or zonal statistics of key indicators and covariates, ideal for modelling infectious diseases.



cube4health builds earth observation data cubes for health response systems, allowing users to visualise and share data of their study area obtained through the other tools.

All tools are accessible through the tools website, which also hosts comprehensive documentation and training sessions delivered to over 20 researchers, academics, and local government partners at a training course in Rio de Janeiro in November 2025. While only part of the suite is currently published, users can already work with the tools by following the online training materials. Full public release of the complete 4health toolkit is planned for February 2026.



CONCLUSÃO

The HARMONIZE toolkit offers a practical, scalable solution to a longstanding bottleneck in data-driven disease surveillance: the time-consuming task of gathering and harmonizing multi-source datasets. By making this process faster, more reproducible, and more accessible, especially in low-resource, high-risk settings, the toolkit will support stronger, data-informed responses to infectious disease threats. As climate and environmental crises intensify, tools like this will be essential to building more resilient surveillance systems and protecting public health where it's needed most.

REFERÊNCIAS

Project website: <https://www.harmonize-tools.org/>
Tools website: <https://harmonize-tools.github.io/>