Getting Started with the Climate Adaptation Data Platform

Brian Lee Yung Rowe

2024-02-12

Overview

The Climate Adaptation Data Platform (CADP) accelerates climate adaptation initiatives by bringing together multiple stakeholders to addresss multiple climate-related challenges. The primary stakeholders are policymakers/governments, farmers, and individual members of the community. The CADP provides tangible value to each stakeholder.

The magic of the CADP is that it breaks through policy paralysis by creating the device and data infrastructure that policymakers need to drive widespread climate adaptation efforts. Our hypothesis is that it's unrealistic for municipalities to deploy and operate sensor networks. It's also too complex an undertaking for community groups to take on.

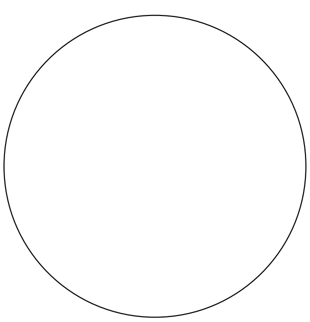
CADP provides an alternate path where individuals and businesses purchase devices that incrementally build out the infrastructure needed for policymakers.

This document discusses the different use cases that CADP is designed for and the value people get from these use cases. Where applicable, a distinction will be made between the value individuals get from the platform versus the value policymakers get.

After describing the use cases, we move into more technical territory so people can learn how to use and develop the platform further. First is a high-level view of the physical infrastructure and the relationship between a device network and the CADP. Devices are responsible for monitoring environmental conditions and sending the collected data to the platform.

Use Cases

Every device powers multiple simultaneous use cases. We call this use case stacking. This is different from re-using or re-purposing a device. With use case stacking, multiple applications run concurrently. The supported use cases depend on the sensors attached. The two main public health use cases only need the base set of sensors, which means anybody with an account can get hyperlocal heat risk and mosquito abundance information.



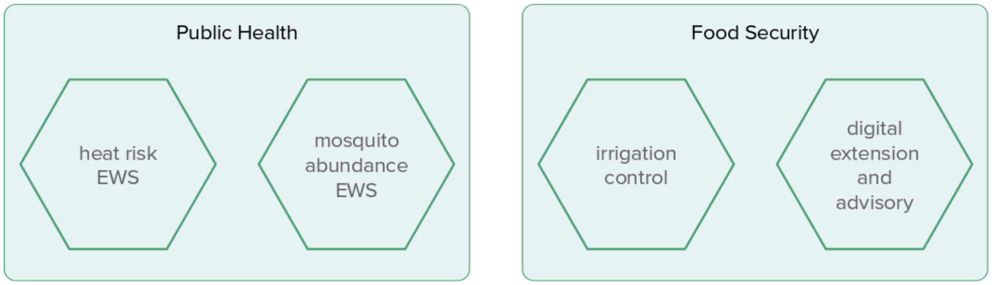


Figure 1: Use case stacking enables the CADP to power multiple simultaneous applications across public health and food security.

Public health

Early warning systems

The genesis of this project was the death of 11,000 people in Libya due to a medicane (Mediterranean hurricane) that caused a dam to fail. People were caught off guard because as a failed state, Libya has no functioning weather agency. It turns out that most of Africa has limited weather stations. Most official weather agencies only send out seasonal forecasts.

Weather forecasts give people critical time to plan and prepare for major weather events. Building on top of weather forecasts are early warning systems (EWS). These systems apply rules, or risk protocols, to forecasts to determine the level of risk people are exposed to. EWSes can operate on any type of forecast and includes heat risk, air quality risk, the spread of wildfire, the spread of infectious disease, and the spread of pests like the ash borer.

The UN declared a mandate that everyone should be covered by an early warning system by 2030. Some $3 billion dollars have been earmarked for this effort. While many countries have early warning systems, many are outdated, no longer work, or cover a limited number of risks.

The CADP provides the infrastructure to quickly build EWSes and fulfill the UN's vision.

Heat risk

Survivability is the upper limit of temperature and humidity that humans can withstand. Humidity has a significant impact on the maximum temperature that humans can survive in. With high humidity, the body loses its ability to regulate temperature via perspiration.

In some places, climate change is making heat waves more extreme and pushing temperatures to the survivability level. Even if that limit isn't reached, whether those conditions are livable is another matter. Livability limits help people understand what activities are possible given temperature and humidity (and sun exposure). Figure **??** compares livability limits for two population groups and whether the activity takes place in direct sun or shade.

Heat risk builds on the concept of livability. The goal is to provide an EWS that helps people understand their heat risk expoure and make decisions about what activities to do or how to mitigate heat risk during certain activities.

Individuals can get personalized heat risk guidance for any device in CADP. Alerts can be delivered via email or via the mobile app.

Users can enter some health information to get better alerts. This information is saved but segregated from personally identifiable information (PII).

Relevant health information includes

- age

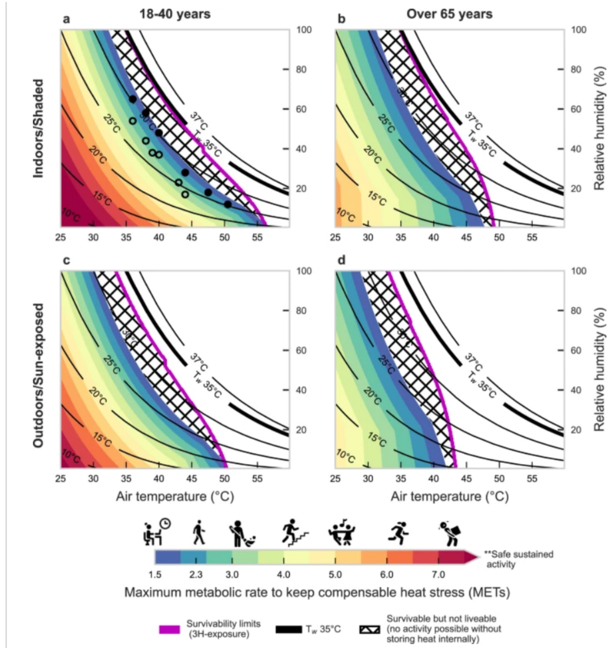


Figure 2: Livability limits for different METs at different temperature and humidity levels. Source:

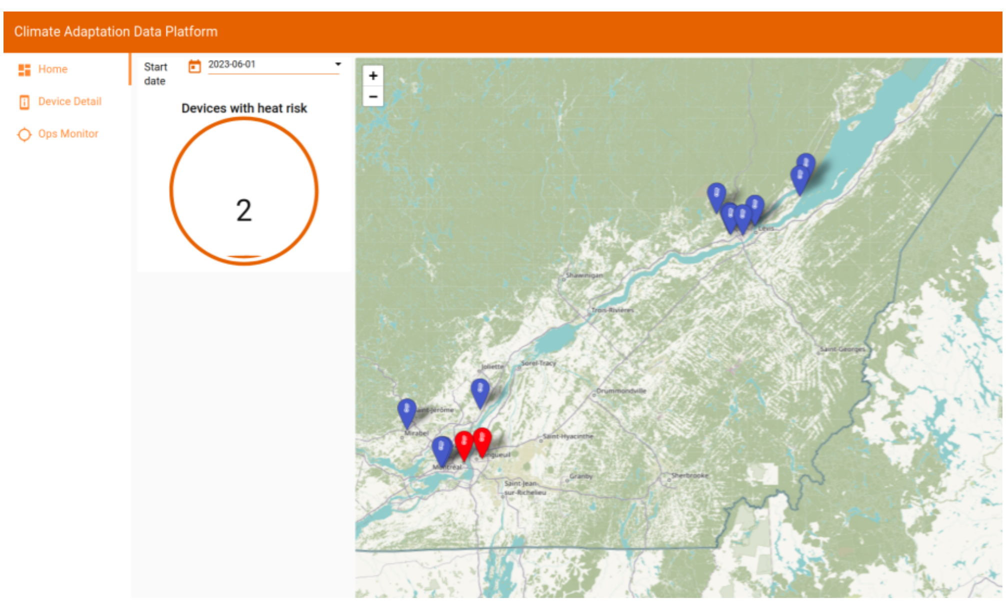


Figure 3: A screenshot of a policy dashboard showing active heat risk alerts for two devices.

- BMI (weight + height)
- diabetes
- heart issues

The conversion formula between METs (M) and calories is

$$M\frac{3.5m}{200} = kcal/min, \tag{1}$$

where m is mass in kilograms.

Mosquito-borne infectious disease

Mosquitos are a common disease vector and can carry numerous infectious diseases, such as malaria, dengue fever, zika virus, yellow fever. Disease outbreaks are driven by the growth of a mosquito population. To limit the severity of an outbreak, mosquito surveillance is conducted to monitor the spread of mosquitos. By tracking mosquito abundance, it is possible to predict how large an outbreak might be.

Direct observation of mosquitos is an involved process. It requires specialized devices to attract and trap mosquitos. Usually these devices need periodic cleaning.

A simpler approach is to use a mosquito abundance model and build risk protocols on that model. These can be calibrated with physical traps (future). Figure 4 shows a system of partial differential equations (PDEs) that describe the change in mosquito abundance given a few parameters. Of importance is that both temperature and humidity drive the population growth of mosquitos.

$$\begin{aligned}
\frac{dE}{dt} &= \beta A_o - (f_E + \mu_E)\,E \\
\frac{dL}{dt} &= f_E E - \mu_L \left(L + \frac{L}{k_L}\right) - f_L L \\
\frac{dP}{dt} &= f_L L - (f_P + \mu_P)\,P \\
\frac{dA_{em}}{dt} &= P f_P \sigma e^{\mu_{A_{em}}\left(1+\frac{P}{k_P}\right)} - (\mu_{A_{em}} + \mu_r)\,A_{em} + f_{A_{em}} A_{em} \\
\frac{dA_b}{dt} &= f_{A_{em}} A_{em} + f_{A_o} A_o - (\mu_A + \mu_r)\,A_b - f_{A_b} A_b \\
\frac{dA_g}{dt} &= f_{A_b} A_b - \mu_A A_g - f_{A_g} A_g \\
\frac{dA_o}{dt} &= f_{A_g} A_g - (\mu_A + \mu_r) A_o - f_{A_o} A_o
\end{aligned}$$

Figure 4: A system of partial differential equations that model mosquito population growth based on environmental conditions.

We can thus use the temperature and humidity forecasts available in the CADP to make mosquito abundance forecasts. Figure 5 shows a graph of a device in the CADP. The green line represents observational data, while the orange time series is the forecast produced by the CADP.

1. Install device in observation area
2. Enable mosquito abundance alerts

Precision agriculture

Dry soil forecasting

Forecasting dry soil is the first step in irrigation control. The forecast is simply the moisture content of the soil.

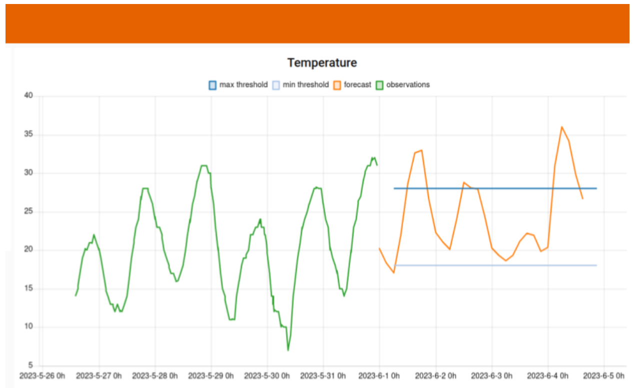


Figure 5: A device detail view showing historical temperature, a temperature forecast, plus heat risk thresholds.

Watering schedules

Watering depends on the type of plant growing. Some plants need a lot of water while others don't. Many plants don't like wet roots, so watering cannot be too frequent. We need to develop these control models for each plant.

Crop development

Water needs change based on the maturity of the plant. Beans need moist soil when pods are forming. To form large roots, beets need minimal water during early growth stages to promote root development.

We can extract this data from various agricultural sites, such as the Farmers Almanac. The information needs to be encoded into programmatic rules.

Risk protocol for dry soil

Knowing whether soil needs irrigation depends on a few factors. First, the crop determines the general watering schedule. Decisions are made based on reconciling watering needs with current environmental conditions. We need to know the current soil moisture level. We also want to know the future soil moisture level, which factors in weather forecasts. There's no point watering today if it will rain tomorrow.

Irrigation control

Automated irrigation control builds on top of the water needs forecasting model.

Freeze warnings

Farmers and gardeners in temperate or similar climates need to know when there are frosts that could damage fragile seedlings. Uncertainty leads to wasted time preparing for non-events or stress worrying about whether plants will be okay.

1. Install device in observation area
2. Enable freeze warning alerts

Hotter Times

This site is used to display public data related to the sensor network. Hotter Times thus becomes a public service where everyone can benefit from the device network.

All of the layers discussed above can be displayed on Hotter Times.

Mobile app

A mobile application is used to mediate the network connection for devices in remote areas. It is also used to deliver alerts to users. Using the device as an alert delivery mechanism ensures users have the device installed and able to act as a data transport for devices.

Physical Infrastructure

Devices can connect with the CADP via different means. Figure 6 shows two different approaches. The simplest approach is to connect directly via WiFi. In this approach, a device communicates directly with the CADP MQTT server. This approach works well when reliable WiFi is available.

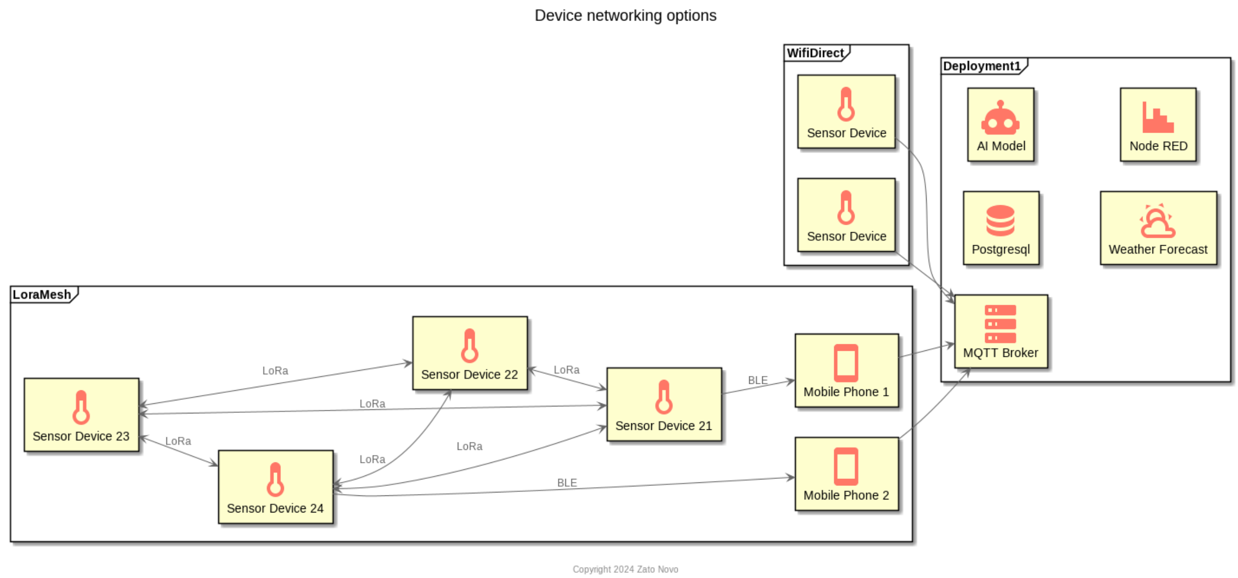


Figure 6: Different network configuration of system

The method used for remote areas with limited infrastructure utilizes a mobile phone as an intermediary. In this approach we assume that mobile phones (with a network connection) exist in proximity to devices but they are no fixed internet gateways. We call the mobile phones ephemeral gateways because they may disappear at any time.

Ephemeral gateways offer a temporary network connection. The EG is responsible for pushing data to the CADP MQTT server. If a network connection is unavailable, it will cache the data on the mobile phone until a network connection is available. Once a connection is established, the EG must push the data to the server.

Devices are connected together via the LoRa wireless protocol. When a particular device connects with an ephemeral gateway, it broadcasts a message to other connected devices and notifies them that it has a network connection. This device will attempt to upload as much data to the mobile device as possible.

Devices that receive this message will begin broadcasting data to be received by the device connected to the EG. The broadcast includes the ID of the device connected to the EG. Only the device connected to the EG can broadcast an ack back.

In the event that multiple EGs are available, each device is responsible for choosing the device to use as the target EG.

Device Operation

1. Install device in activity area

2. Pair device via Bluetooth
3. Set location for device

CADP Architecture

The architecture of the CADP comprises:

- a low power and long range sensor devices;
- a data collection and storage facility;
- a pluggable predictive models for producing forecasts and predictions;
- a reporting dashboard for decisionmakers and policymakers; and
- a delivery mechanism to provide alerts and suggested interventions to individual stakeholders.

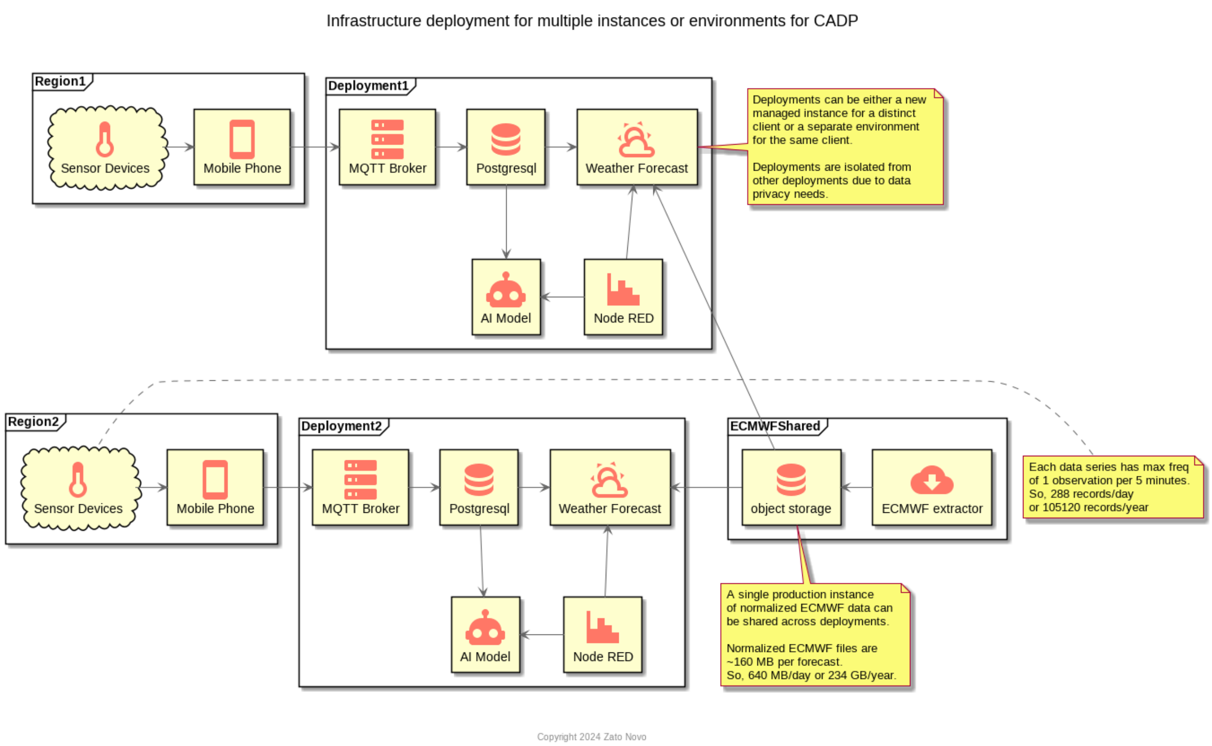


Figure 7: System architecture to generate weather forecasts and monitor devices.

The system has two primary sources of data. The first is baseline weather forecast data, produced by the European Center for Medium-range Weather Forecasts (ECMWF). This is a global, gridded dataset with 0.4 degree resolution. These forecasts are archived in Amazon S3 and are publicly accessible.

The second dataset comes from the devices themselves. The sensor data is used to tune the gridded forecast according to local conditions.

Design principles

- Plan for 18 month shelf life
- Focus on end-to-end infrastructure
- Stub components where necessary
- Use concurrency where appropriate to minimize wall time
- Minimize data movement
- Minimize cost

Database

Sensor data ingestion

Raw sensor data

Stored as an unnormalized delimited file. These files have no header, and each row is self-contained.

The general format of the file is `key,timestamp,values`, where `values` represents an arbitrary number of additional fields.

For example, the following snippet shows actual sensor data in this format.

```
air_t_h_p,2024-01-15T00:03:26,-5.34,85.30,1048.80
battery,2024-01-15T00:08:23,0.8398108
air_t_h_p,2024-01-15T00:13:20,-5.19,83.84,1049.22
air_t_h_p,2024-01-15T00:23:15,-5.11,82.51,1049.31
battery,2024-01-15T00:28:12,0.8361486
air_t_h_p,2024-01-15T00:33:09,-6.98,83.39,1049.73
air_t_h_p,2024-01-15T00:43:03,-6.97,82.98,1050.29
battery,2024-01-15T00:48:00,0.8359045
air_t_h_p,2024-01-15T00:52:57,-6.99,84.48,1050.52
air_t_h_p,2024-01-15T00:57:44,-5.05,84.28,1050.53
battery,2024-01-15T01:02:44,0.8312657
air_t_h_p,2024-01-15T01:02:45,-6.87,84.85,1050.91
air_t_h_p,2024-01-15T01:12:39,-5.05,84.56,1050.78
air_t_h_p,2024-01-15T01:22:33,-5.06,84.41,1051.12
battery,2024-01-15T01:22:34,0.8307775
air_t_h_p,2024-01-15T01:32:28,-6.99,83.29,1051.41
air_t_h_p,2024-01-15T01:42:22,-6.87,82.33,1051.89
battery,2024-01-15T01:42:22,0.8363928
air_t_h_p,2024-01-15T01:52:16,-6.61,80.60,1052.68
air_t_h_p,2024-01-15T02:01:55,-6.27,80.06,1053.42
battery,2024-01-15T02:01:56,0.831754
air_t_h_p,2024-01-15T02:11:50,-6.34,77.78,1053.46
air_t_h_p,2024-01-15T02:21:44,-6.39,76.10,1053.47
battery,2024-01-15T02:21:45,0.8322423
air_t_h_p,2024-01-15T02:31:39,-6.17,73.03,1053.98
air_t_h_p,2024-01-15T02:41:33,-6.08,71.05,1054.36
battery,2024-01-15T02:41:33,0.8361486
```

Two observation types are included. The first is `air_t_h_p`, which represents air observations. The key hints that there are three value fields. The second observation type is `battery` and contains a single reading for the battery level.

Data issues

When the device cannot connect to the Internet for a while, it loses it's internal time, and the clock resets to 2000-01-01. These data points can be dropped.

Even if the clock doesn't reset, there can be clock drift. This can usually be ignored since the observations are aggregated anyway.

Weather forecasting

The system uses the ECMWF global gridded forecast data as a baseline forecast. Local sensor data is integrated into this forecast to produce a more accurate forecast for specific locations.

ECMWF forecasts

ECMWF forecasts are updated every six hours. The publicly available AWS archive[1] has a complete history of data. Usage for the bash script that downloads the data appears in Appendix A.

There are up to 48 forecast steps per forecast. Each forecast step ranges between 27 MB - 50 MB, which implies a minimum of 1.3 GB of data is retrieved every six hours (per update). To save space, these files need to be slimmed down by removing all unnecessary data.

Early warning system

Plugins

Forecasting plugins

Risk protocols

References

Vanos, J., Guzman-Echavarria, G., Baldwin, J.W. et al. A physiological approach for assessing human survivability and liveability to heat in a changing climate. Nat Commun 14, 7653 (2023). https://doi.org/10.1038/s41467-023-43121-5

https://www.nature.com/articles/s41598-021-98316-x

[1] https://registry.opendata.aws/ecmwf-forecasts/