

# Automatically Survey Frogs Using Raspberry Pis, Jetstream Cloud, and Machine Learning

Sheri Sanders, Eliza Foran, Emmanuel Guido, Jazzy Anderson, Thomas Slayton, Thomas G. Doak

## Introduction

Field stations increasingly leverage remote sensors for large scale environmental data collection. Here we demonstrate a proof-of-concept workflow of data collection from remote sensors to presentation of summary results on a remote - and therefore fast and stable - cloud server. Acoustic (frog call) and environmental data is collected via Raspberry Pis and the data is streamed to the server on XSEDE's Jetstream, housed in part at Indiana University, through low-bandwidth messaging. Alternatively, data from the Pis can be picked up via drive-by or drones if collection sites are outside of a network, and then uploaded to the cloud server. The Jetstream cloud server then does all the heavy lifting, exporting the data into a database, running automatically updating summary scripts to produce graphs, and hosting a Drupal-based website to present the data to collaborators or the public. An additional server image was produced to categorize acoustic data (frog calls) using three different machine learning algorithms. This publicly available image serves as starting template and documentation for a variety of cheap sensor projects that may be useful to field stations or citizen science projects, such as FrogWatch USA.

## Remote Data Collection

Data collection is driven by sensors and microphones linked to Raspberry Pis. Text based data (temperature, humidity) is pushed to Jetstream using small messages, which is forgiving of limited bandwidth and connection. Larger data types need more bandwidth/pick-up

## Data Analysis

We are developing support images that can be used with different types of data (i.e. acoustic), allowing for analysis and presentation of data to the website. For example, we have used Convolutional Neural Networks that can identify four frog species accurately 97.35% of the time.

## Remote Pick-up

We have now enabled data pick up via a designated raspberry pi, which can be used in a drone fly-by or drive-by to gather data from multiple sensors (great if you have a limited network on site).

## Citizen Science

The same set up can be used to aggregate data from citizen science projects at diverse locations. The data is sent to Jetstream, tagged, and added to the full data database. We are currently working on automating the full FrogWatch USA collection protocol!

## Focal Animal Studies

We have used this set up to collect daily data on individual animals, adding an additional dimension to "nest cams". This set up allows researchers to monitor bee hive temperatures or nest choice/success factors

## Technical Details

The technical details of setting up the Raspberry Pis as well as the Jetstream image are included in the documentation and on the image itself.

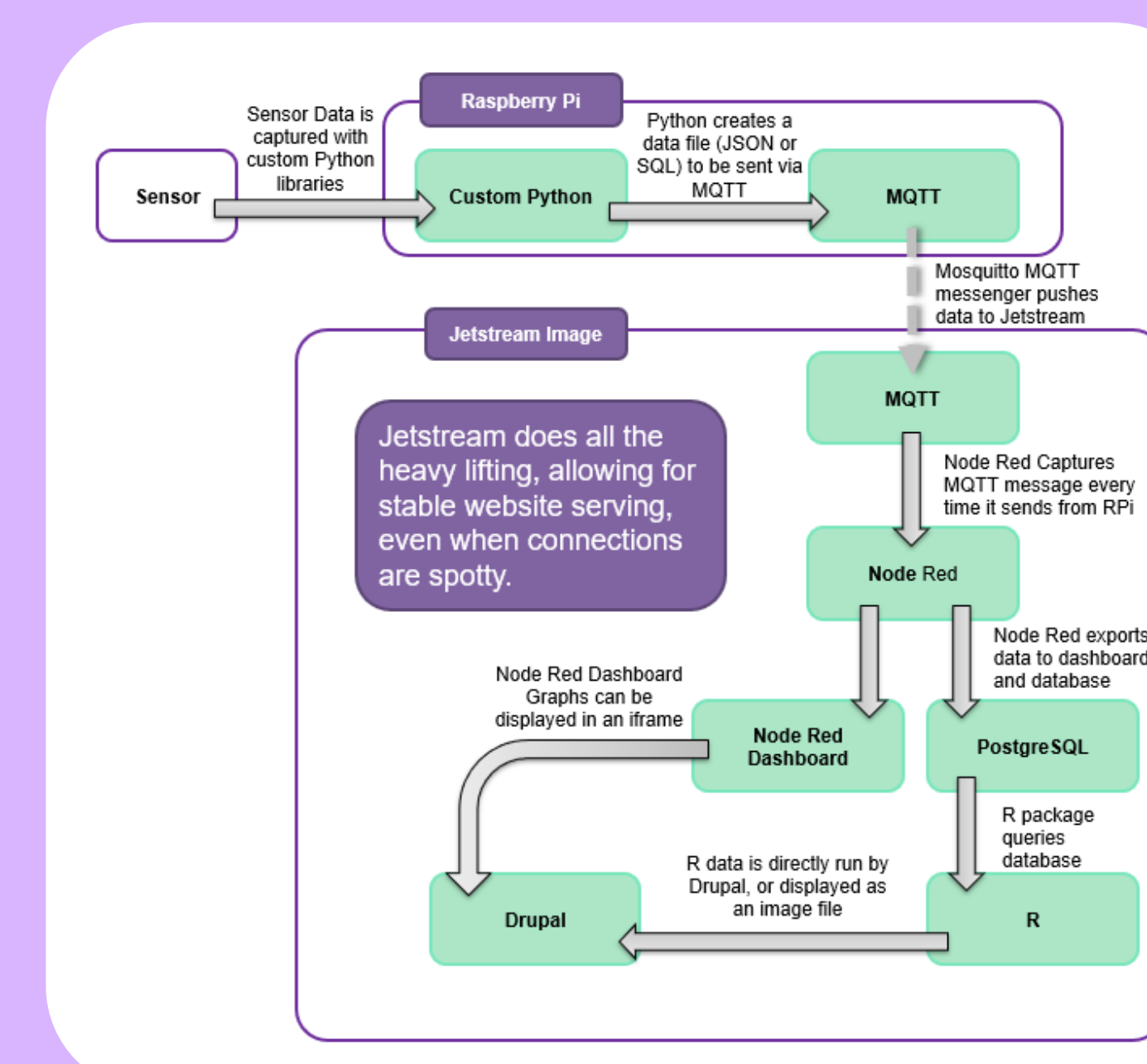
This setup was initially developed during a Jetstream REU project in the Summer of 2018 at Indiana University in conjunction with NCGAS, and further developed by 2018-2019 Center of Excellence for Women in IT (CEWIT) REUs mentored by NCGAS.

## NEW Machine-Learning Based Analysis of Frog Calls

Jetstream REUs built an analysis image for bioacoustics using frog call data, but this image could be adapted to other audio or image data as well.

The students tested three types of analysis, spectrograph images with a Convolutional Neural Network (CNN), direct audio analysis with CNN, and direct audio analysis with Recurrent Neural Network (RNN). The networks were tested for accuracy on Spring Peepers versus Chorus Frogs (Data from Cornell's Macaulay Library archives of wildlife sound), and then again with additional data from Green Frogs and American Toads.

The CNN with images was the most resilient to more species, and we will be continuing to improve and optimize this networks, however all three options will be available on github, as tutorials, and on Jetstream images.



Proportion of data selected for model training versus prediction

	CNN-image	CNN-audio	RNN-audio
<b>2 species</b>	Training: 1244 (86.15%) Testing: 200 (13.85%)	Training: 1015 (83.54%) Testing: 200 (16.46%)	Training: 1015 (83.54%) Testing: 200 (16.46%)
<b>4 species</b>	Training: 2428 (85.86%) Testing: 400 (14.14%)	Training: 1911 (82.69%) Testing: 400 (17.31%)	Training: 1911 (82.69%) Testing: 400 (17.31%)

Prediction accuracy of the three models (avg ± std dev %) after 10 runs

	CNN-image (avg ± std dev. %)	CNN-audio (avg ± std dev. %)	RNN-audio (avg ± std dev. %)
<b>2 species</b>	97.72 ± 0.80	99.50 ± 0.00	99.45 ± 0.16
<b>4 species</b>	97.35 ± 0.38	88.88 ± 0.89	89.83 ± 0.69

## Visualization for Collaboration or Public

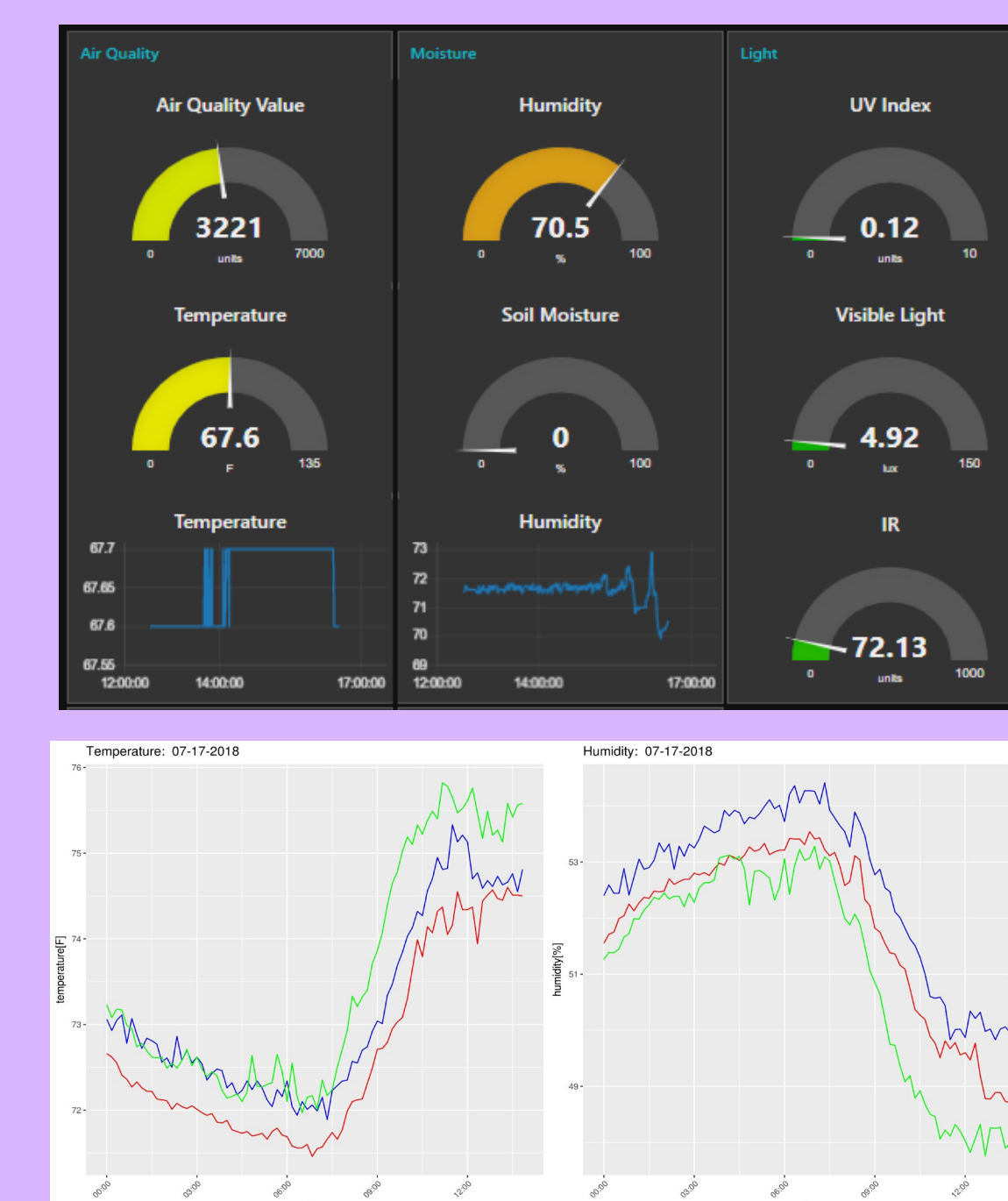
We developed two default options for presenting data on a pre-built Drupal website. The first is using Node Red to take a flat file (JSON format) and produce real time data and graphs without the need for a database (though it can be saved as one). Alternatively, data can be delivered directly to a database on the Jetstream instance (a more secure option). Data can then be queried using R or other scripts and results presented on the site.

Drupal allows for page-by-page control of access, allowing for publicly viewable sections as well as collaborator sections—controlled by a simple point and click interface.

## Demo Data

We implemented both options on our demo site – the first being a single Raspberry Pi with seven sensors attached, monitoring an animal habitat. This set up uses the direct stream option and Node Red to produce the graphs (top).

Additionally, we used three separate Raspberry Pis to stream single sensor data to one database, and used R to produce daily graphs (bottom).



## Getting Started using NCGAS tools

Get started on [ncgas.org](http://ncgas.org) – we have tutorials on starting to use Jetstream, how to gain quick access to our field station allocations, and demos of all of our tools. We also offer support for workshops, consulting, and software set up! Feel free to ask us for demos of our webtools!

This project was funded by NSF grant ACI-1445604 supporting the Jetstream Research Experience for Undergraduate Program and the NSF grant ABI-1062432 supporting the National Center for Genome Analysis Support.