## Research Notes for the Developer(s)

## Scraping and Combining Public Data

The raw inputs of this project are a) one or more JSON-format public feeds; and b) publicly available camera images that specifically show a pictorial representation of the data. Both items are specific to “The Bureau of Meteorology”. These items are updated regularly (every 30 minutes for the weather observations) and are eminently suitable for this project without filtering or processing. Below are some URLs to investigate:

* [BOM Weather observations for Toowoomba Airport](http://www.bom.gov.au/fwo/IDQ60901/IDQ60901.95551.json)
* [Toowoomba Airport Weather Camera point North-East](https://weathercams.airservicesaustralia.com/wp-content/uploads/airports/041529/041529_045.jpg)

The home page for the Air Services Australia web cameras appears at:

[Airservices Australia – Live Weather Cameras](https://weathercams.airservicesaustralia.com/)

Other weather data sources may include Wundergroud:

[Personal Weather Station Dashboard | Weather Underground (wunderground.com)](https://www.wunderground.com/dashboard/pws/ITOOWOOM10)

Another example of publicly available web cameras may be found here:

[Weather Cameras (airportweathercams.com)](https://airportweathercams.com/YCFN/)

## Python Libraries

This project will rely on Python HTTP **requests** so the **requests** library needs to be installed from PIP or another source. Libraries include:

* Requests (as noted above)
* Date Time.

As Splunk has multiple Python libraries, one project outcome is to learn to use a library to communicate with the Splunk instance rather than attempting to craft a HTTP GET or POST request from scratch. This would be necessary with JavaScript, for example. Python has these abstractions so they should be used.

Below is a link to a tutorial on using the Splunk HTTP Event collector:

[Set up and use HTTP Event Collector in Splunk Web - Splunk Documentation](https://docs.splunk.com/Documentation/Splunk/9.1.2/Data/UsetheHTTPEventCollector)

## Coding Questions

1. Assuming more than one data source is going to be queried, how should the details about the data sources be stored?
2. What other values in the code should be configuration items that appear by themselves and how should this data be stored?
3. How often should each data source be queried?
4. If a problem occurs and the data can’t be retrieved or is corrupted, what should happen?