## Step 2: Data Collection and Preparation [Week 3-5]

Data is instrumental for the success of an AI project. Consider the following:

* **Data Sources**
  + <https://datasetsearch.research.google.com/>
  + <https://science.gc.ca/site/science/en/open-science/datasets-and-portals#wb-auto-4>
  + <https://search.open.canada.ca/opendata/>
  + <https://data.gov/>
  + <https://mira.mcmaster.ca/resources/data-repos-datasets/>
  + <https://www.tableau.com/learn/articles/free-public-data-sets>
  + <https://www.kaggle.com/datasets>
  + <https://data.worldbank.org/>
  + Assess the accuracy and reliability of the sources
* **Collection Method**
  + We will be gathering datasets from websites listed in the above section.
* **Data Attributes**
  + Document the specific features or attributes you need for your AI model.
* **Data Cleaning**
  + Explain how you will handle missing values, outliers, and inconsistencies.
  + Detail how you will split the data into training, validation, and testing sets.
  + Identify what other tasks you need to do for the dataset.