**Data Management Plan**

****

**Team: TP31**

**Table of Contents**

[**1.0 Data Sets Overview 3**](#_lccraoumqmos)

[1.1 Open Data Sources 3](#_m75nudcixtu2)

[**2.0 Iteration 1 4**](#_5a2n9ltq2ng4)

[2.1 Data Sources Used 4](#_igbqmlagyigb)

[2.2 Data Usage 4](#_c7kp56pujup0)

[2.3 Data Preparation 5](#_a0ksq2hvdj95)

[2.4 Data Storage 5](#_im2uyuoggole)

[2.5 Data Design 5](#_yimmaft6g8ca)

[2.6 Data Analytics 6](#_8cl8boufxm26)

[2.7 Ethical, Legal, and Privacy Issues 7](#_p3khryghq8nu)

[**3.0 Iteration 2 8**](#_kgwtl41umw1w)

[3.1 Data Sources Used 8](#_b072co8ndp54)

[3.2 Data Usage 9](#_3xbmlc28qifd)

[3.3 Data Preparation 9](#_3ym0pry5ic5u)

[3.4 Data Storage 11](#_9b3ifng0ss9p)

[3.5 Data Design 11](#_9i48dzr1jk8z)

[3.6 Data Analytics 13](#_8csupigwy603)

[3.7 Data Modelling 14](#_nptgdk32rjza)

[3.8 Ethical, Legal, and Privacy Issues 14](#_i570v9fqih02)

[**4.0 Iteration 3 15**](#_kwfrq9f8lskq)

[4.1 Data Sources Used 15](#_54xdcbsd1xxp)

[4.2 Data Usage 16](#_orsh5f91ecgc)

4.3 Data [Preparation](#_3ym0pry5ic5u) 16

[4.4 Data Storage 17](#_ybcahi5j9kew)

[4.5 Data Design 18](#_ce80pb9nj6q4)

[4.6 Data Analytics 18](#_j30rwft6mljk)

[4.7 Ethical, Legal, and Privacy Issues 19](#_grqq28bbhaca)

# 1.0 Data Sets Overview

## 1.1 Open Data Sources

| **ID** | **Source** | **Physical Format Used** | **Update Frequency at Source** | **Granularity** | **Copyright Details** | **Last Updated** | **Usage** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | [Stormwater priority area](https://discover.data.vic.gov.au/dataset/hws2018-stormwater-priority-areas) | Spatial | Monthly | High - Catchment Level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 30/08/2024 | Iteration 1 |
| 2 | [Waste collected by month](https://data.melbourne.vic.gov.au/explore/dataset/waste-collected-per-month/table/) | CSV | Yearly | High - Waste type | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 14/12/2023 | Iteration 3 |
| 3 | [Water flow routes](https://data.melbourne.vic.gov.au/explore/dataset/water-flow-routes-over-land-urban-forest/information) | Spatial | Yearly | High - Road level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 14/2/2024 | Iteration 2 |
| 4 | [Bin fill strength](https://data.melbourne.vic.gov.au/explore/dataset/smart-bins-argyle-square/table/?disjunctive.serial&sort=time) | CSV | NA | High - Bin level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 4/06/2022 | Iteration 3 |
| 5 | [Stormwater pits](https://data.melbourne.vic.gov.au/explore/dataset/stormwater-pits/table/?location=16,-37.79022,144.92126&basemap=mbs-7a7333) | CSV | Yearly | High - Pit level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 26/03/2023 | Iteration 2 |
| 6 | [Waste projection model](https://discover.data.vic.gov.au/dataset/victorias-waste-projection-model) | CSV | Quarterly | High - Suburb Level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 13/06/2024 | Iteration 2 |
| 7 | [Air Temperature Victoria](https://discover.data.vic.gov.au/dataset/air-temperature-observations) | CSV / GeoJSON | Quarterly | High - Water station level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 26/06/2024 | Iteration 2 |
| 8 | [Water Gauging Stations Victoria](https://discover.data.vic.gov.au/dataset/streamflow-water-gauging) | CSV | Quarterly | High - Water station level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 13/06/2024 | Iteration 2 |
| 9 | [Pavement Condition Data](https://discover.data.vic.gov.au/dataset/pavement-condition-data) | CSV / GeoJSON | NA | High - pavement level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 18/08/2021 | Iteration 2 |

# 2.0 Iteration 1

## 2.1 Data Sources Used

| **ID** | **Source** | **Physical Format Used** | **Update Frequency at Source** | **Granularity** | **Copyright Details** | **Last Updated** | **Usage** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | [Stormwater priority area](https://discover.data.vic.gov.au/dataset/hws2018-stormwater-priority-areas) | Spatial | Monthly | High - Catchment Level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 30/08/2024 | US 1.3 |

## 2.2 Data Usage

* We used the Priority Stormwater dataset from Victoria to **create a map** that highlighted areas in the state **prone to higher pollution** due to factors like industrial activity, urban density, and rainfall patterns.
* This data was instrumental in developing a heat map that visually represented these priority pollution zones, helping us identify regions needing more focused stormwater management.
* Key indicators, such as infiltration rates and annual water harvesting, were incorporated to measure the effectiveness of existing stormwater practices.
* Additionally, we analysed catchment and sub catchment areas, highlighting them on the map to increase user awareness of nearby creeks and waterways.
* This approach will help users understand their proximity to potential pollution sources and encourage them to adopt practices to reduce stormwater pollution in their surroundings.

## 2.3 Data Preparation

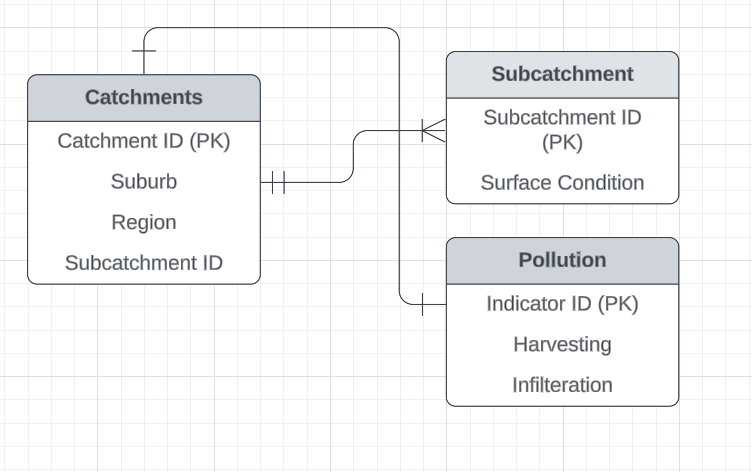
**Data Wrangling:**

* We cleaned the dataset by removing any **rows with null values**, particularly in catchment areas, subcatchments, and pollution indicators, to maintain data accuracy for mapping and analysis.
* We reviewed and corrected data types, ensuring that water infiltration rates were properly formatted as numerical data.
* We conducted a logic check to ensure that the relationships between pollution levels and proximity to industrial zones were consistent. For example, we confirmed that areas with higher industrial activity and urban density logically corresponded with higher pollution indicators.

## 2.4 Data Storage

* The database was set up in **Oracle cloud**.
* The cleaned and validated dataset was first uploaded to Oracle Cloud using Oracle Cloud Infrastructure (OCI) Data Transfer.
* Within the Oracle Autonomous Database, we defined a relational table structure that aligned with the dataset’s schema. This included creating tables for key entities such as Catchments, Subcatchments, and Pollution\_Indicators.
* Primary keys were established for each table, such as catchment\_id for the Catchments table and subcatchment\_id for the Subcatchments table.
* The cleaned dataset was then ingested into the corresponding tables within the Oracle Autonomous Database.

## 2.5 Data Design



## 

## 

## 

## 

## 

## 2.6 Data Analytics

#### **Hindsight:**

#### We analysed historical data from the Priority Stormwater dataset to understand past trends in pollution levels across various regions in Victoria.

#### The analysis revealed specific catchment and subcatchment areas that have been significantly affected by industrial runoff and urban stormwater.

#### We assessed the effectiveness of past stormwater management practices, such as infiltration and water harvesting, by comparing pollution levels before and after the implementation of these measures.

#### **Insight:**

#### Using the most recent data, we created a heat map highlighting current pollution hotspots. These areas, identified by high levels of pollutants and poor water quality indicators, are now the focus for immediate stormwater management efforts.

* The analysis provided insight into the current impact of various factors on stormwater pollution.
* By mapping catchment and sub catchment areas, we made users aware of their proximity to polluted waterways. This insight is crucial for community engagement, as it encourages residents to adopt practices that reduce stormwater pollution, such as proper waste disposal and rainwater harvesting.

#### **Foresight:**

#### The foresight gained from the analysis informs strategic planning for stormwater management. For instance, regions projected to experience increased industrial growth can be prioritised for the implementation of advanced stormwater filtration systems and stricter pollution controls.

## 2.7 Ethical, Legal, and Privacy Issues

### **Ethical Issues**

* **Data Accuracy:** Ensuring that all data used is accurate and reliable.
* **Transparency:** Being open about where data comes from and how it’s used.
* **Environmental Impact:** Considering the broader effects of water management solutions on the environment.

### **Legal Issues**

* **Data Protection Laws:** Complying with regulations like GDPR and local privacy laws.
* **Intellectual Property:** Using data and technology legally, with proper permissions.
* **Liability:** Defining who is responsible if there are issues with the data or the website.

### **Privacy Issues**

* **Data Confidentiality:** Keeping user data secure and private.
* **Consent:** Making sure users know what data is collected and how it’s used, and getting their permission.
* **Data Retention:** Managing how long data is kept and ensuring it’s properly disposed of when no longer needed.
* **Third-Party Sharing:** Carefully handling any data shared with other organisations and ensuring they also protect user privacy.

# 3.0 Iteration 2

## 3.1 Data Sources Used

| **ID** | **Source** | **Physical Format Used** | **Update Frequency at Source** | **Granularity** | **Copyright Details** | **Last Updated** | **Usage** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | [Water flow routes](https://data.melbourne.vic.gov.au/explore/dataset/water-flow-routes-over-land-urban-forest/information) | Spatial | Yearly | High - Road level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 14/2/2024 | Epic 3 & 4 |
| 2 | [Stormwater pits](https://data.melbourne.vic.gov.au/explore/dataset/stormwater-pits/table/?location=16,-37.79022,144.92126&basemap=mbs-7a7333) | CSV | Yearly | High - Pit level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 26/03/2023 | Epic 3 & 4 |
| 3 | [Waste projection model](https://discover.data.vic.gov.au/dataset/victorias-waste-projection-model) | CSV | Quarterly | Medium - Suburb Level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 13/06/2024 | Epic 3 & 4 |
| 4 | [Air Temperature Victoria](https://discover.data.vic.gov.au/dataset/air-temperature-observations) | CSV / GeoJSON | Quarterly | High - Water station level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 26/06/2024 | Epic 3 & 4 |
| 5 | [Water Gauging Stations Victoria](https://discover.data.vic.gov.au/dataset/streamflow-water-gauging) | CSV | Quarterly | High - Water station level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 13/06/2024 | Epic 3 & 4 |
| 7 | [Pavement Condition Data](https://discover.data.vic.gov.au/dataset/pavement-condition-data) | CSV / GeoJSON | NA | High - pavement level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 18/08/2021 | Epic 3 & 4 |

## 3.2 Data Usage

* **Water Flow Routes:** To predict and understand how stormwater runoff travels through Victorian suburbs, using key geographical points and flow connections between nodes.
* **Stormwater Pits:** Analyze the placement and construction of stormwater pits to model their effectiveness in managing runoff, particularly focusing on materials and geographic locations.
* **Waste Projection Model:** Evaluate how different waste streams contribute to stormwater pollution, with a focus on waste recovery rates and waste types that impact runoff quality.
* **Air Temperature Victoria:** Use air temperature to estimate evaporation rates, which affect the stormwater volume and flow predictions across various locations.
* **Water Gauging Stations Victoria:** Monitor water levels and overflow data from key gauging stations to identify potential high-risk areas for stormwater flooding.
* **Pavement Condition Data:** Assess the impermeability of surfaces, such as parking lots, to estimate runoff levels based on surface conditions like roughness, texture, and cracking.

## 3.3 Data Preparation

1. Each data table was cleaned to remove invalid and null values to ensure that the data populates accurately.
2. All data columns in every table was not needed for this project and hence, only the needed columns were retained. Please see the below table for an insight into the columns that were eliminated.

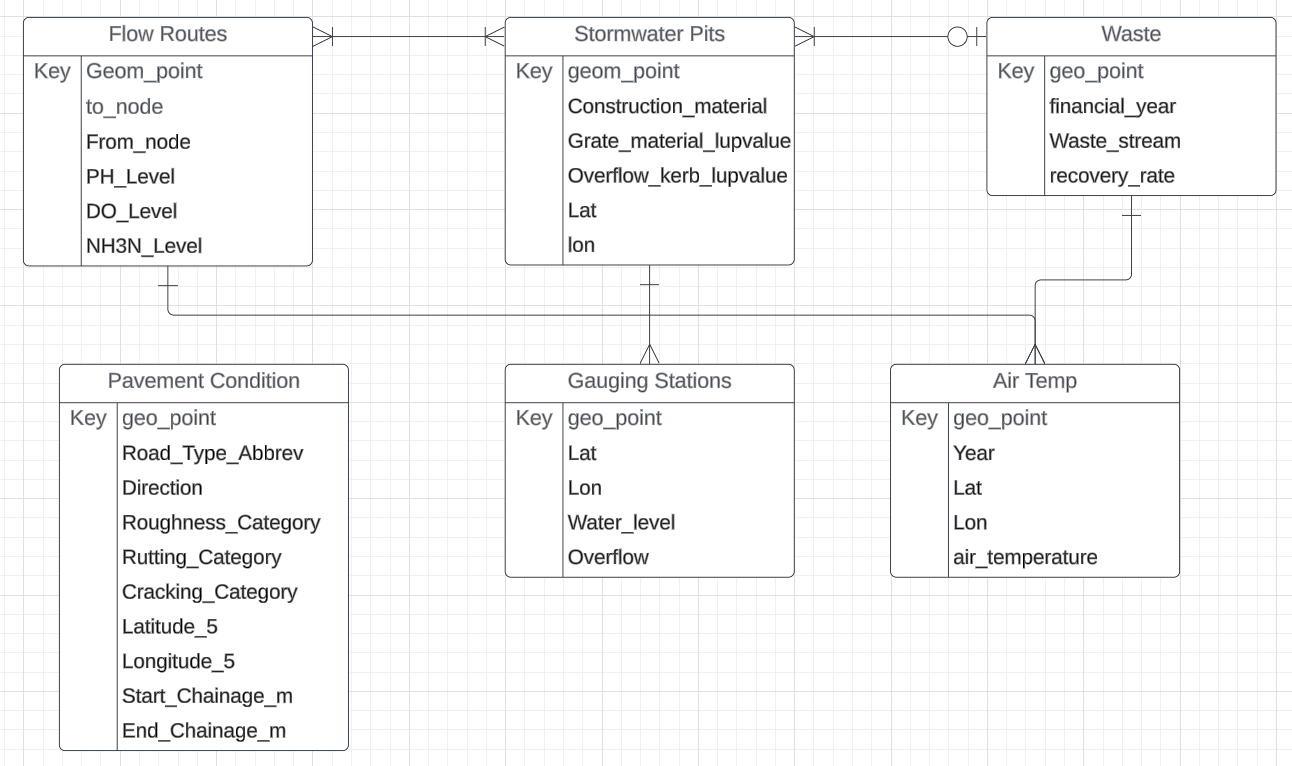
| **Dataset** | **Intended Use** | **Initial Columns** | **Columns Retained** |
| --- | --- | --- | --- |
| [Water flow routes](https://data.melbourne.vic.gov.au/explore/dataset/water-flow-routes-over-land-urban-forest/information) | This data is crucial to understand and predict the runoff routes through Victorian suburbs. | Geom\_point  Geo\_shape  Grid\_code  Source  To\_node  From\_node  PH Level  DO Level  NH3N Level | Geom\_point  To\_node  From\_node  PH Level  DO Level  NH3N Level |
| [Stormwater pits](https://data.melbourne.vic.gov.au/explore/dataset/stormwater-pits/table/?location=16,-37.79022,144.92126&basemap=mbs-7a7333) | Critical to analyse placement of the pits in the model | Asset\_number  Asset\_desc  Construction\_material  Grate\_length  Grate\_material\_lupvalue  Grate\_width  Model\_descr\_lupvalue  Model\_no\_lupvalue  Object\_type\_lupvalue  Overflow\_kerb\_lupvalue  Lat  lon | Asset\_number  Construction\_material  Grate\_material\_lupvalue  Overflow\_kerb\_lupvalue  Lat  lon |
| [Waste projection model](https://discover.data.vic.gov.au/dataset/victorias-waste-projection-model) | Waste generation and its impact on the runoff causing polluted stormwater | financial\_year  wpm\_material\_type  wpm\_material\_name  waste\_stream  recycled  waste\_to\_energy  export\_interstate  export\_international  disposal  recovery\_rate | financial\_year  Waste\_stream  recovery\_rate |
| [Air Temperature Victoria](https://discover.data.vic.gov.au/dataset/air-temperature-observations) | Air temperatures help understand the evaporation rates for the model | date\_time  location\_description  Lat  Lon  air\_temperature | Year  Lat  Lon  air\_temperature |
| [Water Gauging Stations Victoria](https://discover.data.vic.gov.au/dataset/streamflow-water-gauging) | Gauging stations are important to understand the overflow levels | Year  Lat  Lon  Description  Water level  Overflow  geo\_point | Year  Lat  Lon  Water level  Overflow |
| [Pavement Condition Data](https://discover.data.vic.gov.au/dataset/pavement-condition-data) | Parking lots and other surfaces around water stations are key to runoff levels. The more impermeable the surface, the more runoff. | OBJECTID  Classified\_Road\_Number  Name\_Part  Road\_Type\_Abbrev  Direction  Region  RMA\_Class  Federal\_Road\_Network  Date  Roughness\_Category  HATI\_Category  Texture\_\_LWP\_\_Category  Texture\_\_BWP\_\_Category  Rutting\_Category  Cracking\_Category  Latitude\_5  Longitude\_5  Start\_Chainage\_m  End\_Chainage\_m | Name\_Part  Road\_Type\_Abbrev  Direction  Region  Roughness\_Category  HATI\_Category  Texture\_\_LWP\_\_Category  Texture\_\_BWP\_\_Category  Rutting\_Category  Cracking\_Category  Latitude\_5  Longitude\_5  Start\_Chainage\_m  End\_Chainage\_m |

## 3.4 Data Storage

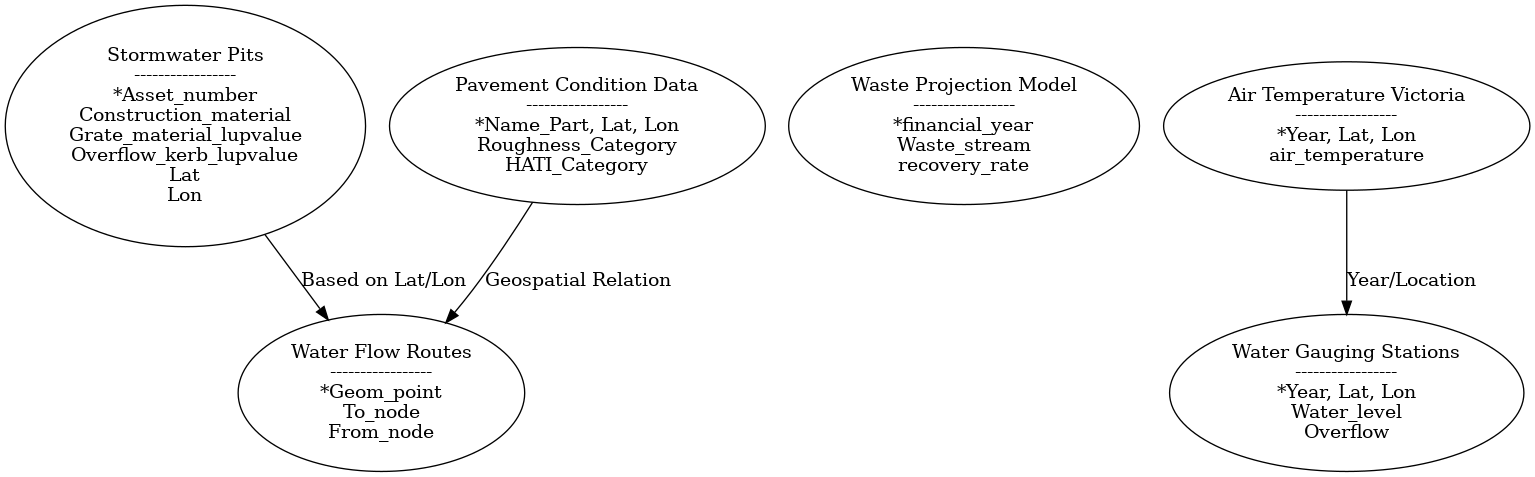
* Data is stored on an **AWS server** for scalable and secure management, with datasets like **Water Flow Routes** and **Stormwater Pits** stored in **S3 buckets** for raw files and **AWS RDS** for structured data.
* A **MySQL database** is used for structured storage and fast querying, ensuring efficient access to datasets such as **pits placement**, **waste stream recovery rates**, and **water flow routes**.
* The database was created by defining a schema that maps each dataset to tables with relevant columns. Primary keys and indices are set for location-based queries like **lat/lon** data.
* Datasets were ingested into the MySQL database from CSV files (in AWS) or via APIs.
* The **website** is connected to the database via a **secure API**, allowing real-time data retrieval for model training.

## 3.5 Data Design

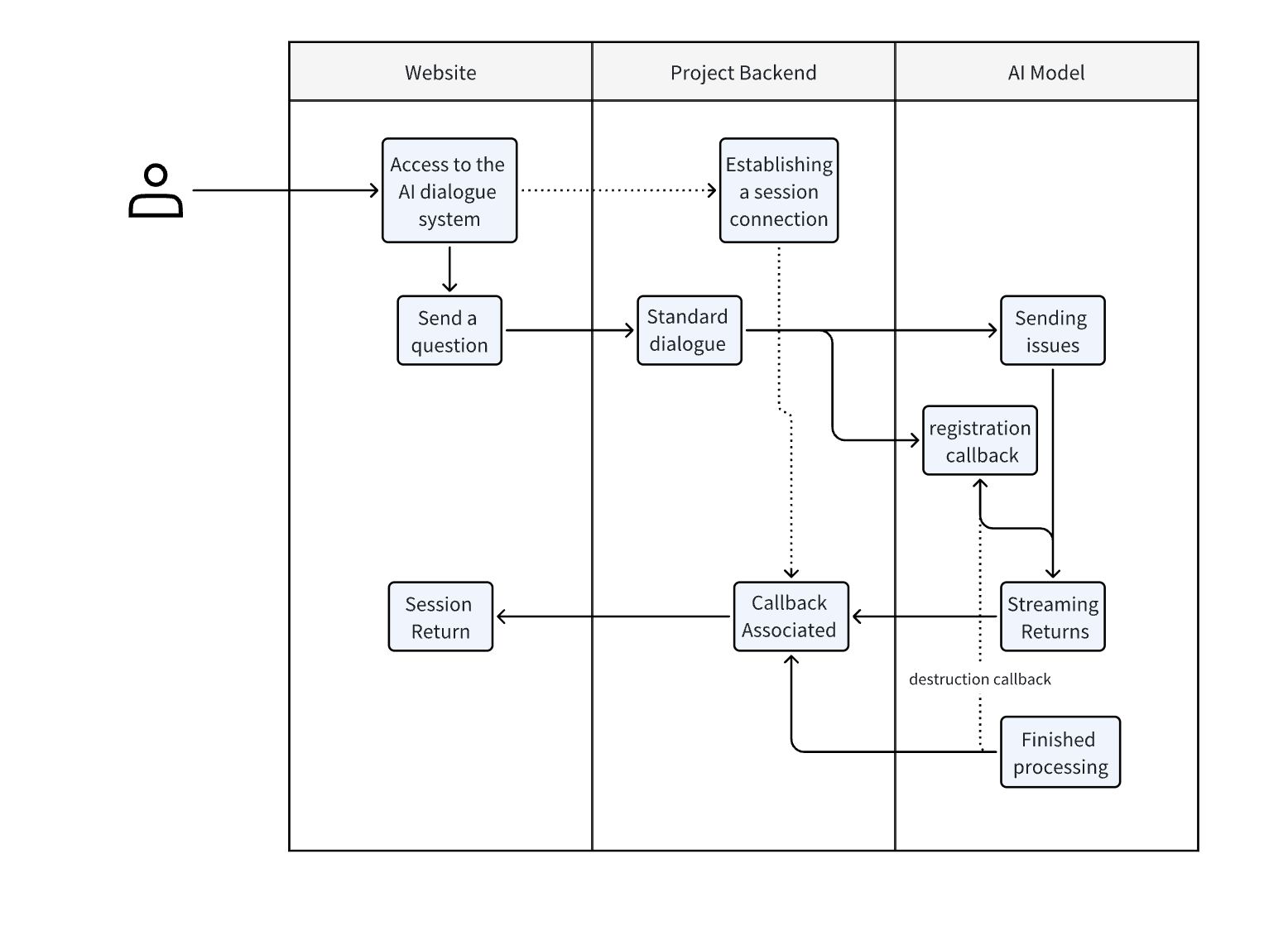
1. ERD Model



1. Model Diagram



1. Chat Bot Integration



## 3.6 Data Analytics

### **Hindsight**

Hindsight focuses on understanding past data to learn from previous patterns and events. Using historical datasets in your project, such as:

* **Air Temperature Victoria** and **Water Gauging Stations**: Analyzing past water levels and temperature patterns helps identify seasonal trends, historical flood risks, and the impact of temperature on water evaporation.
* **Waste Projection Model**: Reviewing historical waste generation can show how waste disposal has affected stormwater quality over the years.

### **Insight**

Insight derives meaning from the current data, offering explanations for the present situation:

* **Stormwater Pits** and **Water Flow Routes**: Insights into the current placement of stormwater pits and flow routes can highlight existing vulnerabilities in stormwater management, such as areas prone to overflow or insufficient drainage infrastructure.
* **Pavement Condition Data**: The current state of parking lots and roads allows for real-time understanding of surface runoff issues, as impermeable surfaces may lead to higher runoff, contributing to water pollution.

### **Foresight**

Foresight involves using data to predict future outcomes and make informed decisions:

* **Water Flow Routes** and **Waste Projection Model**: Using current and past data on water flow and waste patterns, the model can predict future flood risks, pollution hotspots, and areas requiring infrastructure upgrades.
* **Air Temperature Victoria**: By forecasting future air temperature patterns, the model can estimate the evaporation rates and water availability, helping design effective stormwater management strategies for different seasons.

## 3.7 Data Modelling

Please visit the [Data Modelling Document](https://docs.google.com/document/d/1jChdjN13U1Tss6q0RrTkgnD3O6atbXnzr03FLGVSug8/edit) to learn the details of the modelling used in this project.

## 3.8 Ethical, Legal, and Privacy Issues

**Ethical Considerations:**

* Ensured the accuracy and reliability of data used to avoid misleading conclusions and potential negative impacts on community planning and infrastructure decisions.
* Made responsible use of environmental models and data to promote sustainable practices and avoid any practices that could unfairly penalise communities.
* Applied the data to support effective environmental and urban planning, avoiding biases that could affect certain neighbourhoods or groups unfairly.

**Legal Considerations:**

* Complied with the Creative Commons Attribution 4.0 International licence for all datasets, providing proper attribution.
* Followed relevant regulations and standards for environmental data use and reporting, ensuring that all legal requirements for data usage were met.
* Managed data in compliance with data protection laws and regulations applicable to the project’s jurisdiction.

**Privacy Considerations:**

* Ensured that no personally identifiable information (PII) was included in the datasets or inadvertently revealed through data integration and visualisation.
* Implemented security measures for data storage and access on AWS and MySQL, including encryption and access controls, to protect against unauthorised access and data breaches.
* Practised data minimization by retaining only the data necessary for the project, avoiding the collection or storage of unnecessary or sensitive information.

# 4.0 Iteration 3

## 4.1 Data Sources Used

| **ID** | **Source** | **Physical Format Used** | **Update Frequency at Source** | **Granularity** | **Copyright Details** | **Last Updated** | **Usage** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | [Stormwater priority area](https://discover.data.vic.gov.au/dataset/hws2018-stormwater-priority-areas) | Spatial | Monthly | High - Catchment Level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 30/08/2024 | US 1.3 |
| 2 | [Waste collected by month](https://data.melbourne.vic.gov.au/explore/dataset/waste-collected-per-month/table/) | CSV | Yearly | High - Waste type | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 14/12/2023 | Iteration 3 |
| 3 | [Bin fill strength](https://data.melbourne.vic.gov.au/explore/dataset/smart-bins-argyle-square/table/?disjunctive.serial&sort=time) | CSV | NA | High - Bin level | [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) | 4/06/2022 | Iteration 3 |

## 4.2 Data Usage

## **Catchment and Creek Data (Table 1)**

## **Purpose:** This data provides geographic locations of catchments and creeks, helping users identify nearby water bodies where they can swim.

## **Fields:** Catchment ID, Creek Name, Geo-coordinates (latitude, longitude), Area, Waterbody Type, Catchment Size, Nearby Locations.

## **Usage:** The geographic locations will be used to map water bodies in proximity to the user’s location.

## **Waste Dataset (Stormwater Pit Strength) (Table 2)**

## **Purpose:** This table contains data on stormwater pits and their capacity to manage waste, helping assess stormwater pollution risks.

## **Fields:** Pit ID, Catchment ID, Strength Rating (1-10), Waste Load, Pit Condition, Installation Date, Last Maintenance Date.

## **Usage:** This data will be analyzed to determine the likelihood of stormwater overflow, contributing to the pollution level calculations.

## **Rainfall and Temperature Data (API Data for the Day) (Table 3)**

## **Purpose:** This provides real-time environmental conditions, essential for assessing water safety.

## **Fields:** Date, Time, Rainfall (mm), Temperature (°C), Location (Geo-coordinates), Forecast Status.

## **Usage:** This data will be combined with stormwater pit data to predict pollution levels and suggest whether it is safe for swimming.

## **User Tips and Recommendations (Table 5)**

## **Purpose:** This table provides safety tips and pollution prevention advice based on the predicted pollution levels.

## **Fields:** Pollution Level (Good/Fair/Poor), Safety Tips, Pollution Prevention Measures, Health Risk Precautions, Local Recommendations.

## **Usage:** Depending on the pollution prediction, users will be presented with tailored tips for a safe swim and guidance on minimizing pollution.

## 

## 4.3 Data Preparation

## **Data Cleaning:**

## **Catchment and Creek Data:** Verify geo-location accuracy, remove any duplicates, and ensure correct mapping of catchments to creeks.

## **Waste Dataset (Stormwater Pit):** Check for missing maintenance data, and standardise waste load values to a common unit (grams).

## **Rainfall and Temperature Data:** Ensure the API fetches complete records for the required time intervals.

## **Data Transformation:**

## **Waste Dataset (Stormwater Pit):** Normalize stormwater pit strength to a scale (e.g., 1-10) for uniformity across regions. Prepare data for integration with rainfall data.

## **Data Integration:**

## **Waste Dataset + Rainfall and Temperature Data:** Link stormwater pit data with rainfall to create a pollution risk model. Establish thresholds for pit strength, rainfall, and temperature to calculate predicted pollution levels.

## **Decision and Recommendation System:**

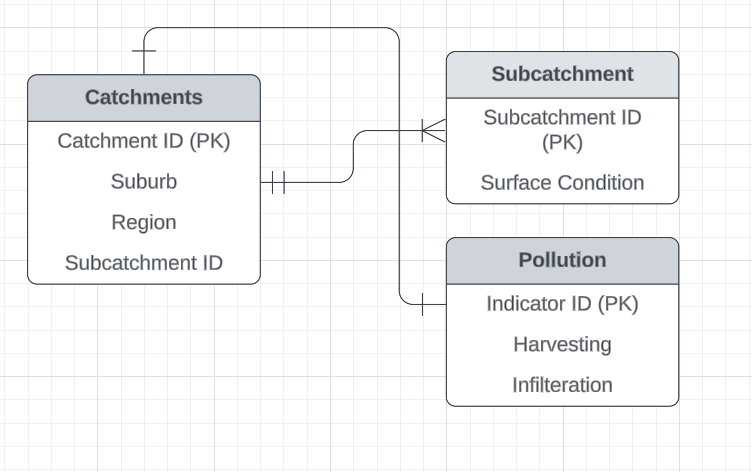
## **Pollution Prediction:** Use historical data and safe condition research (from accredited research information) to predict pollution levels based on integrated datasets (stormwater, rainfall, temperature).

## **User Tips and Recommendations:** Link pollution levels to corresponding tips and suggestions in the tips table.

## 4.4 Data Storage

* Data is stored on an **AWS server** for scalable and secure management, with datasets like **Water Flow Routes** and **Stormwater Pits** stored in **S3 buckets** for raw files and **AWS RDS** for structured data.
* A **MySQL database** is used for structured storage and fast querying, ensuring efficient access to datasets such as **pits placement**, **waste stream recovery rates**, and **water flow routes**.
* The database was created by defining a schema that maps each dataset to tables with relevant columns. Primary keys and indices are set for location-based queries like **lat/lon** data.
* Datasets were ingested into the MySQL database from CSV files (in AWS) or via APIs.
* The **website** is connected to the database via a **secure API**, allowing real-time data retrieval for model training.

## 4.5 Data Design



## 4.6 Data Analytics

#### **Hindsight**

#### **Historical Swim Safety Data**: By analyzing past decisions made by the tool, we can identify how often and in what conditions waterbodies were considered safe or unsafe for swimming. This includes tracking pollution levels, stormwater overflow, and user interactions during different weather conditions.

* **Stormwater Pit Performance History:** Review the past effectiveness of stormwater pits in managing waste and reducing runoff pollution. Track instances of pit failures or overflow that led to poor water quality.
* **Weather-Related Water Quality Changes:** Analyze how past rainfall and temperature affected pollution levels. This includes mapping periods of heavy rainfall or high temperatures to increases in pollution.
* **User Engagement Trends:** Evaluate user behavior in terms of their swimming choices and how closely they followed the tool’s recommendations. Were warnings heeded, or were there patterns of users swimming despite "Poor" water conditions?

#### **2. Insight**

* **Pollution and Weather Correlation:** Insight into how rainfall, temperature, and stormwater infrastructure failures led to spikes in pollution. For instance, is there a threshold rainfall (e.g., >10mm) that consistently leads to poor water quality?
* **Stormwater Pit Condition and Pollution Levels:** Analysis of how unmaintained or weak stormwater pits directly contribute to pollution during heavy rain events.
* **Geographical Impact:** Some regions may consistently show higher pollution levels due to poor infrastructure or dense urban areas. Insight here can explain why certain catchments are more vulnerable.
* **User Response to Recommendations:** Understand why users may have ignored or adhered to the tool’s recommendations. Are they more likely to take precautions in certain weather conditions, or is the advice followed based on pollution severity alone?

#### **3. Foresight**

#### **Pollution Level Predictions:** Using patterns from rainfall, stormwater pit data, and temperature trends, predict water pollution levels in real time and in the near future. This will help forecast whether swimming will be safe in the next 24 hours or week.

* **Impact of Future Weather Conditions:** Predict upcoming pollution risk based on expected rainfall and temperature changes. For example, the tool can give advance warnings for higher pollution risks during stormy weather.
* **Stormwater Pit Maintenance Forecast:** Predict future pollution risks based on scheduled stormwater pit maintenance, or the lack thereof. Regions with poorly maintained pits could be flagged for higher future pollution risks.
* **User Behavior Forecast:** Based on historical user engagement data, predict how users will interact with the tool during specific conditions. For example, during summer, users may need more reminders to consider pollution levels due to increased swim activity.

## 4.7 Ethical, Legal, and Privacy Issues

#### **Ethics Considerations**

* **Transparency:** It’s essential to ensure users fully understand how decisions are made about water safety. The tool should provide clear, easy-to-understand explanations about pollution risks, how data is used, and the basis for "Good/Fair/Poor" swimming recommendations. Misleading users about potential risks could have severe health implications.
* **Health and Safety:** The primary ethical obligation is to safeguard public health. If there is uncertainty in the pollution predictions, the tool should err on the side of caution and provide adequate warnings to avoid health risks like infections or illnesses from polluted water.
* **Environmental Responsibility:** Encouraging users to follow pollution prevention tips is crucial for minimizing the impact of human activities on natural water systems. Ethical design should foster environmental consciousness and promote sustainable behavior to reduce overall stormwater pollution.

#### **2. Legal Considerations**

* **Liability for Health Outcomes:** The tool must include disclaimers, clearly stating that it provides recommendations based on the best available data but cannot guarantee complete safety. Users must acknowledge this limitation and understand that they are responsible for their decisions, which helps mitigate legal risks in case of illness due to water exposure.
* **Data Accuracy and Regulatory Compliance:** The tool should comply with local environmental regulations and standards for water safety. If it uses government data (e.g., stormwater, pollution levels), ensure that it is up-to-date and accurate, and abide by any data usage restrictions set by regulatory bodies.
* **Compliance with Open Data Licenses:** If the catchment and creek data, waste datasets, or weather data are sourced from government or open datasets, the tool must comply with the licensing requirements, which often include attribution and use limitations.

**3. Privacy Considerations**

* User Location Data Protection: Since the tool collects users' geo-location data to recommend nearby swimming spots, strict privacy protections must be in place. Ensure that location data is only used for the specific purpose of the tool, and not shared with third parties without user consent.
* Anonymization and Data Security: Personal and usage data should be anonymized to protect users’ identities. The tool should implement strong encryption and data security measures to prevent unauthorized access to sensitive information, including user location and behavioral data.
* User Consent: The tool must explicitly ask for consent to collect location and environmental data. Users should be able to opt-out of data collection features without losing access to the essential functionality of the tool.
* Data Retention Policy: Ensure that user data, particularly sensitive geo-location data, is only retained for as long as necessary. Establish clear data deletion policies to avoid unnecessary storage of private information.