Data Collection Plan

**Project Title:** [Name of Your AI Project]

**Project Manager:** [Your Name]

**Date:** [Date of Plan Creation]

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Author | Description |
| 1 | 13 July 2022 | Glynn Williams | Original |
|  |  |  |  |
|  |  |  |  |

**Distribution**

|  |  |  |
| --- | --- | --- |
| Organisation | Name | Position |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Copyright**

This document remains the property of Emphasys Pty Ltd. It cannot be copied or disclosed to any party without the express permission of Emphasys Pty Ltd.

**Contact**

Any queries relating to this document should be addressed to:

Name: Glynn C. Williams

Emphasys Pty Ltd

Telephone: 1300 810 797

Email: [glynn.williams@emphasys.com.au](mailto:glynn.williams@emphasys.com.au)

Table of Contents

[1 Document Version Control 2](#_Toc146721707)

[Objectives of Data Collection 4](#_Toc146721708)

[Data Sources 4](#_Toc146721709)

[Data Types 4](#_Toc146721710)

[Data Volume 4](#_Toc146721711)

[Data Quality 5](#_Toc146721712)

[Data Collection Methods 5](#_Toc146721713)

[Data Collection Schedule 5](#_Toc146721714)

[Data Cleaning and Pre-processing 5](#_Toc146721715)

[Data Storage 6](#_Toc146721716)

[Data Documentation 6](#_Toc146721717)

[Data Ethics and Compliance 6](#_Toc146721718)

[Data Backup and Recovery 6](#_Toc146721719)

[Data Retention: 7](#_Toc146721720)

[Data Ownership and Governance 7](#_Toc146721721)

[Data Collection Team 7](#_Toc146721722)

[Monitoring and Quality Control 7](#_Toc146721723)

[Data Collection Costs 8](#_Toc146721724)

[Risk Assessment 8](#_Toc146721725)

[Data Collection Plan Approval 8](#_Toc146721726)

[References 8](#_Toc146721727)

Objectives of Data Collection

*[Define the specific goals and objectives of data collection for the AI project. What insights or patterns are you seeking to discover from the data?]*

Data Sources

*[Identify the sources from which you will collect data. This may include databases, APIs, sensors, surveys, or any other relevant sources.]*

## Data Types

*[List the types of data to be collected, such as structured, semi-structured, or unstructured data. Specify data formats (e.g., CSV, JSON, text) and any special requirements.]*

Data Volume

*[Estimate the volume of data required for the project. Include details such as the number of records, rows, or files to be collected.]*

## Data Quality

*[Describe the criteria for data quality, including accuracy, completeness, consistency, and relevance. Specify how data quality will be assessed and maintained throughout the project.]*

## Data Collection Methods

***[****Outline the methods and tools that will be used to collect data. This may involve data scraping, data entry, data streaming, or data acquisition from third-party providers.]*

## Data Collection Schedule

***[****Create a timeline that specifies when data collection will begin, how often it will occur, and when it is expected to conclude.]*

## Data Cleaning and Pre-processing

***[****Describe the steps for data cleaning and pre-processing. This includes handling missing values, outliers, and data transformation procedures.]*

## Data Storage

*[Specify where and how the collected data will be stored. Address data security, encryption, and access control measures.]*

## Data Documentation

*[Develop a data dictionary or metadata documentation that explains the meaning and structure of collected data variables.]*

## Data Ethics and Compliance

*[Ensure compliance with data privacy regulations (e.g., GDPR) and ethical considerations. Document consent and anonymization processes if applicable.]*

## Data Backup and Recovery

*[Implement data backup and recovery procedures to prevent data loss.]*

## Data Retention:

*[Define the data retention policy, including how long data will be retained and when it should be archived or deleted.]*

## Data Ownership and Governance

***[****Clarify data ownership and governance responsibilities within the project team.]*

## Data Collection Team

*[Identify team members responsible for data collection, cleaning, and pre-processing. Assign roles and responsibilities.]*

## Monitoring and Quality Control

*[Establish procedures for monitoring data collection processes and ensuring data quality throughout the project's lifecycle.]*

## Data Collection Costs

*[Estimate the budget required for data collection, including tools, personnel, and any third-party data acquisition costs.]*

## Risk Assessment

*[Identify potential risks associated with data collection and outline mitigation strategies.]*

## Data Collection Plan Approval

***[****Obtain approval from relevant stakeholders and project sponsors for the data collection plan.]*

## References

*[List any external sources or references consulted while creating the data collection plan.]*