# Dataset Documentation Card Template

# Dataset Documentation Card Template

**Dataset Name**: [Descriptive name for the dataset] **Version**: [Version number, e.g., v1.0] **Date Created**: [YYYY-MM-DD] **Last Updated**: [YYYY-MM-DD] **Created By**: [Names and affiliations] **Contact**: [Email for questions about this dataset]

## Dataset Overview

### Purpose and Scope

**Primary Purpose**: [Why was this dataset created?] **Research Questions**: [What research questions does this dataset address?] **Intended Use Cases**: [How is this dataset intended to be used?] **Scope**: [What does this dataset cover? Time period, geographic area, population, etc.]

### Dataset Summary

**Total Records**: [Number of observations/participants/data points] **Data Collection Period**: [Start date] to [End date] **Geographic Coverage**: [Where data was collected] **Population**: [Who or what is represented in the data] **Data Types**: [Survey responses, behavioral data, text, images, etc.]

## Data Collection

### Collection Methodology

**Data Collection Method**: [Survey, experiment, observation, archival, etc.] **Collection Instruments**: [Surveys, interview guides, measurement tools, etc.] **Data Collection Team**: [Who collected the data] **Quality Control Measures**: [How data quality was ensured]

### Sampling Strategy

**Target Population**: [Who was the intended population] **Sampling Method**: [Random, convenience, stratified, etc.] **Sample Size Calculation**: [How sample size was determined] **Recruitment Method**: [How participants were recruited] **Response Rate**: [If applicable]

### Inclusion/Exclusion Criteria

**Inclusion Criteria**: - [Criterion 1] - [Criterion 2] - [Criterion 3]

**Exclusion Criteria**: - [Criterion 1] - [Criterion 2] - [Criterion 3]

## Data Structure and Variables

### File Structure

dataset-name/  
├── data/  
│ ├── raw/ # Original, unprocessed data  
│ ├── processed/ # Cleaned and processed data  
│ └── analysis-ready/ # Final datasets for analysis  
├── documentation/  
│ ├── codebook.md # Variable definitions and coding  
│ ├── data-card.md # This document  
│ └── collection-protocol.md # Data collection procedures  
├── code/  
│ ├── cleaning/ # Data cleaning scripts  
│ ├── processing/ # Data processing scripts  
│ └── analysis/ # Analysis code  
└── README.md # Quick start guide

### Key Variables

| Variable Name | Type | Description | Values/Range | Missing Data Code |
| --- | --- | --- | --- | --- |
| [var\_name] | [numeric/categorical/text/date] | [Description] | [Possible values] | [How missing data is coded] |
| [var\_name] | [numeric/categorical/text/date] | [Description] | [Possible values] | [How missing data is coded] |
| [var\_name] | [numeric/categorical/text/date] | [Description] | [Possible values] | [How missing data is coded] |

### Data Formats

**Primary Format**: [CSV, JSON, Excel, etc.] **File Encoding**: [UTF-8, etc.] **Date Format**: [YYYY-MM-DD, etc.] **Missing Data Representation**: [NA, NULL, -999, etc.] **Categorical Coding**: [How categories are coded]

## Data Quality and Limitations

### Data Quality Assessment

**Completeness**: [Percentage of complete records, missing data patterns] **Accuracy**: [How accuracy was verified] **Consistency**: [Internal consistency checks performed] **Validity**: [How validity was assessed]

### Known Limitations

**Sampling Limitations**: - [Limitation 1 and its implications] - [Limitation 2 and its implications]

**Measurement Limitations**: - [Limitation 1 and its implications] - [Limitation 2 and its implications]

**Temporal Limitations**: - [Time-specific factors that may affect generalizability]

**Other Limitations**: - [Any other important limitations]

### Data Cleaning and Processing

**Cleaning Steps Performed**: 1. [Step 1: Description of cleaning procedure] 2. [Step 2: Description of cleaning procedure] 3. [Step 3: Description of cleaning procedure]

**Outlier Treatment**: [How outliers were identified and handled] **Missing Data Treatment**: [How missing data was handled] **Data Transformations**: [Any transformations applied to variables]

## Ethical Considerations

### Human Subjects Protection

**IRB Approval**: [ ] Yes [ ] No [ ] Not Required **IRB Number**: [If applicable] **Consent Process**: [How informed consent was obtained] **Participant Rights**: [How participant rights were protected]

### Privacy and Confidentiality

**Personally Identifiable Information**: [What PII is included, if any] **De-identification Process**: [How data was de-identified] **Data Security Measures**: [How data is secured] **Access Controls**: [Who has access to what data]

### Potential Risks and Harms

**Privacy Risks**: [Potential privacy risks and mitigation] **Re-identification Risks**: [Risk of re-identification and mitigation] **Bias and Fairness**: [Potential biases in data and implications] **Misuse Potential**: [How data could be misused and safeguards]

## Usage Guidelines

### Recommended Uses

**Appropriate Analyses**: - [Type of analysis 1]: [Why appropriate] - [Type of analysis 2]: [Why appropriate] - [Type of analysis 3]: [Why appropriate]

**Research Questions Well-Suited for This Data**: - [Research question 1] - [Research question 2] - [Research question 3]

### Discouraged Uses

**Inappropriate Analyses**: - [Type of analysis]: [Why inappropriate] - [Type of analysis]: [Why inappropriate]

**Cautions**: - [Important caution about interpretation] - [Important caution about generalization]

### Citation Requirements

**How to Cite This Dataset**: [Provide full citation format]

**Acknowledgments**: [Required acknowledgments for funding, institutions, contributors]

## Technical Information

### Software Requirements

**Minimum Requirements**: - [Software/language]: [Version] - [Required packages/libraries]

**Recommended Tools**: - [Tool 1]: [Why recommended] - [Tool 2]: [Why recommended]

### File Specifications

**File Sizes**: [Approximate sizes of data files] **Storage Requirements**: [Total storage needed] **Download Information**: [Where and how to access data] **Checksums**: [If provided for data integrity verification]

## Maintenance and Updates

### Version History

| Version | Date | Changes | Updated By |
| --- | --- | --- | --- |
| v1.0 | [YYYY-MM-DD] | Initial release | [Name] |
| v1.1 | [YYYY-MM-DD] | [Description of changes] | [Name] |

### Maintenance Plan

**Update Schedule**: [How often data/documentation will be updated] **Maintenance Responsibility**: [Who is responsible for maintenance] **End-of-Life Plan**: [When and how dataset will be archived]

### Contact for Updates

**Primary Contact**: [Name and email] **Institution**: [Affiliation] **Alternative Contact**: [Name and email]

## Related Resources

### Associated Publications

* [Citation 1]: [Brief description of how dataset was used]
* [Citation 2]: [Brief description of how dataset was used]

### Related Datasets

* [Dataset name]: [Relationship to this dataset]
* [Dataset name]: [Relationship to this dataset]

### Code Repositories

* [Repository name]: [URL] - [Description of code]
* [Repository name]: [URL] - [Description of code]

### Documentation

* [Document name]: [Description and location]
* [Document name]: [Description and location]

## Appendices

### Appendix A: Data Collection Instruments

[Include or reference surveys, interview guides, etc.]

### Appendix B: Detailed Variable Descriptions

[Extended codebook information if needed]

### Appendix C: Quality Control Procedures

[Detailed description of quality control measures]

**Document Status**: [ ] Draft [ ] Under Review [ ] Final **Review Date**: [When this document should be reviewed next] **Approved By**: [Name and title of approver]