**I-GUIDE DATA CARD**

The I-GUIDE Data Card is an easy-to-use tool that will allow you to create documentation for each dataset that you create or use in a project.

Using this tool will help facilitate transparency and reproducibility of your project. It will also help you comply with data management and sharing policies of journals, funding agencies, and universities.

The Data Card applies to:

1. **Secondary Datasets**: Data sourced from external repositories or other researchers;
2. **Primary Datasets**:
   * Data collected through experiments, fieldwork, or user-generated sources;
   * Data obtained via web scraping, API collection, or similar automated means.

**Data Card Attribution**

This Data Card template is an adapted version of materials originally developed by Google’s *Data Cards Playbook*, available at https://pair-code.github.io/datacardsplaybook/.  
Adaptations have been made to support the I-GUIDE platform, with a focus on simplifying structure, tailoring prompts for geospatial data, and aligning with ethical and FAIR data practices in scientific research.

This work is shared under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).  
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

A logo with text on it

AI-generated content may be incorrect.

* + 1. **BASIC INFORMATION**

|  |  |
| --- | --- |
| Data Card ID Number | *(Suggested format: DC-01, DC-02, etc.)* |
| Dataset Name | *(Include a short, descriptive name)* |
| Dataset Version | *(Specify version number and release date)* |
| Persistent Identifier | *(Provide a DOI)* |
| Outputs Supported | *(List publications whose findings are supported by this dataset)* |
| Data Card Author | *(State name, affiliation, contact information)* |

* + 1. **DATASET OVERVIEW**

|  |  |
| --- | --- |
| Dataset Owners and Publishers | *(Individuals, organizations, or institutions responsible for dataset; add contact information)* |
| Source and Acquisition Method | ☐ Primary dataset from digital sources or automated extraction  ☐ Primary dataset from direct measurements or observations  ☐ Primary dataset from simulation or synthetic methods  ☐ Secondary dataset (from external source) |
| Terms of Use, or Data Sharing Agreement | *(E.g., CC-BY license, CC0 license, proprietary)* |
| Storage Location | ☐ Repository: *(Name and link to dataset)*  ☐ Project-specific storage: *(Describe location)* |
| Access Control Policies | ☐ Open  ☐ Embargoed: *(Describe release timeline)*  ☐ Restricted: *(Describe access criteria)* |

1. **DATASET CHARACTERISTICS**

|  |  |
| --- | --- |
| Data Subjects | ☐ Natural phenomena (e.g., weather patterns, water levels)  ☐ Objects (e.g., buildings, infrastructure, vehicles)  ☐ People (e.g., demographic data, survey responses)  ☐ Places (e.g., geographical region, urban area)  ☐ Systems (e.g., transport networks, ecological interactions)  ☐ Other: *(Specify)* |
| Dataset Size | *(File size, number of records, variables, resolution etc.)* |
| Spatial Data | ☐ Yes  ☐ No  If “Yes”:  Coordinate Reference System (CRS): *(If applicable)*  Spatial Resolution: (*Grid size, cell size etc.)*  Temporal Resolution: *(Time intervals, data format, update frequency)* |
| Data Modality | ☐ Graph  ☐ Image  ☐ Tabular  ☐ Text  ☐ Time series  ☐ Multimodal: *(Specify)*  ☐ Other: *(Specify)* |
| Variables | *(Complete for each variable. Table for recording this information is appended at the end of this document)*  Variable name:  Brief description: |

1. **PROVENANCE**

|  |  |
| --- | --- |
| Methods of Collection | ☐ API  ☐ Artificially generated  ☐ Crowdsourced – paid  ☐ Crowdsourced – volunteer  ☐ Scraped or crawled  ☐ Sensor based  ☐ Surveys, forms, interviews, or polls  ☐ Other: *(Specify)* |
| More About Methods | *(Include sampling methods, criteria for inclusion and exclusion of data points, known limitations of methods)* |
| Tools and Libraries Used | *(List software, programming libraries, and any scripts used for data collection, with links where possible)* |
| Collection Policies (if data collected using web scraping or other digital methods) | *(Relevant policies governing data collection, e.g., platform terms of use)* |

1. **SENSITIVE DATA**

|  |  |
| --- | --- |
| Human Subject Identifiability | ☐ None  ☐ Personally identifiable information  ☐ Pseudonymous data  ☐ Anonymous data |
| Other Sensitivity Factors | ☐ Commercially sensitive data  ☐ Health data  ☐ Data about children  ☐ Data about marginalized group: *(Specify)*  ☐ Location sensitive data  ☐ Military or security related data  ☐ Restricted government data  ☐ Surveillance data |
| Measures Taken to Handle Sensitive Data | *(Describe de-identification, anonymization, encryption, and/or access restrictions)* |
| Demographic Variables Represented in Dataset | ☐ Age  ☐ Culture  ☐ Disability status  ☐ Ethnicity  ☐ Gender  ☐ Language  ☐ Nationality  ☐ Race  ☐ Socio-economic status  ☐ Other: *(Specify)* |
| Correlation with Demographic Variables | *(Describe any variables that correlate with demographic data, and provide correlation metrics where applicable)* |
| Dataset Representativeness | *(Describe any known issues regarding demographic and geographic representativeness, e.g., underrepresentation or overrepresentation of specific demographic groups)* |
| Information About Ethical Oversight | ☐ Not subject to Institutional Review Board (IRB) approval: *(Briefly explain why not)*  ☐ Subject to Institutional Review Board (IRB) approval  Name of IRB:  Link to IRB website: *(URL)*  Primary IRB contact: *(Name, Email)*  Approval dates: *(From, To)*  Reference number: |
| Informed Consent Processes | (*Describe how consent was obtained, key elements covered, and extra measures taken to facilitate consent (e.g., with special populations). If waived, note the reason)* |

1. **TRANSFORMATIONS**

|  |  |
| --- | --- |
| Transformations Applied | ☐ Anomaly detection  ☐ Cleaning mismatched values  ☐ Cleaning missing values  ☐ Converting data types  ☐ Data aggregation  ☐ Dimensionality reduction  ☐ Joining input sources  ☐ Redaction or anonymization  ☐ Other: *(Specify)* |
| Description of Transformations | *(Complete one version for each transformation)*  Transformation applied:  Field(s) transformed: *(List affected variables)*  Reason for transformation:  Who carried out transformation:  Methods applied:  Platform, tool, or libraries used (including link): |

1. **SUITABLE AND UNSUITABLE USES OF DATASET**

|  |  |
| --- | --- |
| Suitable Uses | *(Describe suitable types of use cases for this dataset)* |
| Unsuitable Uses | *(Describe unsuitable types of use cases for this dataset, e.g., cases for which the data is of inadequate quality, or may pose risks of misuse or misinterpretation, or may pose risks relating to stigmatization or other harm to communities)* |

1. **ANNOTATION TASKS (only complete if dataset includes labeled data)**

|  |  |
| --- | --- |
| Types of Annotation Performed | ☐ Annotation target in data: *(Specify what was being labeled)*  ☐ Crowdsourced  ☐ Human (expert)  ☐ Human (non-expert)  ☐ Machine-generated *(Describe how system generated labels)*  ☐ Other: *(Specify)* |
| Description of Annotations | *(Complete one version for type of annotation)*  Number of unique annotations: *(Total distinct labels/categories)*  Total annotations:  Platforms, tools, or libraries (include link):  Task description*)*  Methods used:  Inter-rater adjudication policy: |

1. **APPLICATIONS AND BENCHMARKS (only complete if dataset used for AI purposes)**

|  |  |
| --- | --- |
| Relevant AI Model(s) | *(List models trained, tested, or validated on this dataset, or for which this dataset has served as input data)* |
| Use in AI | ☐ Training  ☐ Testing  ☐ Validation  ☐ Fine-tuning |
| Key AI Tasks | *(Describe purposes of the AI application, e.g., forecasting)* |
| Evaluation Results | *(Provide accuracy, precision-recall, F1-score, or other performance benchmarks)* |

**APPENDIX: VARIABLES (see “Variables” field in Section 2, “Dataset Characteristics”; add extra rows if necessary)**

|  |  |
| --- | --- |
| **VARIABLE NAME** | **BRIEF DESCRIPTION** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |