**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 | DC-01 |
| Dataset Name | ERA5 Reanalysis Product |
| Dataset Version | ERA5 v5.0, released 2019 (updated monthly) |
| Persistent Identifier | https://doi.org/10.24381/cds.adbb2d47 |
| Outputs Supported | Climate analyses, drought and flood risk assessments, precipitation and temperature trend studies. |
| Data Card Author | IGUIDE- Team 1: Jennifer Marlon, Deepika Pingali, Surabhi Upadhyay, Emine Senkardesler, Pratyush Tripathy, Okikiola Michael Alegbeleye |

* + 1. **DATASET OVERVIEW**

|  |  |
| --- | --- |
| Dataset Owners and Publishers | European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the Copernicus Climate Change Service (C3S) - https://cds.climate.copernicus.eu |
| Source and Acquisition Method | ☑ Secondary dataset (from external source) |
| Terms of Use, or Data Sharing Agreement | Copernicus open data license (free and unrestricted use with attribution). |
| Storage Location | ☑ Repository: [*https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview*](https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview) |
| Access Control Policies | ☑ Open |

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 | Varies by request; global coverage, hourly data from 1940–present. Each monthly file ≈ 12 GB (netCDF). |
| Spatial Data | ☑ Yes  ☐ No  If “Yes”:  Coordinate Reference System (CRS): *WGS84 (EPSG:4326)*  Spatial Resolution: 0.25° × 0.25° (~31 km at equator)  Temporal Resolution: Hourly, aggregated to monthly or yearly as needed |
| Data Modality | ☐ Graph  ☐ Image  ☐ Tabular  ☐ Text  ☑ Time series  ☐ Multimodal: *(Specify)* |
| Variables | *(Complete for each variable. Table for recording this information is appended at the end of this document)*  Variable name: tp (Total Precipitation Mean, m/day) and t2m (2m Air Temperature Maximum, K)  Brief description: tp: Sum of all liquid and frozen water (rain + snow + other) reaching the surface, in meters (m) of water equivalent and t2m: Represents the mean air temperature at 2 meters above the surface over the specified period, in Kelvin which reflects near-surface thermal conditions. |

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 | *CDS API, Python (cdsapi, xarray, netCDF4) for download and processing.* |
| 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: Data aggregation  Field(s) transformed: ERA5 daily/hourly precipitation and temperature variables  Reason for transformation: To summarize high-resolution ERA5 climate variables into monthly aggregated forms for alignment with survey data.  Who carried out transformation: Team 1  Methods applied: Temporal aggregation from daily data to monthly statistics (e.g., total precipitation, mean temperature)  Platform, tool, or libraries used (including link): Python with xarray (https://docs.xarray.dev/), geopandas (https://geopandas.org/), and rasterstats (https://pythonhosted.org/rasterstats/) |

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

|  |  |
| --- | --- |
| Suitable Uses | Climate analysis at district, state, or regional level for research and policy applications.  Aggregating precipitation and temperature data for use as covariates in environmental, agricultural, or disaster risk models.  Historical climate variability assessment and integration into vulnerability mapping. |
| Unsuitable Uses | Real-time weather forecasting or operational decision-making (ERA5 is reanalysis, not live).  Localized microclimate assessments at household or street level — resolution is too coarse. |

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** |
| **Tp (Total Precipitation), t2m (2m Temperature)** | **ERA5 total precipitation aggregated to district and time period and ERA5 mean 2m air temperature aggregated to district and time period** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |