

Data Dictionary Template

Use this template to document detailed field specifications after completing your data requirements. This template includes the PHI, PII, and sensitivity classifications from Step 2.

Dataset Information

Dataset Name: [Your dataset name]

Version: 1.0

Last Updated: [Date]

Use Case: [Personalized Health / Manufacturing Analytics / Fraud Detection]

Classification Legend

Classification	Description
PHI	Protected Health Information - HIPAA covered health data
PII	Personally Identifiable Information - Can identify an individual
Public	Can be shared openly without restrictions
Private	Internal use only, not for public distribution
Confidential	Need-to-know basis, business-sensitive
Restricted	Highest protection level, requires special access

Example 1: Personalized Health - Oura Ring API

Field Name	Data Type	PHI	PII	Sensitivity	Description	Valid Values	Example
user_id	string	No	Yes	Restricted	Unique user identifier from Oura	UUID format	usr_a1b2c3d4
date	date	No	No	Private	Date of the recorded metrics	ISO 8601	2026-01-12
sleep_score	integer	Yes	No	Confidential	Overall sleep quality score	0-100	85
hrv_avg	integer	Yes	No	Confidential	Average heart rate variability (ms)	10-150	42
resting_hr	integer	Yes	No	Confidential	Resting heart rate (bpm)	35-100	58
deep_sleep_mins	integer	Yes	No	Confidential	Deep sleep duration in minutes	0-300	90
rem_sleep_mins	integer	Yes	No	Confidential	REM sleep duration in minutes	0-200	105

Field Name	Data Type	PHI	PII	Sensitivity	Description	Valid Values	Example
readiness_score	integer	Yes	No	Confidential	Daily readiness score	0-100	78

Example 2: Manufacturing - IoT Sensors

Field Name	Data Type	PHI	PII	Sensitivity	Description	Valid Values	Example
sensor_id	string	No	No	Public	Unique sensor identifier	Pattern: SNS-[0-9]{4}	SNS-0042
equipment_id	string	No	No	Public	Machine/equipment identifier	Pattern: CNC-[0-9]{3}	CNC-042
timestamp	datetime	No	No	Public	Time of sensor reading	ISO 8601	2026-01-12T14:30:00Z
temperature_c	float	No	No	Public	Equipment temperature in Celsius	20.0-150.0	72.4
vibration_hz	float	No	No	Public	Vibration frequency (Hz)	0-500	145.2
pressure_psi	float	No	No	Public	Pressure reading (PSI)	0-200	85.5
oil_level_pct	float	No	No	Public	Oil level percentage	0-100	78.5
runtime_hours	float	No	No	Public	Total runtime hours	0-100000	4523.5

Example 3: Fraud Detection - Transaction Stream

Field Name	Data Type	PHI	PII	Sensitivity	Description	Valid Values	Example
transaction_id	string	No	No	Private	Unique transaction identifier	UUID	TXN-2026011214300001
account_id	string	No	Yes	Restricted	Customer account number	Pattern: ACC-[0-9]{8}	ACC-00045821
timestamp	datetime	No	No	Private	Transaction time	ISO 8601	2026-01-12T14:30:00Z
amount	float	No	No	Confidential	Transaction amount in USD	0.01-1000000	47.52
merchant_category	string	No	No	Public	MCC code for merchant type	4-digit MCC	5411

Field Name	Data Type	PHI	PII	Sensitivity	Description	Valid Values	Example
merchant_name	string	No	No	Public	Name of merchant	Free text	SAFEWAY #1234
location	string	No	Yes	Private	Transaction location	City, State format	Seattle, WA
channel	string	No	No	Public	Transaction channel	card_present, card_not_present, ach, wire	card_present

Template: Your Data Dictionary

Data Source: [Name]

Field Name	Data Type	PHI	PII	Sensitivity	Description	Valid Values	Example

Data Quality Rules

Validation Rules (customize for your use case)

- Identifiers:** Must be unique across the dataset
- Dates/Times:** Must be in ISO 8601 format
- Numeric Ranges:** Values must fall within Valid Values range
- Required Fields:** PHI=Yes or PII=Yes fields require encryption
- Cross-field:** Timestamps should be chronologically consistent

Null Handling

Sensitivity Level	Null Policy
Public	Allowed with default value
Private	Allowed, log missing count
Confidential	Allowed, requires review
Restricted	Never null, reject record

Access Control Matrix

Role	Public	Private	Confidential	Restricted
Data Scientist	✓	✓	✓ (de-identified)	✗
ML Engineer	✓	✓	✓ (de-identified)	✗
System Admin	✓	✓	✗	✗
Compliance Officer	✓	✓	✓	✓ (audit only)

Role	Public	Private	Confidential	Restricted
External Auditor	✓	✗	✗	✗

Change Log

Version	Date	Changes	Author
1.0	[Date]	Initial data dictionary	[Name]

Notes

- All synthetic data in this dictionary is for educational purposes
- PHI and PII fields require encryption at rest and in transit
- Restricted fields require multi-factor authentication for access
- Update this dictionary when schema changes occur