

Datasheets for Datasets Template

Type: Documentation Template | **Audience:** Data Engineers, ML Teams, Data Governance

1. Motivation

Question	Response
Why was this dataset created?	
Who created the dataset?	
Who funded the creation?	
What tasks is it intended for?	

2. Composition

Question	Response
What does each instance represent?	
How many instances total?	
Is this a sample? Of what population?	
What data is included per instance?	
Are there labels? How obtained?	
Is there confidential information?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes: _____
Could content be offensive?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes: _____
Does it relate to people?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes: _____

3. Collection Process

Question	Response
How was data collected?	
Who collected the data?	
Over what timeframe?	
Was consent obtained?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> N/A Details: _____
Ethics review conducted?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes, by: _____

4. Preprocessing / Cleaning / Labeling

Question	Response
What preprocessing was done?	
Is raw data preserved?	<input type="checkbox"/> Yes <input type="checkbox"/> No Location: _____
Software used for preprocessing?	
How was data labeled?	
Who performed labeling?	

5. Uses

Question	Response
Intended uses?	
Prohibited uses?	
Consequences of misuse?	
Other known uses?	

6. Distribution

Question	Response
How is it distributed?	
Is it currently available?	<input type="checkbox"/> Yes <input type="checkbox"/> No
License type?	
Any restrictions (export, regulatory)?	

7. Maintenance

Question	Response
Who maintains the dataset?	
Contact information?	
Will it be updated?	<input type="checkbox"/> Yes <input type="checkbox"/> No Frequency: _____
How will errors be communicated?	
Retention period?	

Dataset Name: _____

Prepared By: _____ **Date:** _____