
The Data Cards Playbook

A toolkit for purposeful and people-centric dataset documentation for transparency in AI systems.

<https://pair-code.github.io/datacardsplaybook/>

#datacardsplaybook



THE DATA CARDS PLAYBOOK

Introduction

01 Ask

02 Inspect

03 Answer

04 Audit



THE DATA CARDS PLAYBOOK

OFTEn Distribution

IN THIS SECTION

Check to see if your Data Card is well-balanced and represents the Data Lifecycle meaningfully.

INSTRUCTIONS

Assign each question an OFTE_n category. Track the distribution and discuss if the Data Card is an appropriate depiction of your dataset's lifecycle.

OUTCOMES

A Data Card template that is representative of your datasets' lifecycles.

ACTIVITY LEVEL

Basic

Definitive Questions

List questions from the Data Card template that fall squarely into **Origins**

1. Add questions here

Provisional Questions

List questions from the Data Card template that might fall into **Origins**

1. Add questions here

Missing & Action Items

What steps must be taken to refine or address shortcomings?

- Action items

Origins

Questions about early stages of a dataset's lifecycle when decisions to create a dataset are made. Includes various planning activities such as defining requirements, collection or sourcing methods, design decisions, and formulating policies which dictate final outcome.

Total number of Questions

Count the number of questions that fall in the two boxes:

Definitive Questions

123456

Provisional Questions

123456

Factuals

Questions about actual data collection processes and outcomes of decisions in Origins. Includes statistical and other factual attributes that describe the dataset, inclusion and exclusion criteria, deviations from the original plan, and any pre-wrangling analysis.

Total number of Questions

Count the number of questions that fall in the two boxes:

Definitive Questions

123456

Provisional Questions

123456

Definitive Questions

List questions from the Data Card template that fall squarely into **Factuals**

1. Add questions here

Provisional Questions

List questions from the Data Card template that might fall into **Factuals**

1. Add questions here

Missing & Action Items

What steps must be taken to refine or address shortcomings?

- Action items

Transformations

Questions about operations like filtering, validating, parsing, formatting, and cleaning that transform raw data into a useful form. Includes summaries of labeling, annotation, or validation tasks, inter-rater adjudication processes, rater descriptions, feature engineering and modifications made to handle privacy, security, or PII.

Total number of Questions

Count the number of questions that fall in the two boxes:

Definitive Questions

123456

Provisional Questions

123456

Definitive Questions

List questions from the Data Card template that fall squarely into **Transformations**

1. Add questions here

Provisional Questions

List questions from the Data Card template that might fall into **Transformations**

1. Add questions here

Missing & Action Items

What steps must be taken to refine or address shortcomings?

- Action items

Experience

Questions about how the dataset might be used or is applied in experimental, production, or research practice. Includes using the data for specific tasks, undergoing access training, making modifications to suit the task, acquiring results and comparing to other similar datasets, and noting any expected/ unexpected behaviors.

Total number of Questions

Count the number of questions that fall in the two boxes:

Definitive Questions

123456

Provisional Questions

123456

Definitive Questions

List questions from the Data Card template that fall squarely into **Experience**

1. Add questions here

Provisional Questions

List questions from the Data Card template that might fall into **Experience**

1. Add questions here

Missing & Action Items

What steps must be taken to refine or address shortcomings?

- Action items

n=1 (Examples)

Questions about actual examples of data from the dataset - or vignettes - representing normal datapoints and outliers. Includes in-and-out-of distribution datapoints, demonstrates noteworthy data points with specific attributes, and where applicable, model outcomes on them.

Total number of Questions

Count the number of questions that fall in the two boxes:

Definitive Questions

123456

Provisional Questions

123456

Definitive Questions

List questions from the Data Card template that fall squarely into **n=1 (Examples)**

1. Add questions here

Provisional Questions

List questions from the Data Card template that might fall into **n=1 (Examples)**

1. Add questions here

Missing & Action Items

What steps must be taken to refine or address shortcomings?

- Action items

Final checks

YOU SHOULD NOW HAVE

—

- ✔ Categorized questions in your Data Card Template using OFTE n
- ✔ Checked to see if your Data Card template is well-balanced across the data lifecycle
- ✔ Identified action items for a more utilitarian and representative template



#datacardsplaybook



[The Data Cards Playbook](#) [↗] is an adaptable toolkit of participatory activities, conceptual frameworks, and guidance that support Responsible AI practices for transparency in dataset documentation.

If you've adapted, implemented, or have feedback for this guidance, we'd love to hear from you at <https://github.com/pair-code/datacardsplaybook> [↗].

Find the complete playbook at
<https://pair-code.github.io/datacardsplaybook> [↗]



The [Data Cards Playbook](#) [↗] by [the People + AI Research Initiative](#) [↗] at [Google Research](#) [↗] is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. You are free to share and adapt this work under the [appropriate license terms](#) [↗].