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

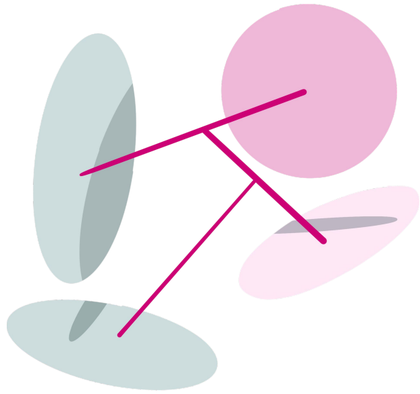
Introduction

01 Ask

02 Inspect

03 Answer

04 Audit



Lens Brainstorm

IN THIS SECTION

Co-create lenses to systematically frame Agent Information Journeys with stakeholders from across your dataset lifecycle for your Data Card(s).

INSTRUCTIONS

Individually brainstorm lenses for each agent in your dataset lifecycle. Share with the group, cluster and refine.

OUTCOMES

A list of agent-specific lenses to take forward in the Scopes Brainstorm.

ACTIVITY LEVEL

Moderate

How to frame a good lens

As a(n) engineer, I want to know about the access restrictions and policies associated with the dataset(s).

As a(n) analyst, I want to know about the data collection process and rationale.

✅ DO

Write a statement

Keep it high-level (applicable to many datasets)

Keep it short and concise

Convey a single idea per lens

Litmus test:
You should be able to break down a statement into multiple follow-up questions.

❌ DON'T

Ask a question

Make it too specific

Add too much detail

Write a statement that can be resolved with a yes/no answer.

Litmus test:
If you can answer it simply, it's a scope, not a lens.

EXAMPLE

What do lenses look like?

As a(n) [perspective], I want to know about the **publishers** of the dataset(s).	As a(n) [perspective], I want to know about the **data transformations** (cleaning, parsing, and processing) in the dataset(s).	As a(n) [perspective], I want to know about **access** restrictions and policies of the dataset(s).
As a(n) [perspective], I want to know about the **wipeout and retention** policies of the dataset(s).	As a(n) [perspective], I want to know about the **data upkeep** in the dataset(s).	As a(n) [perspective], I want to know about the **distributions** in the dataset(s).
As a(n) [perspective], I want to know about the **original or upstream sources** of the data.	As a(n) [perspective], I want to know about the **nature** (data modality, domain, format, etc.) of the dataset(s)	As a(n) [perspective], I want to know what **typical and outlier examples** in the dataset(s) look like.
As a(n) [perspective], I want to know explanations and **motivations for creating** the dataset(s).	As a(n) [perspective], I want to know the **intended applications** of the dataset(s).	As a(n) [perspective], I want to know about the **safety** (risks, limitations, and trade-offs) of using the dataset(s).
As a(n) [perspective], I want to know about any **known patterns** (correlations, biases, skews) within the dataset(s).	As a(n) [perspective], I want to know about the **differences across versions** of the dataset(s).	As a(n) [perspective], I want to know about **expectations of using** the dataset(s) with other datasets or tables.

10 min

[Your name] Brainstorm lenses for the agreed agents

✍️ As a(n) [perspective], I want to know [lens]	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens

10 min

Refine or write 1-3 lenses to represent each cluster:

✍️ As a(n) [perspective], I want to know [lens]	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens
✍️ Lens	✍️ Lens	✍️ Lens

15 min

Cluster your lenses

✍️ CLUSTER NAME

✍️ CLUSTER NAME

✍️ CLUSTER NAME

✍️ CLUSTER NAME

✍️ CLUSTER NAME

Checklist

YOU SHOULD NOW HAVE

–

- ✔ Brainstormed lenses for each agreed agent
- ✔ Explored what the same lens might be for different agents
- ✔ Prioritized which lenses to include in your Data Card for a given agent
- ✔ Everything you need for the scopes brainstorm

#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 ↗.