
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

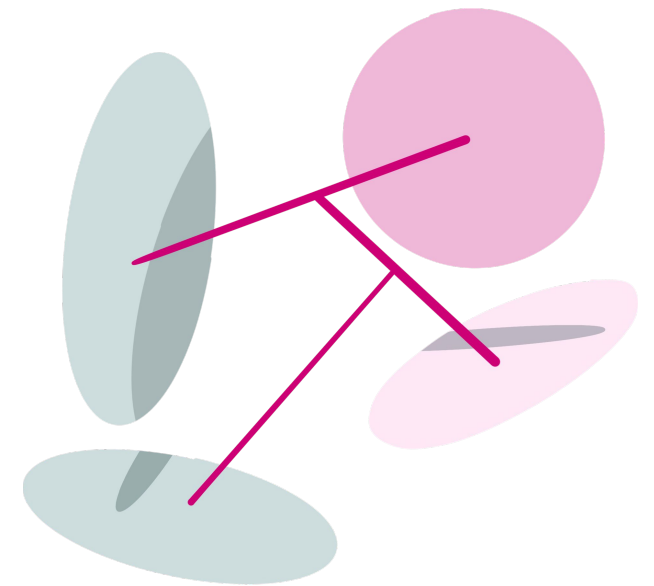
Introduction

01 Ask

02 Inspect

03 Answer

04 Audit



THE DATA CARDS PLAYBOOK

Scopes Brainstorm

IN THIS SECTION

Convert your lenses into scopes – questions that will become the building blocks of your Data Card(s).

INSTRUCTIONS

Break down each lens into sets of telescopic-periscopic-microscopic questions.

OUTCOMES

Questions that that will be prompts for producers to provide information that is most useful for agents.

ACTIVITY LEVEL

Moderate

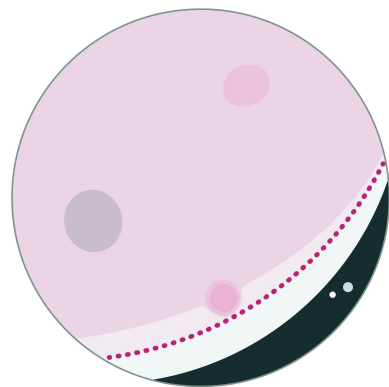
Types of scopes

Telescopic

Questions about attributes or characteristics commonly found across multiple datasets.

For example: Does this dataset contain **S/PII**?

Data Cards with just telescopes will describe obvious and not add any distinct value, but telescopes tell agents where to begin looking.

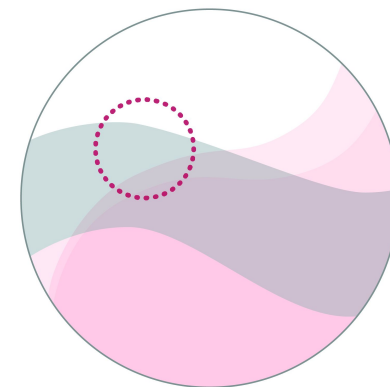


Periscopic

Questions about attributes or observations specific to the dataset being documented.

For example: Which **features in the dataset** contain **S/PII**?

Data Cards with just periscopes can get overly technical, without any details on context, relevance, or importance.

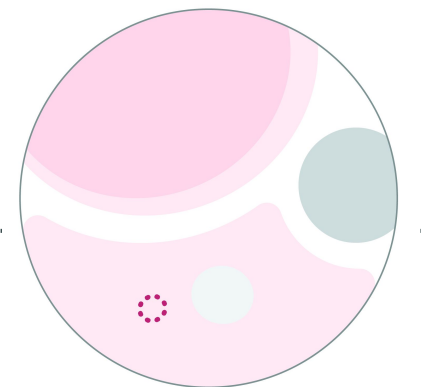


Microscopic

Questions about unobservable aspects of dataset(s), such as decisions, processes, impact.

For example: How was **S/PII handled** in this dataset?

If we only use microscopes, we can easily get lost in the details, and lose sight of the bigger picture.



	TELESCOPIC	PERISCOPIC	MICROSCOPIC
Description	Characteristics	Observations	Explanations
Input Type	Yes/No or multiple choice responses	Key-Value pairs or short descriptions	Detailed, long-form descriptions
Format	Tags, Chips	Text, Tables, and Visualizations w/ Links	Paragraphs, Tables, Visualizations, Links
Content	Universal attributes of datasets	Unique, observable characteristics of the dataset	Rationales, decisions, and policies that shape datasets
Utility	Find information quickly; Indexical – search or scan Keywords, descriptors	Facilitate quick assessments Tactical – what statements	Provide context and links to additional information Rationale and process – why and how statements
Value	Provide an overview and help habituate agents in the Data Card. Make Data Cards easy to index in repositories	Often reproducible from the dataset, prevent re-work, and add more context for agents when stacked	Unobservable attributes and considerations important to agents, but accessible only to producers

As a **data scientist**, I want to know
about the structure of the dataset, so I ask...

... what is the data format?

telescopic

... are there any media in the dataset?

... how many features are in the dataset?

... how many features are engineered?


periscopic


... which features are strongly correlated
and why? Was this intentional?


microscopic





[Your name] Brainstorm scopes for the agreed lenses


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


Telescopic Question: 
Expected Answers:


Periscopic Question: 
Expected Answers:


Microscopic Question: 
Expected Answers:


Telescopic Question: 
Expected Answers:

Periscopic Question: 
Expected Answers:

Microscopic Question: 
Expected Answers:

Telescopic Question: 
Expected Answers:

Periscopic Question: 
Expected Answers:

Microscopic Question: 
Expected Answers:



Cluster your scopes

 CLUSTER NAME

 CLUSTER NAME

 CLUSTER NAME

 CLUSTER NAME

 CLUSTER NAME

Checklist

YOU SHOULD NOW HAVE

—

- ✔ Sets of questions that probe different facets of the dataset
- ✔ A shared intuition around how questions are related to each other
- ✔ Well-organized scopes that are building blocks for your Data Card
- ✔ Expectations around what constitutes a response to the question



#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](#) [↗].