The Data Cards Playbook

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

https://pair-code.github.io/datacardsplaybook/
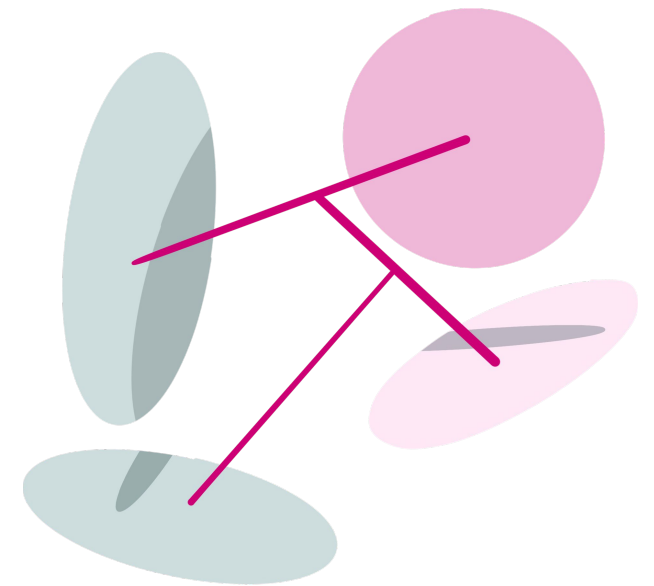
#datacardsplaybook

Assemble your Template

IN THIS SECTION

Structure your lenses and scopes into a Data Card template that works for producers and agents alike.

Organize lenses and scopes into a Data Card template by converting lenses to titles and structuring scopes as questions that Producers of datasets will fill out.
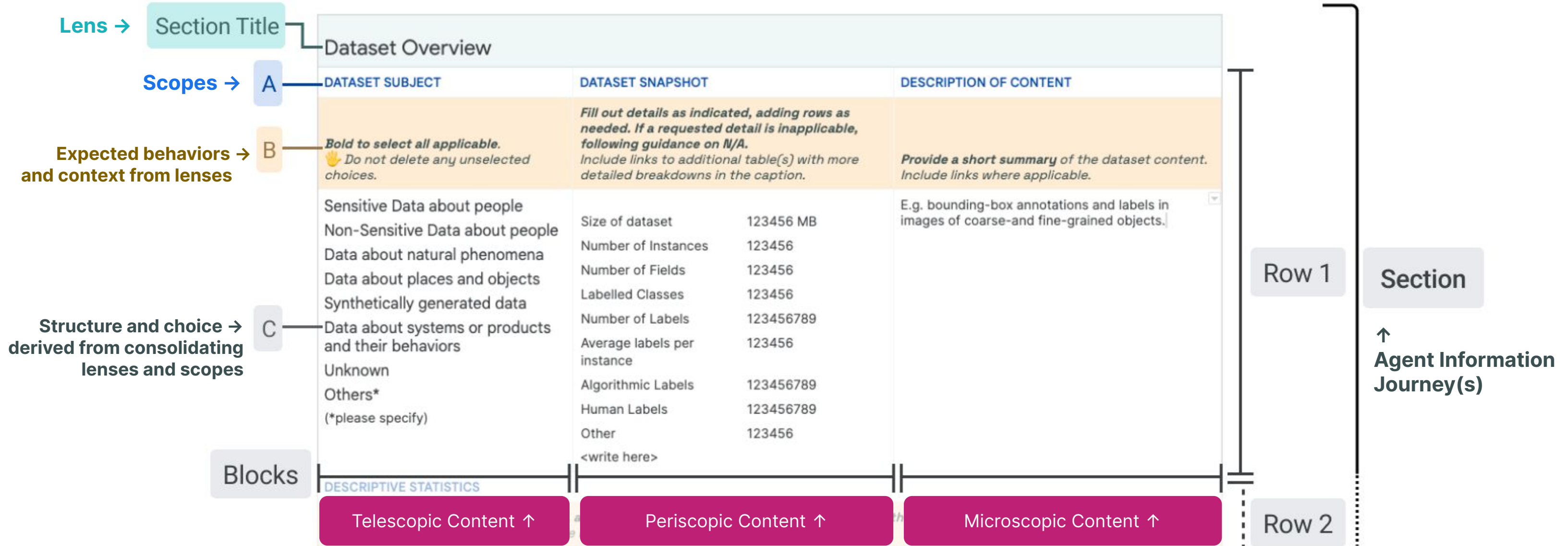
A Data Card template that contains the most useful information for you agents, which can be inspected, tested, and implemented.
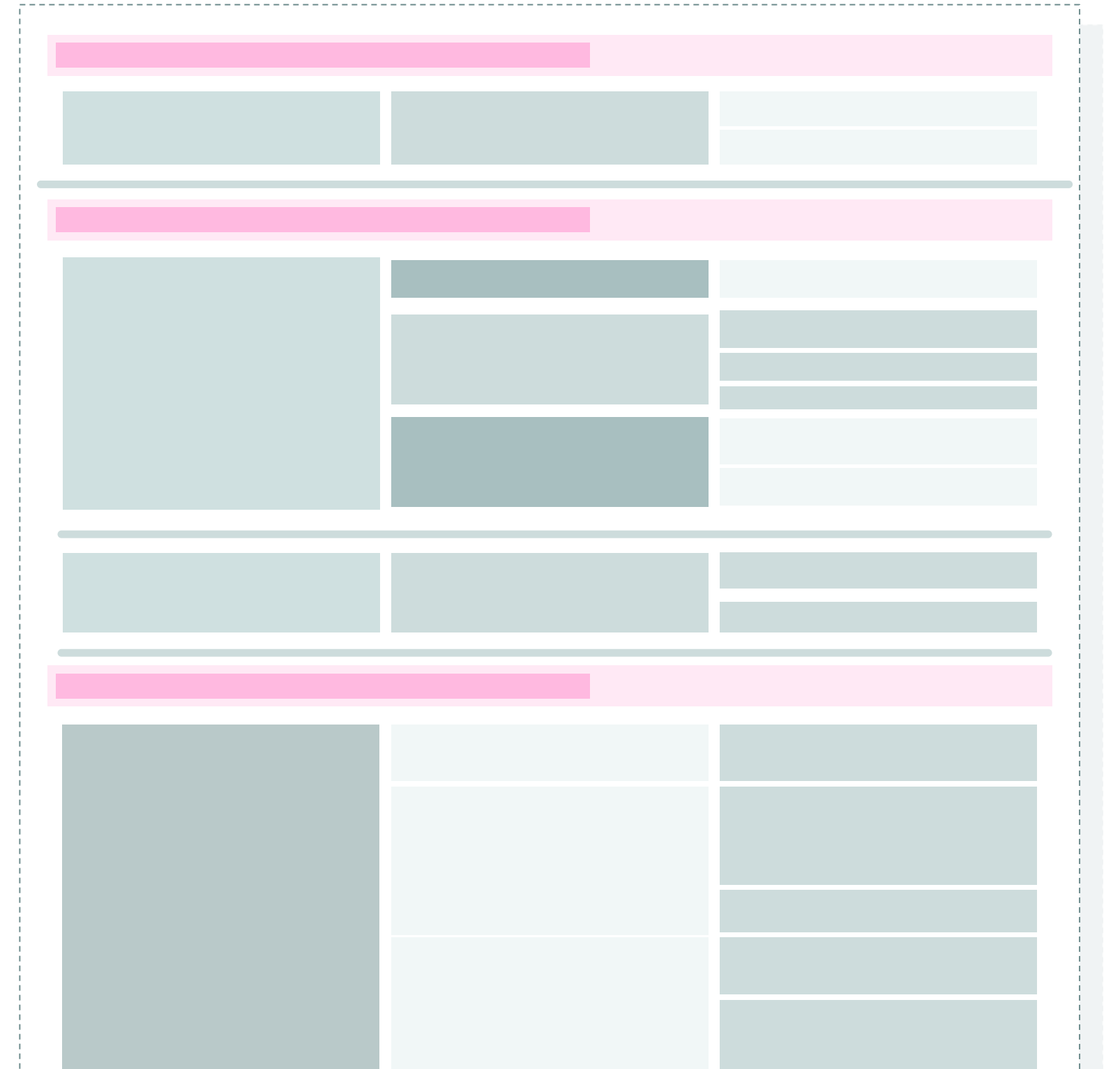
Moderate

Consider this structure

Lens → Section Title

Scopes → A

Expected behaviors →
and context from lenses B

Structure and choice →
derived from consolidating
lenses and scopes C

Blocks

Dataset Overview

DATASET SUBJECT	DATASET SNAPSHOT	DESCRIPTION OF CONTENT
Bold to select all applicable. ✋ *Do not delete any unselected choices.*	*Fill out details as indicated, adding rows as needed. If a requested detail is inapplicable, following guidance on N/A.* *Include links to additional table(s) with more detailed breakdowns in the caption.*	*Provide a short summary of the dataset content.* *Include links where applicable.*
Sensitive Data about people		E.g. bounding-box annotations and labels in images of coarse-and fine-grained objects.
Non-Sensitive Data about people	Size of dataset — 123456 MB	
Data about natural phenomena	Number of Instances — 123456	
Data about places and objects	Number of Fields — 123456	
Synthetically generated data	Labelled Classes — 123456	
Data about systems or products and their behaviors	Number of Labels — 123456789	
Unknown	Average labels per instance — 123456	
Others*	Algorithmic Labels — 123456789	
(*please specify)	Human Labels — 123456789	
	Other — 123456	
	<write here>	

DESCRIPTIVE STATISTICS

| Telescopic Content ↑ | Periscopic Content ↑ | Microscopic Content ↑ |

Row 1

Row 2

Section

↑
Agent Information
Journey(s)

Meaningfully arrange your AIJs into a template for Data Card(s).

Lenses can sections or "blocks" in a data card. Scopes describe the questions in each block. Organize these so readers are able to navigate information reasonable as they read the Data Card from **top to bottom** (across blocks), **hierarchically** (from blocks to questions), and **horizontally** (across questions).

Best practices for assembling Data Card templates

You may need to re-organize your lenses & scopes for a more refined completion and reading experience. Here are some helpful strategies:

Combine scopes that pose similar or related questions, even if they were born from different lenses

Prefix straggler microscopes or periscopes with conditions that are reflected in their telescopes

Convert similar telescopes into answer choices that producers can select from; including "Others (please specify)

Reflect additional context from lenses and related scopes in the suggested structure of answers

Use a nested organization schema so it's easy to fill and navigate completed Data Cards. Consider these two experiences when creating catch-all rows

Add specificity to binary choices by being explicit about conditions, rather than offering Yes/No or True/False as potential choices

Your dataset

Publishers of the dataset

FUNCTIONAL CRITERIA	TELESCOPIC	PERISCOPIC	MICROSCOPIC
Include this when: ✍️ Data Card will be public	What's the name of the institution or organization who published it?	Which team, group, or group of individuals are responsible for publishing the dataset?	Who's (individual) primarily responsible for publishing the dataset?
		What kind of institution or organization is the publishing party?	What is the mission statement of the publishing party?
Include this when: ✍️ Publishers are outside USA		Where are publishers located?	
Include this when: ✍️ Publishers are not the creators of the dataset.		If the publishers are different from creators of the dataset, who are the creators?	How are creators involved with the dataset if different from publishers?
	Can publishers be contacted?	How and can publishers be contacted?	For what purpose should publishers be contacted?

✍️ Convert lens into a section title

15 min

FUNCTIONAL CRITERIA	TELESCOPIC	PERISCOPIC	MICROSCOPIC
Include this when: ✍️ Condition	✍️ Telescopic Question	✍️ Periscopic Question	✍️ Microscopic Question

Final checks

–

✔ Agreed on the top 2-5 agents of the Data Card

✔ Prioritized the Agent Information Journeys addressed by your Data Card (lenses)

✔ Produced specific questions and prompts to answer in your Data Card (scopes)

✔ **Created a Data Card 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 ↗