
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

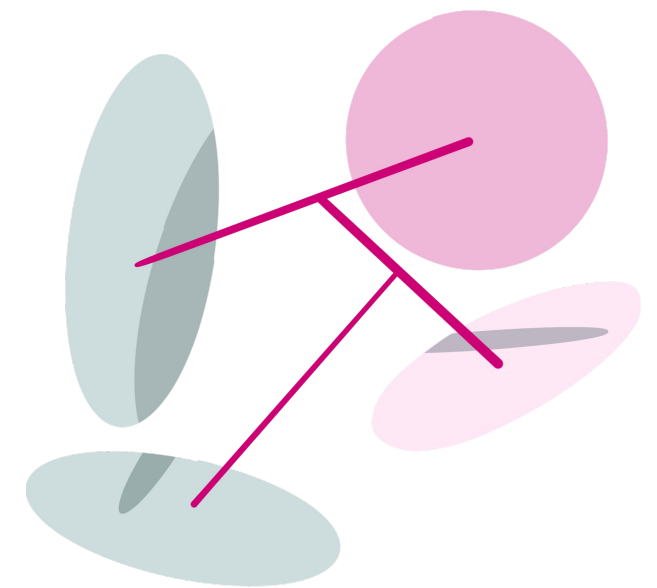
Introduction

01 Ask

02 Inspect

03 Answer

04 Audit





THE DATA CARDS PLAYBOOK

Map your Stakeholders

IN THIS SECTION

Create a map of stakeholders in your dataset(s) lifecycle. Use this to decide the audience of your Data Card(s).

INSTRUCTIONS

Use the Typology of Stakeholders to identify stakeholders who will produce and/or interact with your dataset(s)' documentation. Decide how they may might contribute to the creation of Data Cards(s), or use them.

OUTCOMES

Produce a map of stakeholders in your dataset's lifecycle(s) that highlights key touchpoints where they interact with documentation.

ACTIVITY LEVEL

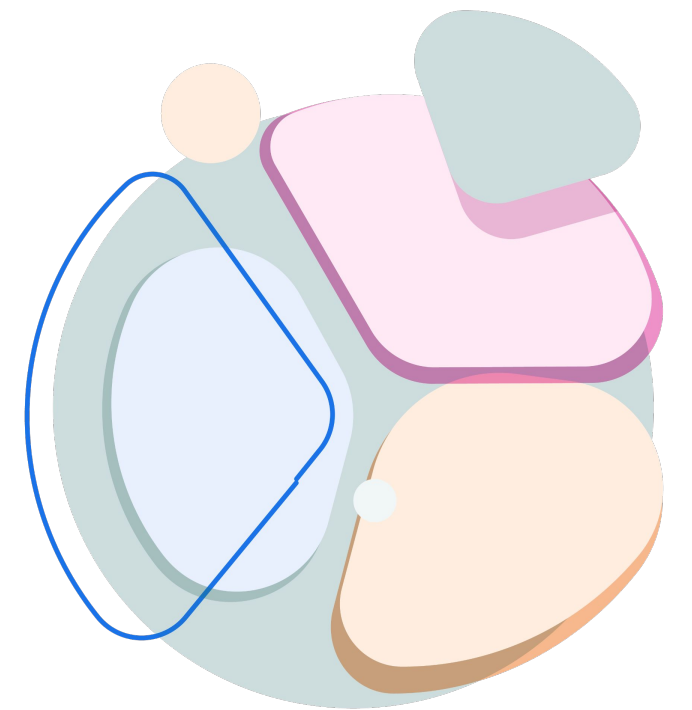
Basic

Stakeholders in a dataset's lifecycle can be bucketed into three categories

Producers = create datasets and/or documentation

Agents = use, evaluate, or determine how the dataset is or should be used

Users = contribute to data and represent demographics who are impacted by the way data is used



of datasets and documentation

use datasets and documentation

interact with products using data



Dataset Use

Pipeline Maintenance

Product Experience Data Contribution



Producers = create datasets and/or documentation

SOURCE

—

People who implicitly or explicitly **contribute data towards a dataset**. The people, behaviors, and cultures represented by a dataset.

Product Users, Data Contributors, Surveyed Population

CORE

—

The team of people responsible for **producing and publishing dataset(s)** and launch, adoption and/or success.

Researchers, Data Scientists, Software Engineers, Managers, Subject Matter Experts

ADJACENT

—

Individuals and groups **recruited to collect or label the data, provide advice on methods or interpretation**, at various points during the data lifecycle.

Surveyors, Raters, Labellers, Validators, 3rd Party Vendors, Domain Experts

IMPACTED

—

Current and future team members, partners, clients, or data-hosting platforms, responsible for **dataset maintenance or upkeep, deploying in production, monitoring**.


Domain Experts, Data Platform Owners, Data Aggregators

Who are Producers in your dataset(s)? Write who will **create** Data Card(s).

SOURCE

—

Who implicitly or explicitly **contributes data towards your dataset?**

 names, job roles, departments, affiliations, etc.

CORE

—

Who all are responsible for **producing, publishing and ensuring success of your dataset(s)?**



ADJACENT

—

Who all have been **recruited to produced data or advice on critical decisions?**



IMPACTED

—

Who are responsible for **dataset maintenance or upkeep, deploying in production, monitoring?**



**Agents = use,
evaluate, or
determine
how the dataset
is used or should
be used**

CORE

—

Industry and academic roles that use dataset(s) in their products, platforms, tools, and research.

Developers, Product Managers, Data Scientists, Creative Coders, Researchers, Teachers, Students

ADJACENT

—

Roles that don't use the dataset, but evaluate and make decisions that can directly affect the goals of the producers or core agents.

Industry Consultants, Policy Experts, Legal Entities, Investigative Journalists, Community Reps, Domain Experts

IMPACTED

—

Professional, expert-system, and domain expert roles whose work is affected by availability, updates, and removal of the data.

Domain Experts, Data Service Providers, Data Aggregators, Production Roles




Who are Agents in your dataset(s)? Write who will read and use Data Card(s).

Use this to determine which group(s) of data users might need quick or more detailed, technical descriptions about the data or the dataset to help them make decisions.

CORE

—

Who will use your dataset(s) in production, tooling and research?

 names, job roles, departments, affiliations, etc.

ADJACENT

—

Who will make critical decisions about the data but may not use it?



IMPACTED

—

Who will be affected by changes, updates, and removal of the data?



**Users =
contribute to
data and
represent
demographics
who are
impacted
by the way data
is used**

TYPICAL

— Individuals or cohorts of users of a product or service that uses the data, and have an as-expected or neutral experience.

Consumers of products, platforms, or services

IMPACTED

— Individuals or cohorts of end users of products and services who are significantly affected (positive or negative) as a result of the data being used in the product or service.

Users with extreme experiences, Non-profit organizations, Legal representatives

CONTRIBUTORS

— Users who produce or opt-in data in the product experience, which is then collected and turned into a dataset. In this case, these are often the same as source producers.

Users who opt-in data, People who operate machines that generate data, Research and Industry partners




Who are Users in your dataset lifecycle?

Use this to determine which group(s) of product users might need easy descriptions about the data or the dataset, or who might be at risk from inappropriate use of the dataset.

TYPICAL

–

Who are end-users who have a **normal or typical experience** of classes of products that use the data?

 names, job roles, departments, affiliations, etc.

IMPACTED

–

Who are end-users who have an **atypical (positive or negative) experience** of classes of products that use the data?













CONTRIBUTORS

–

Who are end-users who **produce or opt-in data in the product experience**, that is used to update the dataset(s)?



Producers, agents, & users in the dataset lifecycle relevant to Data Card(s)

<div>SOURCE PRODUCERS</div> <div> Who, role in Data Card</div>	<div>CORE PRODUCERS</div> <div></div>	<div>ADJACENT PRODUCERS</div> <div></div>	<div>IMPACTED PRODUCERS</div> <div></div>
<div>CORE AGENTS</div> <div></div>	<div>ADJACENT AGENTS</div> <div></div>		<div>IMPACTED AGENTS</div> <div></div>
<div>TYPICAL USERS</div> <div></div>	<div>IMPACTED USERS</div> <div></div>	<div>CONTRIBUTING USERS</div> <div></div>	



Who are most likely to fill out your Data Cards?

PRODUCERS

 Who, role in creating Data Card

AGENTS



USERS





Who are most likely to read your Data Cards?

PRODUCERS

 Who, use for Data Card

AGENTS

 Who, use for Data Card

USERS

 Who, use for Data Card

Summarize: Who are the target readers (agents) of your Data Card?

Top Priority:

1. 🖋️

Medium Priority:

2. 🖋️

3. 🖋️

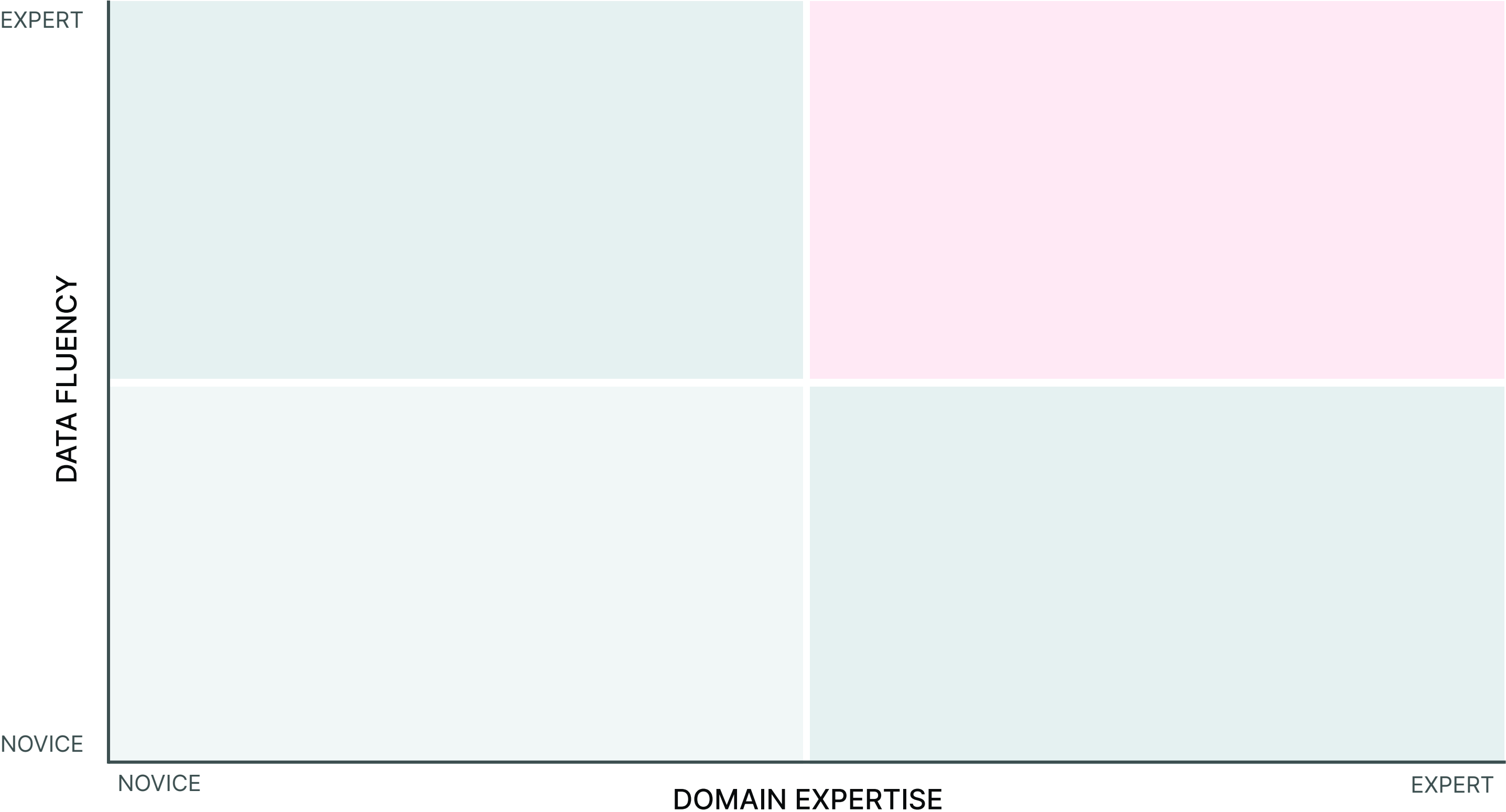
4. 🖋️

5. 🖋️

6. 🖋️



Agent Characteristics: Data Fluency and Domain Expertise



Checklist

YOU SHOULD NOW HAVE

—

- ✔ A map of stakeholders in your dataset's lifecycle
- ✔ Their role in creating Data Cards
- ✔ Their uses for Data Cards
- ✔ An intuition for downstream needs for dataset documentation



#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 ↗](https://github.com/pair-code/datacardsplaybook).

Find the complete Playbook at
[https://pair-code.github.io/datacardsplaybook ↗](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 ↗](#).