



---

THE DATA CARDS PLAYBOOK • RESOURCE

# Taxonomy of Stakeholders

Participatory activities for purposeful, transparent, and people-centric dataset documentation

---

[pair-code.github.io/datacardsplaybook ↗](https://pair-code.github.io/datacardsplaybook)

#datacardsplaybook



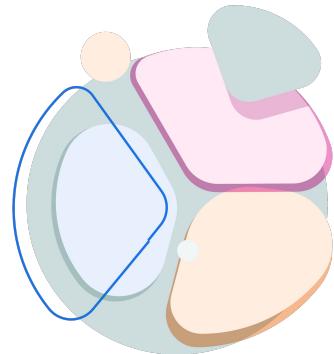
# A Taxonomy of Stakeholders

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

**Producers = create datasets and/or documentation**

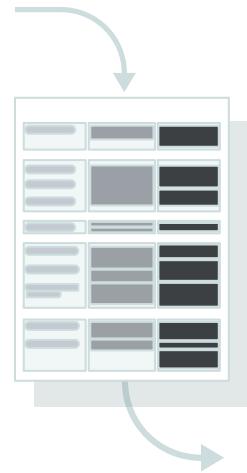
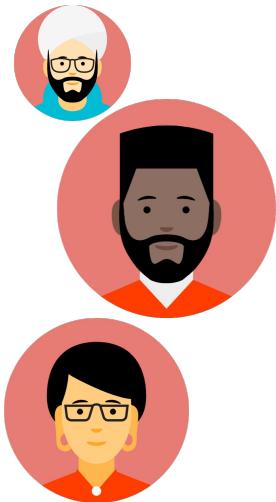
**Agents = use, evaluate, or determine how the dataset is used**

**Users = contribute to data and represent demographics who are impacted by use**



## PRODUCERS

of datasets and documentation



Dataset Ownership  
Launch and Maintenance

## AGENTS

use datasets and documentation



Dataset Use  
Pipeline Maintenance

## USERS

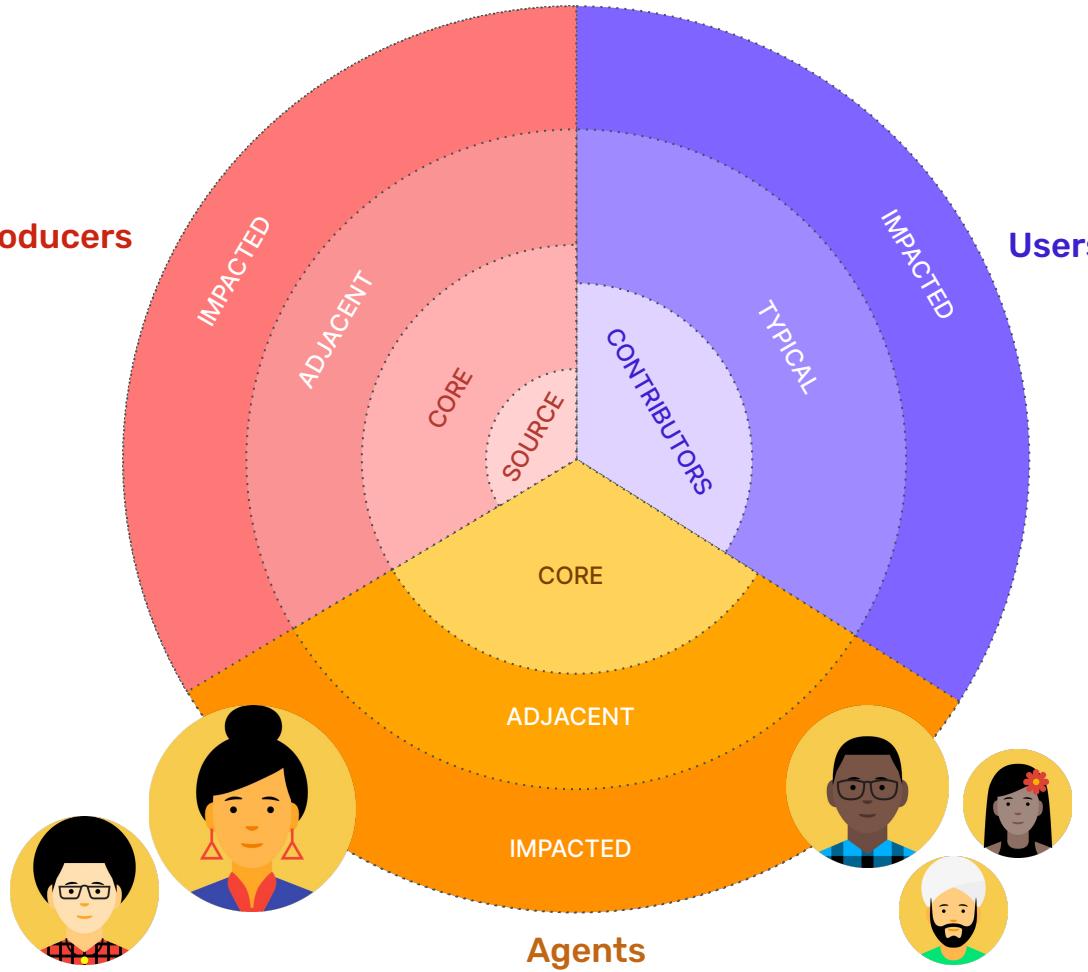
interact with products using data



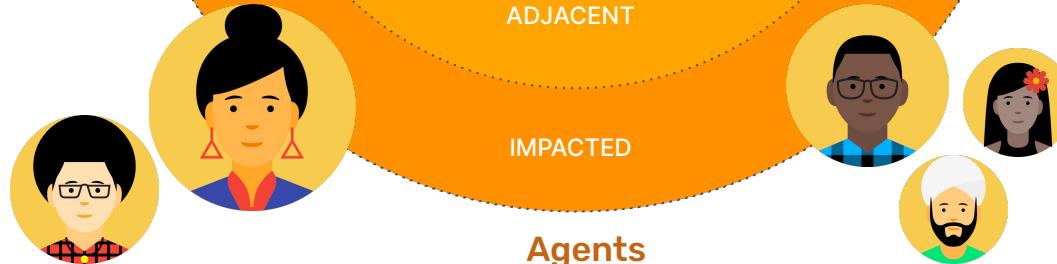
Product Experience  
Data Contribution



Producers



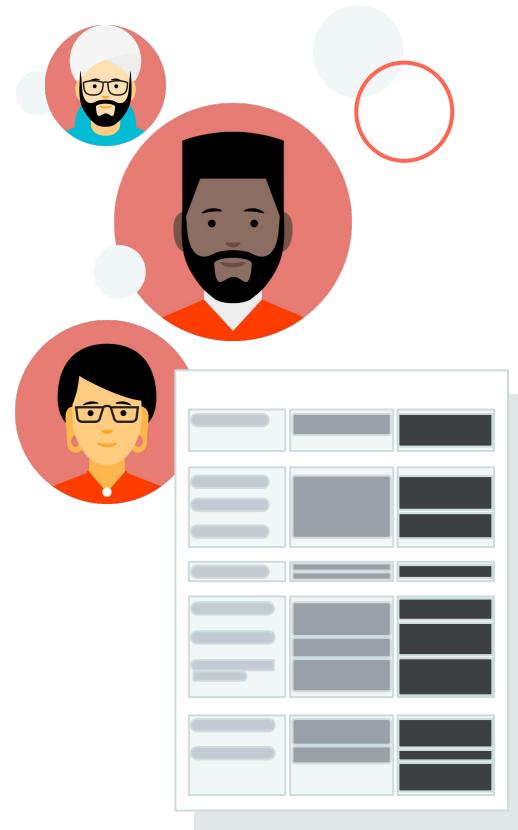
Users



Agents

# Producers = create datasets and/or documentation

- ▶ **Responsible for the dataset's** design, creation, quality testing, documentation, launch, adoption, follow-up maintenance, and future updates
- ▶ **Intended users** of this Playbook
- ▶ **Examples:** Researchers, Data Scientists and Analysts, Software Engineers, Product/Program Managers, etc.
- ▶ **Common tasks:** Dataset adoption, disclosure, future-proofing, fairness & security, improvements



# 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

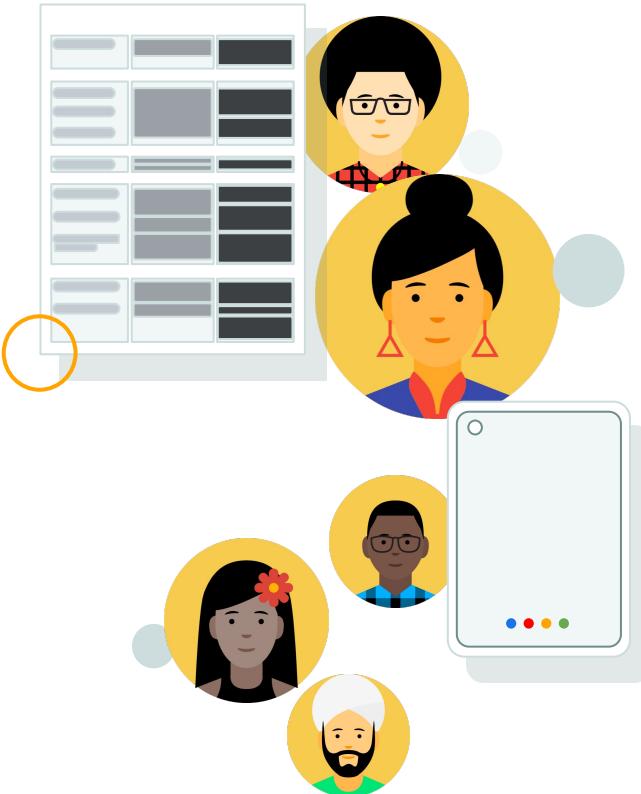
-

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

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

- ▶ **Producer's stakeholders** – people who will evaluate and use the dataset for their work, products, organizations, or communities
- ▶ **Key audience of the Data Card** – but they may or may not interact with the dataset itself
- ▶ **Examples:** Machine Learning or Product Engineers, Researchers, 3rd Party Vendors, Subject Matter Experts, Leadership or Managers
- ▶ **Common tasks:** Manage complexity, accountability, make trade-offs, deploy in production, archive



# Agents = use, evaluate, or determine how the dataset is 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

# Users = contribute to data and represent demographics who are impacted by use

- ▶ Interact with the **products, devices, and applications created by agents** using the producer's datasets
- ▶ **May choose to contribute their data** through products and provide helpful signals for producers and agents
- ▶ **Examples:** Data Contributors, Product Users, Representatives of user cohorts
- ▶ **Common tasks:** Use products, understand data/privacy, provide feedback, raise concerns

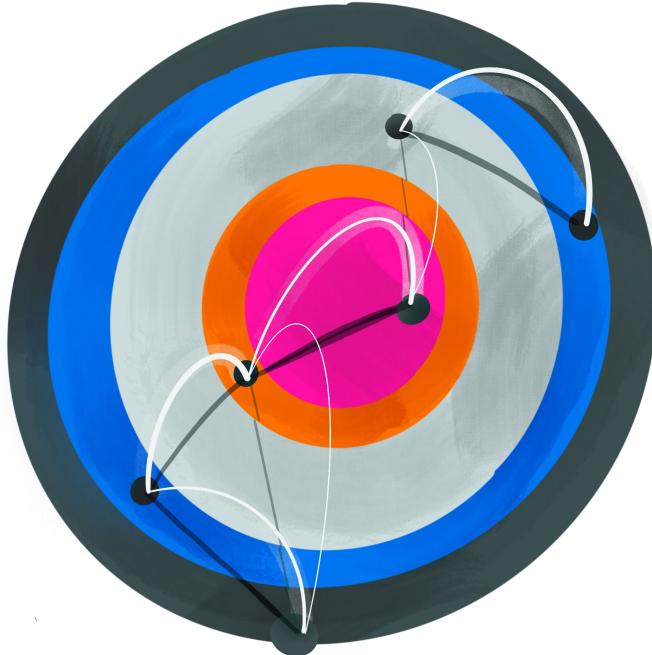


# Users = contribute to data and represent demograph- ics who are impacted by use

TYPICAL	IMPACTED	CONTRIBUTORS
<ul style="list-style-type: none"><li>- Individuals or cohorts of users of a product or service that uses the data, and have an as-expected or neutral experience.</li></ul> <p>Consumers of products, platforms, or services</p>	<ul style="list-style-type: none"><li>- 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.</li></ul> <p>Users with extreme experiences, Non-profit organizations, Legal representatives</p>	<ul style="list-style-type: none"><li>- 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.</li></ul> <p>Users who opt-in data, People who operate machines that generate data, Research and Industry partners</p>

**This taxonomy represents  
a continuum of constantly  
shifting needs and  
expectations from  
datasets and their  
documentation.**

**There is no one-size-fits  
-all solution.**





The [Data Cards Playbook ↗](#) by [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 ↗](#).