MODULE ANSWER | Foundation

Versioning & Changes

Data Cards are a product surface for your dataset.

The best Data Cards are ones that the dataset teams themselves can find useful, often created as a result of regular and continuous work that span both the dataset's lifecycle and the various needs of stakeholders involved. Across the lifecycle, individuals make decisions that affect the final shape and form of the dataset, which need to be systematically captured so readers can account for the "human factors" that are essential for their decisions.

Data Cards act as boundary objects which can be used by different kinds of stakeholders to build a shared understanding of the dataset, resulting in overall fewer assumptions. This can start as early as planning to create a dataset, when the data card acts as a requirements and design document. A Data Card does a lot of work – it aligns stakeholders, enables thoughtful collaboration, and can be used to verify contradictions across a dataset's lifecycle. So regardless of the number of documents it references or how many individuals contribute to it, the Data Card needs to first and foremost be an accurate source of truth.

We may sometimes want to create Data Cards about the same version of a dataset to target different audience groups – but having multiple data cards can cause them to easily fall out of sync without sufficient upkeep. It also needs to be clear to a reader where they can go for reliable information – and the more Data Cards for a given dataset, the more likely readers can get confused.

People + AI Research

Key Takeaways

- Ideally, a Data Card should be created concurrently with the dataset so it accurately describes the human decisions, rationales, and explanations that shape a dataset.

- A Data Card that is useful to both, the team that produces it and the individuals reading it is one that is consistently maintained and updated.

- A Data Card and its versions should be easily available to both upstream and downstream stakeholders in your dataset's lifecycle.

Actions For Your Team

1. **Start early.** Treat the Data Card as a document for design and technical requirements for the most comprehensive and useful Data Card possible. Don't wait to finishing creating the dataset – instead, start filling out your Data Card as early as possible. This reduces re-work, improves accuracy of the Data Card, and is an effective way to anticipate challenges in creating your dataset!

2. **One card, one version.** We recommend a single Data Card per version of the dataset (as your team defines "version"), as opposed to multiple Data Cards describing a single version of the dataset designed for different readers. Update this Data Card in the same way you would treat updates to your dataset.

3. **Tend to multiple readers with different abstraction levels.** Lean into the telescopic-periscopic-microscopic structure of the Data Card to embrace different levels of abstraction that readers from diverse backgrounds can find useful.

Considerations

→ Do you have a plan to create Data Cards for new versions of your dataset? What will happen to older versions of the dataset and their Data Cards (e.g., what is your deprecation or archive strategy)?

→ Are there audience groups that Data Cards will need to be adjusted for in the future, such as when your organization grows?

→ Is there a single place where readers from across the dataset lifecycle can go to find and compare your Data Cards?