MODULE AUDIT | Foundation

Evaluate with Readers

Evaluating your Data Card from the perspective of the reader

A reader uses a Data Card to make assessments about the dataset, so any evaluation of a dataset's Data Card should focus on whether or not the reader can successfully arrive at acceptable conclusions about the dataset.

A first principle is that the information in a Data Card must line up with a reader's experience of using the dataset. This directly impacts the reader's beliefs about the reliability and credibility of the dataset, and subsequently, the reputation and trust in the dataset's authors or publishers. Inversely, a reader's existing beliefs about your dataset, organization, and other datasets published by your organization can also influence how they engage with your Data Card. This is regardless of how discoverable, usable, or well-constructed a Data Card might be.

For example, readers with a positive experience of datasets that previously published by an organization might implicitly place more trust in a new dataset published by the same authors. In this case, there's a chance that the reader might make intuitive leaps and may not read the new Data Card closely enough to have the best possible understanding of the dataset – and specifically, how it differs from an older, similar dataset.

As such, evaluating a Data Card requires approaches that can assess if or not readers can arrive at acceptable conclusions of the dataset within their contexts. These are different from

People + AI Research

evaluating the dataset itself, and can provide different insights depending on how the evaluations are set up and when they are conducted. For example, a user study to understand if your content is understood by different readers can yield directional but actionable insights when you're iterating on your Data Card. After launch, measuring the adoption and efficacy of your data card through user satisfaction, surveys and analytics in the Data Card's implementation can provide insights about its relevance in practice. In that sense, a Data Card can be a useful probe to both drive and evaluate the success of your data set, and to paint a clearer picture of your downstream stakeholder needs.

There are a variety of stakeholders in a dataset's lifecycle, each with different levels of data fluency, domain expertise, and requirements. A requirement is a statement that identifies a product or process operational, functional, or design characteristic or constraint, which is unambiguous, testable or measurable, and necessary for product or process acceptability. (ISO/IEC 2007). The goals of your dataset, stakeholders in the dataset's lifecycle, and the implementation of your transparency efforts play a role in establishing the requirements and evaluation criteria of your Data Card.

For example, multiple product managers, engineers, data scientists, AI designers, and IRB reviewers might use answers in a Data Card. A good evaluation process will have criteria that will relate directly to the functional, operational, usability, and safety requirements for each of these roles.

- **Functional Requirements.** Does your Data Card enable readers to complete their task given their respective roles? *For example, consider a data engineer that is interested in integrating your dataset into their pipeline. Does your Data Card have the information required to successfully implement the infrastructure needed to use the dataset?*

- **Operational Requirements.** Does your Data Card enable readers to identify the essential capabilities, performance measures, and other associated requirements and processes necessary to use the data set effectively? *For example, a machine learning model builder that wants to fine tune a recommender system using your data set. Does your data card have enough information required to determine the constraints as well as the performance needs that must be met?*

- **Usability Requirements.** Can readers easily navigate and interact with your Data Card? Does the implementation of your Data Card meet basic usability heuristics and accessibility standards? *For example, consider a student researcher who wants to use your dataset, but they have limited access to the internet. What kinds of challenges might embedding an interactive, exploratory visualization of your dataset in your data card create? Or, what kinds of UI oversights might prevent a screen reader from translating the data card for a low vision reader? these interactive elements?*

- **Safety Requirements.** Is the information provided in the Data Card useful for practitioners to assess any potential undesirable outcomes associated with your dataset in their domains? *For example, consider machine learning practitioners working in a high risk domain such as healthcare. Does your data card describe the appropriate security, privacy, robustness and compliance requirements that necessarily need to be disclosed to prevent poor patient outcomes?*

Key Takeaways

- Different evaluation methods will yield different insights about the efficacy of a Data Card. Select evaluation methods that can be used throughout the transparency documentation process — from creation to launch and thereafter.

- When designing a study to evaluate your Data Card, consider the unique requirements of different stakeholders in the life cycle of the data set.

- The longevity of any Data Card depends on the reliability, credibility, and trust that can be established through the information in it, and the past reputation of the dataset publishers.

Actions For Your Team

1. **Validate your Data Card with intended readers.** Get your Data Card evaluated by individuals who represent your intended readers in the context of real-world tasks that they might perform. A heuristic evaluation will encourage your readers to articulate with

specificity what they find helpful in your Data Card, and what they don't. This can help you establish success criteria and identify steps to improve your Data Card. Take specific note of places where experts tend to gloss over important information.

2. **Test your Data Card with a low-proficiency agent**. Test your Data Card on a simplified task with a layperson audience or a reader with low levels of data fluency and domain expertise. A [contextual inquiry](#) that requires the participant to articulate their thoughts when interacting with the Data Card to perform the task is an easy way to identify where information is too dense or confusing.

3. **Assess the overall performance of your Data Card.** Use a method like [NASA's Task Load Index (NASA-TLX)](#) to measure the performance of your Data Card across multiple reader groups in the context of the tasks that they might perform. This is particularly useful if you publish multiple Data Cards and want to continuously improve your transparency effort.

Considerations

→ When selecting an evaluation method, what insights might you expect to learn and how might these contribute to the overall success of your Data Card?

→ How might you measure the impact of the Data Card on the perceived reliability, credibility, trustworthiness, and reputation of the dataset, similar datasets, or the publishing organizations?

→ What are the resources (for example, time, study material, and availability of participants) necessary to conduct a study? Can this study be successfully performed for multiple Data Cards?

→ Do evaluations of your Data Card account for study participants' existing beliefs (reliability, credibility, trustworthiness, and reputation) about similar datasets or the publishing organizations?