



1.1 Plan the work

A clear plan and distribution of work to complete the Data Card sets you up for success.

Datasets are often created or curated by teams over long periods of time. Information is distributed across many people and the documents they maintain.

This is especially true for large projects with complex upstream dependencies – where AI models or third-party vendors have been used to label data. Unsurprisingly, one of the top hindrances to completing Data Cards is fragmented and missing information. Before you start filling out your Data Card, take a few minutes to create a plan and organize your existing documentation as a team.

Key Takeaways

- Transparency is a magnet for complexity. While there's no such thing as a quick, easy or one-size-fits-all solution, approaching transparency in a thorough way can reduce knowledge asymmetries and reveal opportunities for more responsible dataset practices.
- With a little bit of planning and collaboration, you can prepare for missing documents, incomplete analyses, or outdated information about datasets.

Actions For Your Team

1. **Start with a joint review of the work ahead.** Creating a data card can be a complex project. Ask everyone on the project to review the transparency template together. Gather and consolidate knowledge about your dataset that may be stored in different places before you begin.
2. **Commit to accuracy across the dataset lifecycle.** Identify a fact-checker who can answer questions as the “source of truth” about your dataset when distributing work. Get reviewers who can verify the accuracy and authenticity of responses in your Data Card.
3. **Build in slack.** Account for slow communications, back and forths, and time to conduct additional analyses in your timelines as necessary.
4. **Engage experts.** In one of your earliest reviews of the template, identify any trusted experts and resources you will need to navigate complex or new areas, or to conduct additional research.
5. **Plan the distribution.** Identify specific formats or platforms (e.g., papers with code, a website, a repository) that allow your Data Card to live with your dataset and its versions, and how it will be shared as a stand-alone artifact.



Considerations

- Do you have access or open lines of communication to collaborators on the dataset project?
- Is it clear which questions are relevant to your dataset, and what resources (analysis, documents, people) are needed to answer them?
- Have you enlisted the help of lateral stakeholders or subject matter experts to help as necessary?
- Have you set up a clear timeline with milestones and multiple check-ins?



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 ↗](#).

