

# COLLECTING

## PURPOSE

The Collecting stage involves the responsible generation or compilation of data that will be used in the collaborative.

## Following Best Practices

**1. Proportional Data Collection:** Have you ensured that the amount of data collected or made accessible is the minimum necessary to address the problem or question?

- Key Stakeholders: Partner

**2. Respecting Rights & Dignity:** Have you ensured that data collection processes are respectful of data subjects' rights and conducted in a way that prioritizes their dignity?

- Key Stakeholders: Legal, Data Engineering, Data Subjects, Intended Beneficiaries

**3. Practicing Privacy-by-Design:** Have you taken a “privacy-by-design” approach, in which technical solutions provide security and reduce the risk of data exploitation?

- Key Stakeholders: Data Engineering

**4. Adhering to Regulations:** Have you ensured data collection adheres to relevant data protection regimes?

- Key Stakeholders: Legal, Data Engineering, Data Subjects, Intended Beneficiaries

**5. Obtaining Consent:** Have you obtained meaningful consent to collect data?

- Key Stakeholders: Legal, Data Subjects, Intended Beneficiaries

**6. Protecting Data Sources:** Have you ensured that any sensitive, personally or demographically identifiable data is protected?

- Key Stakeholders: Data Engineering

**7. Ensuring Congruence:** Have you taken steps to ensure that the eventual use (and reuse) of data aligns with data subjects' consent and expectations at the collecting stage?

- Key Stakeholders: Partner, Management, Data Engineering

**8. Continually Reviewing Collection Processes:** Are processes for data collection and consent upheld and continually reviewed throughout the data collaborative?

- Key Stakeholders: Data Engineering

### Evaluating Data Assets

**9. Constructing a Data Inventory:** Have you audited and inventoried datasets that could support the work and potentially negate the need for new data collection?

- Key Stakeholders: Data Science/Analytics

**10. Evaluating Data Accuracy:** Have you assessed the context in which data was collected to ensure that data is applicable to the current problem or situation?

- Key Stakeholders: Data Engineering

**11. Evaluating Data Accuracy:** Have you assessed the relevance, accuracy, and timeliness of collected data?

- Key Stakeholders: Data Science/Analytics

**12. Evaluating Data Completeness:** Have you assessed how complete and representative the data is in relation to the data collaborative's focus?

- Key Stakeholders: Partner, Data Engineering

**13. Evaluating Data Consistency:** Have you ensured that the data conforms to the syntax of its definition?

- Key Stakeholders: Partner, Data Engineering

**14. Evaluating Data Limitations and Biases:** Have you assessed limitations in the data, and engaged external experts to evaluate any unperceived data biases?

- Key Stakeholders: Partner, Data Science/Analytics

**15. Evaluating Data Timeliness:** Have you determined whether data collection is frequent and timely enough to inform effective analysis and decision-making?

- Key Stakeholders: Data Engineering

**16. Preventing Bad Data:** Have you mitigated risks of producing bad data, e.g., technological challenges and misconfigurations, variable norms or quality standards, legal confusion or gaps, and misaligned incentives or interests?

- Key Stakeholders: Partner, Data Engineering, External Expert

**17. Introducing Data Safeguards:** Have you introduced human oversight and technical safeguards to minimize the risk of transcription errors or data manipulation?

- Key Stakeholders: Data Science/Analytics

**18. Developing Documentation:** Have you implemented processes for recording data collection methods, unique features, historical events, omissions, biases, and metadata?

- Key Stakeholders: Data Engineering