How to Evaluate a New Data Source: Checklist

from We Dig Data

This checklist is from the **Practical Data Foundations** newsletter.

Interested in developing real-world data-savvy that will help accelerate your career growth?

Subscribe at wedigdata01.substack.com

STOP: Does the data's scope match internal needs? If not, now's the time to walk away.

Checklist: New Data Source Evaluation

@2025 We Dig Data

2. Data Frequency, History, and Change Goal: Determine whether the historical data and regular updates align with your needs. ☐ Does the data set come with time periods and trend data? ☐ How often is the data updated with the latest data? ☐ How often does the data provider send updated data to the client? ☐ How much historical data is available? ☐ How does the data provider handle revisions or market changes to historical data? To go-forward data? Ex: company or brand name changes, new product classifications, etc. **STOP:** If this doesn't line up with internal needs, check competitors or assess workarounds. 3. Data Collection Goal: Assess the stability and quality of the data source. ■ Where do you get the data? If the data provider collects or creates the data (like surveys, webscraping, or data input teams), ask: What is the collection process? How many surveys/items input/sites scraped per time period? Who does the collection for you? If the data provider acquires data assets from other companies (like geolocation, google trends, or app usage data): • How many data suppliers and data feeds feed this data set? Who are the suppliers and what is the business relationship? How long has the data provider been working with each supplier? • What does the data provider do when one supplier drops out and another joins? • What happens when a supplier delivers data late or sends the wrong data?

4. Data Cleaning and Processing

☐ What happens when there is a problem?

Goal: Understand the extent of the data provider's data handling and value-add activities.

☐ How does the data provider ensure quality in data collection? How does the data

provider check that data files received are complete and correct?

 Walk me through the process for cleaning and processing the data. What complace for quality assurance? What fields are calculated or 'interpreted'? What data manipulation or calculations are applied to the raw data? (balance demographics, project for missing parts of the landscape, etc) Is the data provider using AI or Machine Learning, and what is it for? Where a does the data provider monitor quality? How often? Changes and maintenance on the data: Does the data provider apply those chistorical data or just the go forward? If historical, how far back are the changes. 	e for and how hanges to
5. Logistics, Data Delivery, and Customer Support	
 After the time period closes, when is the data available to clients? How is data delivered to the client? File format? Are these negotiable? What is the data provider's process to notify clients of data delays or change For planned data enhancements or maintenance, how far in advance is the conotified? How does a client contact the data provider with questions or concerns with If a problem is identified, what is the data provider's target response time? Ty resolution time? 	lient the data?
Note: Build response times, delivery times, and notifications of delays or changes in contract.	to the
LAST STOP: Compare the data provider's answers to the the internal needs list. I data source still look like a fit?	Does this
Next Steps: Is there a good fit?	
f there is a match between the company's needs and the data provider's responses, ypical next steps:	these are
 Schedule a follow-up call within 1-2 weeks. Request test files, data documentation, and draft bid or contract terms. Bring in technical partners to discuss integration and testing (if you've not done). Who else do you need to involve in the organization? (Budget holder, users, proclegal, etc.) 	- ,