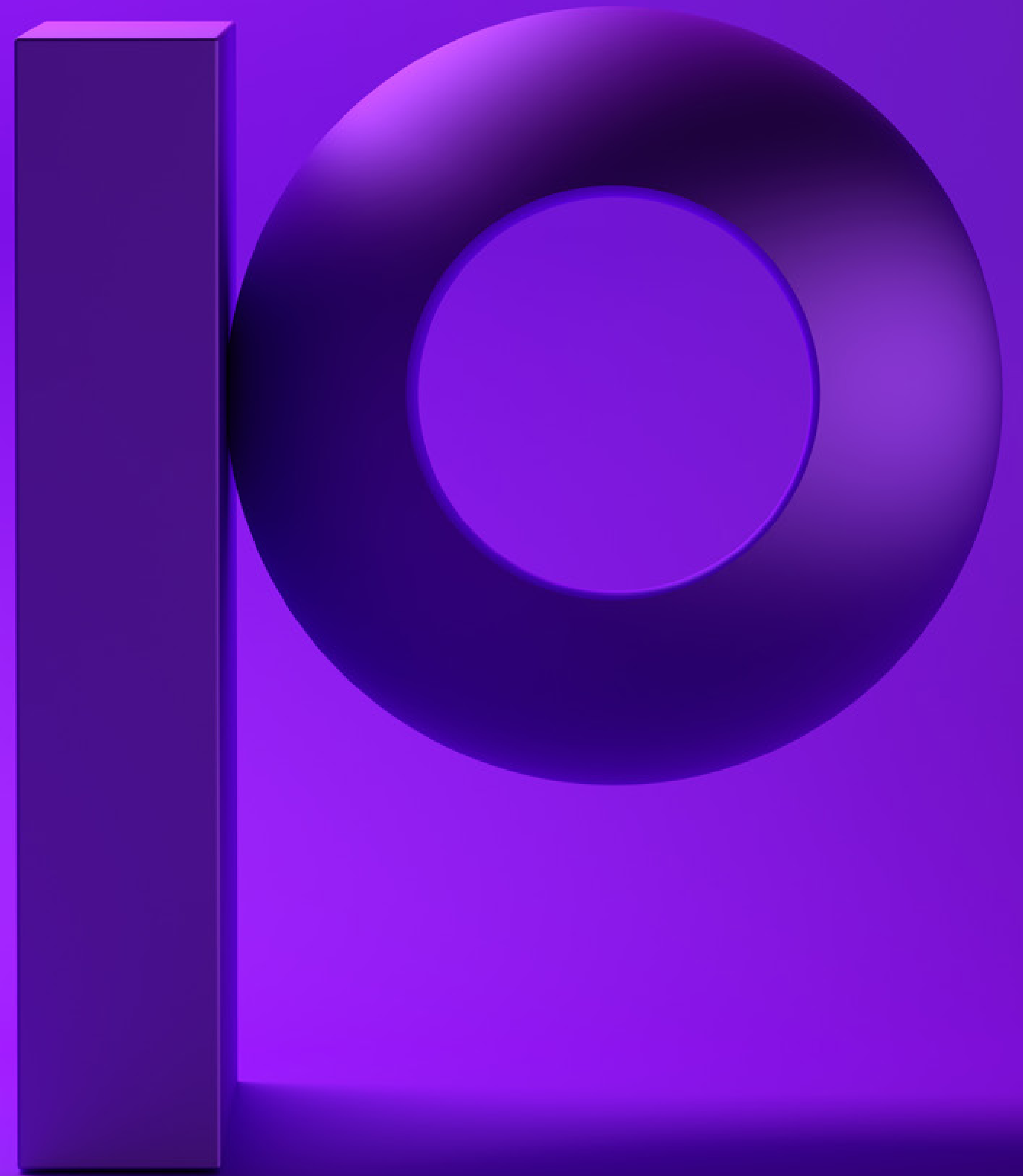precisely

5 things to consider when choosing a data provider

Spend less time prepping data, more time generating business insight

How much time does your company spend on data prep?

Your data professionals spend about 80% of their time finding, prepping, and managing data[1]. That leaves only about 20% of the workday for applying data to business operations: running models to illuminate risks and opportunities, creating efficiencies, enhancing the customer experience, and improving business outcomes.

What if you could flip those ratios?

What if your staff could easily select, incorporate, enrich, and interact with data — your own, and that developed by third parties? What perspectives would you gain? What business opportunities would you uncover?
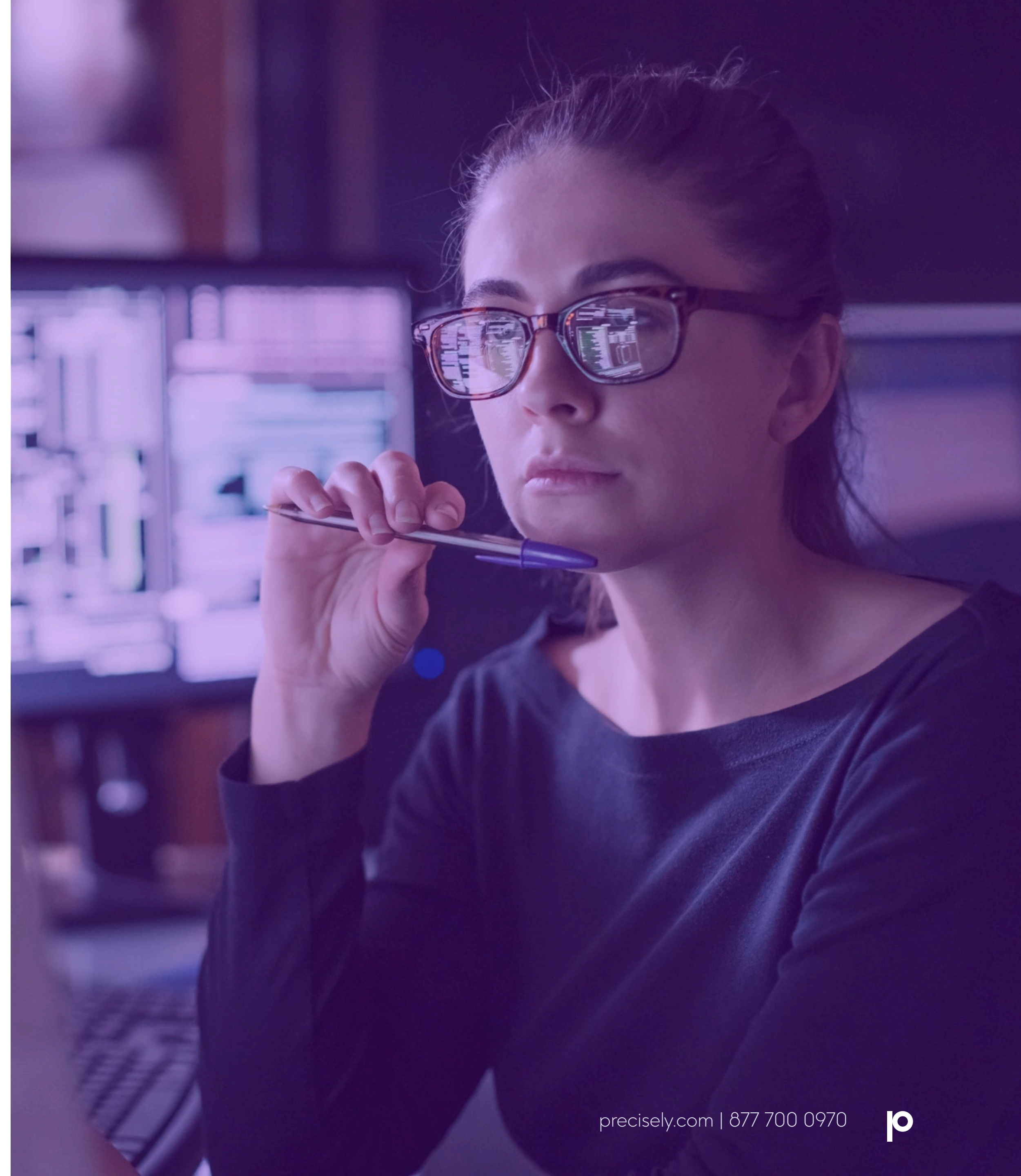
Consider working with a data partner

Obtaining ready access to trusted, easy-to-use data is a complicated process. It requires deep expertise in accessing data from siloed, back-end systems, standardizing and cleansing that data, and enhancing it with the type of context that can reveal relationships among people and places, assets, and opportunities. Businesses must also develop a thorough understanding of how to align data with operational needs to better govern data deployed throughout the organization.

This work requires significant time commitments and specialized skills not always available in-house. That's why many companies choose to work with a trusted third-party data provider.

What characteristics should you look for in a data provider? Here are Precisely's Top 5.

1. Match datasets to your desired outcomes

Invest in data that aligns with pre-determined goals and objectives. Some projects require a high level of accuracy. Others require data to be aggregated into broader categories or geographies.

"You need to understand datasets and their associated quality," says Patrick Mottram, Precisely senior director of product management. For example, companies working in the field of commercial insurance need to know if the depth and breadth of data they're using is sufficient for accurate underwriting. Mottram believes that "a consistent scoring framework is required to assess quality across the data portfolio."[2]

> ## "You need to understand datasets and their associated quality... A consistent scoring framework is required to assess quality across the data portfolio."
>
> — Patrick Mottram, Senior Director, Project Management, Precisely

Determine the level of exactitude needed to meet your business objectives, then find a data partner who can provide it.

This will take some vetting. Look for data suppliers that are transparent about:

- The amount and type of context the data provides
- How data is gathered and captured
- How the information presented in each dataset is derived or calculated
- How precision and accuracy are determined
- How data is sourced. (If your data partner sources data from an outside company, consider your partner's processes for vetting and managing those sources.)
- Whether your partner's data and processes align with your use cases

If a potential data partner fails to provide this information, probe harder. Is the information unavailable? Is it unknown? Or is your prospective data partner just unwilling to tell you? Answers to these questions will determine how confident you can be in the data provided. The data partner's inability to offer complete information about the data being sold may indicate that the data is not as valuable as the provider claims.

2. Provide quality data

Not all data is created for your intended purpose, and no official body sets standards for data quality. That's why, according to [Data Integrity Trends: Chief Data Officer Perspectives](#), 50 percent of companies are concerned with whether third-party data meets internal quality standards. Forty-nine percent worry about whether third-party data is updated regularly and consistently.[3] Therefore, when vetting data providers, it is essential to look for companies that meet your criteria for the following:

Coverage:

Examine the attributes of each dataset. Do they align with your needs? Is each data record as detailed as necessary to meet business goals? If your retail company, for example, plans to build a toy store outside of Little Rock, AR., can you get by with demographic information listing the number of households in a 10-mile radius? Or do you need specific information on the number of households with children ages 12 and under? A provider's ability to match the detail you need enables confident decision-making and may even save money for additional project areas.

Completeness:

Each data record can contains many fields. Consider the fill rate for each field within the dataset. The more fields that are blank or contain null values, the less valuable the data.

Correctness:

How accurate is the data your provider offers? Consider working with a provider that allows statistical sampling of large and vital datasets, enabling you to cross-check sampled data with authoritative information. This can help you determine the error rates of each dataset. Remember that not all types of data will have the same level of correctness in all geographies. Ask your provider to help you set expectations before you spend a lot of time testing data.

Currency:

Determine how frequently your provider updates data. Ask how long it takes for datasets to reflect real-world changes. Use an example based on your use case. For example, you may want to ask questions such as *"If a farmhouse on a 50-acre parcel is demolished, and a developer builds 200 townhouses on that site, how long does it take for updated information on this parcel to be reflected your dataset?"*

Consistency:

Ask your data provider questions such as "What is your on-time delivery rate?" "How standardized is the data format, and when was the last time it was changed?" "How interoperable is this dataset with our existing datasets and software?"

While evaluating data providers, remember that you may have to adjust your expectations for quality depending on the region examined. Location information in North America and Western Europe is far more mature than location information in vast swathes of Asia and Africa. Some countries lack contiguous house numbers. Some regions lack even physical addresses or postal codes. To overcome these barriers to insight, consider partnering with a data provider that offers the most accurate data possible now and actively works to improve data quality for those regions over time.

3. Make data easy to use

The less time you spend finding and prepping data, the quicker your insight and the faster your ROI. Consider working with a provider that offers readily accessible, easy-to-incorporate data. Datasets should be easy to get, easy to load, easy to join, and easy to understand.

And easy to update. Since the best datasets update frequently, consider working with a data provider with established processes for delivering high-quality updates in a reliable, repeatable fashion.

This is harder than it sounds. Eighty percent of the data executives surveyed in Data Integrity Trends say that consistently enriching data at scale is either "very challenging" or "quite challenging." Seventy percent find it hard to achieve a consistent view of data across multiple data formats.[4]

When choosing a data provider, look for automated data cleansing, integration, transformation, and deduplication processes. Invest in data files designed and built to work well together. These should be delivered as planned, consistent in content, and have a consistent layout. When vetting datasets for correctness, it is vital to ask for documentation and the ability to test data samples. Those samples should enable you to examine data content and format. Focus on interoperability and how well datasets align with your operations.

4. Make addresses even more meaningful by appending unique identifiers

Addresses act as linkage points for establishing location and connecting data. But addresses are complex and often provide an incomplete view of the location.

Different administrative bodies — from municipalities and developers to postal authorities — are responsible for various address components. (See Figure 1.) Any one of those entities may change part of the address name. A single old mansion may be divided into apartments. An individual condo owner may buy two units and combine them. New addresses arise as a result.

Furthermore, while a physical address is vital to postal delivery, it does not indicate where a property sits relative to other locations — flood plains, specific tax districts, or the nearest fire station.

Joining datasets using addresses as property identifiers can therefore be challenging. You may want to work with a data partner who uses unique, persistent identifiers to label properties. These identifiers can enable data stewards to amend thousands of data points to specific geolocations based on latitude and longitude. This process leads to a deeper understanding of each property. Think about it this way. It's good to know that there's a mixed-use building at 100 Main Street. It's even more informative to attach parcel, building footprints, building attributes, demographics, and socio-economic data to that building. Unique, persistent identifiers help in this effort.
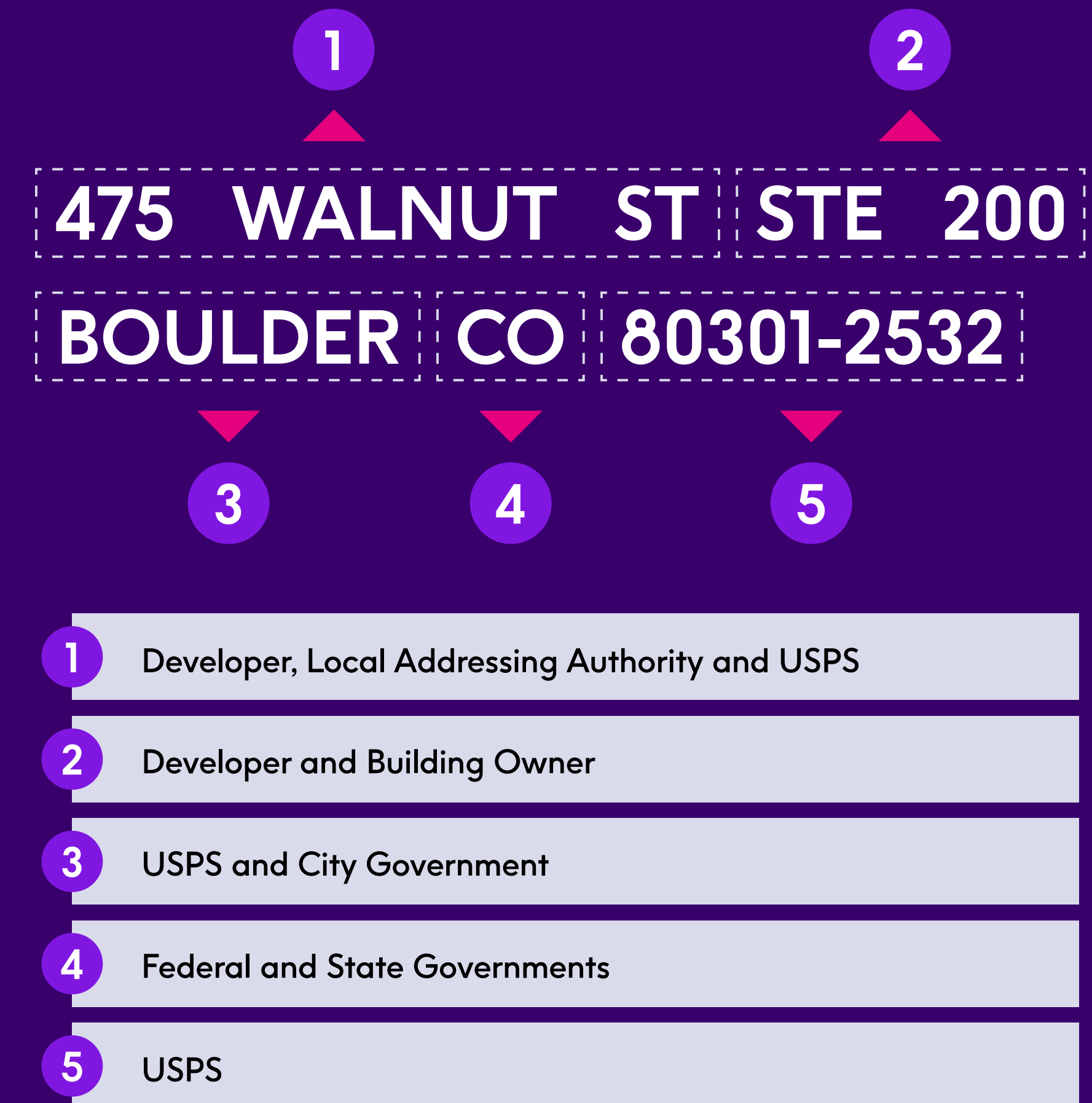
1 475 WALNUT ST 2 STE 200

BOULDER CO 80301-2532
3 4 5

1 Developer, Local Addressing Authority and USPS

2 Developer and Building Owner

3 USPS and City Government

4 Federal and State Governments

5 USPS

Figure 1: This figure illustrates the administrative groups who may control various portions of a street address.

5. Provide a variety of data-delivery options

Find a provider who offers different data delivery modes. These should align with your business needs. Remember that — to balance data-processing times, cloud costs, security requirements, and other factors — many organizations find it helpful to utilize more than one type of data delivery.

Data can be delivered via:

- **On-premises private clouds**

- **Enterprise public clouds**

- **Hybrid clouds:** These are environments that use both private clouds and enterprise public clouds.

- **Cloud APIs:** These are provider cloud environments used to store data. The provider gives customers a data-access point, the API, that allows the customer's application to submit requests for data to be transferred to their own systems. This model is best used to process a small number of requests and to answer pointed questions with up-to-date data.

- **Cloud software-as-a-service environments:** Snowflake, Databricks, and other SaaS environments provide massive amounts of computing power. This power enables users to process enormous datasets for big picture questions — everything from studying the market for new-product introductions to determining how climate change is likely to affect certain regions and parcels — and make decisions accordingly.

Why Precisely?

Precisely is the global leader in data integrity, providing accuracy and consistency in data for 12,000 customers — including 97 of the Fortune 100 — in more than 100 countries. Precisely's data integration, data quality, data governance, location intelligence, and data enrichment products power better business decisions to create better outcomes.

Building on more than 20 years of data-domain expertise, we provide all the capabilities discussed earlier in this eBook. Our products and services enable our customers to spend more time using data to improve business outcomes, and less time sourcing, preparing, quality checking, and updating information. We offer each delivery mode discussed above, including on-premises private clouds. This capability enables you to download bulk datasets directly to your on-site environment.

Since there are no set standards for data quality, we believe working with a data provider you trust is vital. We understand this need. We've experienced it ourselves. Our products are built on data from more than 100 suppliers. Our work evaluating these suppliers saves you the time and complexity of vetting them yourself. The result? Hundreds of interoperable datasets, including customized datasets, designed to meet your needs.

Creating linkages is an integral part of any data effort. For this, we offer PreciselyID. PreciselyID is a unique and persistent identifier, based on a property's latitude and longitude. It can enrich geolocations with more than 9,000 attributes. PreciselyID also allows companies to update, correct, and process records in bulk easily. This not only saves you processing time and expense but increases overall data value and accuracy.

Try our datasets yourself with Precisely Data Experience data-sampling system. Find, evaluate, and sample data for free. The Precisely Data Experience also allows you to explore mapping tools and access training resources.

Many companies now seek quality, easy-to-use datasets that align with business goals. Is yours one of them? If so, you need to partner with a data provider that empowers you to choose your own deployment methods. And you want unique identifiers to help you add context to data. There are many data providers on the market. Choose wisely. Choose Precisely.

Get started today. Learn more with our Data Guide.

1. The 80/20 Data Science Dilemma IDG/InfoWorld, 2021.
2. Commercial Lines Innovation Europe: Kickstart Digital Transformation for Commercial Lines. Panel discussion, 2021
3,4 Data Integrity Trends, Chief Data Officer Perspectives in 2021, Precisely and Corinium, 2021.

precisely

About Precisely

Precisely is the global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 97 of the Fortune 100. Precisely's data integration, data quality, data governance, location intelligence, and data enrichment products power better business decisions to create better outcomes. Learn more at www.precisely.com.

www.precisely.com