

An abstract graphic consisting of several thin, white, curved lines that originate from the left and curve downwards and to the right, ending near the center of the page.

The Data Leader's Guide to External Data Automation

How to accelerate speed to insights by
automating external data integration,
transformation, and observability.

cruxdata.com

Introduction

If you're reading this, the odds are that you're a leader facing some challenges with external data integration. You could be faced with a demand from your internal teams who need more data, expectations from your bosses or even the board with maintaining a competitive edge with external data, better support for strategic decision-making, or trying to keep up with ever-changing regulations.

The standard response to demand is to hire. More data engineers, sustainability directors, or security administrators could be on that list. But at a certain point, you either max out your hiring budget or run out of candidates to fill your pipeline—or maybe even both, but the demand for more data doesn't stop. So what do you do?

Automate.

Automation is a buzzword thrown around constantly, regardless of your industry, as a solution to any problem. But when it comes to external data integration, what should and shouldn't be automated is vital so that your current teams' bandwidth is expanded, their workload is lightened, and their daily tasks are augmented. The goal is to give them back more time to spend on value-add tasks for your company and remove some of the tedious, manual work they do to get there. Unfortunately, this isn't as simple as manipulating internal data, where you can drop a new field into Salesforce and begin collecting that information immediately.

Automating External Data Transformation

Data transformation is the biggest roadblock between raw data and analytic insights. Having the data accessible within your infrastructure isn't enough—it has to be manipulated into a readable and helpful structure. External data isn't meant to answer questions directly but to provide stepping stones toward that solution. It's up to you to combine, format, fill, calculate, and otherwise reshape your data to meet your needs because the default format is whatever the supplier feels like—no rhyme or reason required.

Automating these three key pieces of the external data integration puzzle can transform your organization's experience with external data. These three categories are uniquely challenging, so let's explore how and what you can automate each step of the way to make a difference.

Tips for Automation

Automating External Data Integration

There are two categories of integration for external data—inbound and outbound. Inbound integrations are links directly to your data source. These datasets can come in all shapes and sizes, from an API to SFTP to even a plain email. Your external data infrastructure has to be able to ingest data from any of these types of sources, and the variation adds an additional challenge to the equation.

Outbound integrations are usually to a cloud provider or other storage method. Once data is in your possession, it must be delivered to its storage place, which will likely require formatting and delivery adjustments.

Creating both of these integrations can quickly consume hours, weeks, and even months your data engineers could spend on more critical tasks. Setting up an automation process that delivers datasets from their source to your preferred cloud storage creates more bandwidth for your current team without adding to your headcount.

Additionally, security must be a key consideration for automating external data integrations. The external data you're consuming likely contains sensitive information, and each new dataset introduces a new risk or entry point into your infrastructure. Automating your integrations can reduce your attack surface and human error, making your external data pipeline safer for your business.

Automating External Data Transformation

Data transformation is the biggest roadblock between raw data and analytic insights. Having the data accessible within your infrastructure isn't enough—it has to be manipulated into a readable and helpful structure. External data isn't meant to answer questions directly but to provide stepping stones toward that solution. It's up to you to combine, format, fill,

calculate, and otherwise reshape your data to meet your needs because the default format is whatever the supplier feels like—no rhyme or reason required.

Automating External Data Observability

Automating external data monitoring and observability is like having a smoke detector in your house. The siren should go off when there's just a tiny amount of smoke instead of not noticing until your house is engulfed in flames. By then, it's too late to do anything but damage control, and monitoring external data is conceptually the same.

Monitoring your data pipeline manually—instead of automating it—can also increase the likelihood of human error in your external data process. By relying on people to spot and identify minute data errors—sometimes as tedious as an extra space—you increase your risk level exponentially. Another argument for automating this portion of your process is the downstream consequences unnoticed changes can cause. Data teams are built of talented individuals who are required to drop everything to diagnose what the problem is and where it is coming from and then fix the issues as they arise. Because this type of monitoring is often manual and isn't the priority, bad data can slip through the cracks and create an endless series of fire drills. Bad data isn't as nefarious as it sounds, but its consequences are. Something as small as a one-letter typo, a schema change, missing information, or an incorrect classification for a stock can set off a ripple effect. For example, an error in a pricing file can materially impact performance and attribution reports for internal and external clients. With many manual processes involved in the production of dashboards, portfolio reconciliations, and client portfolio performance and attribution, the risk of errors is severe. It can result in monetary or job loss, regulatory fines, and significant reputational damage to institutional client relationships.

Is There Anything I Can't Automate?

The short answer is yes.

The goal of automating the integration, transformation, and observability isn't to remove humans from the process but rather to refocus them on what truly needs a human touch. By augmenting the external data integration process, data scientists can reallocate their time to actual data science instead of ongoing maintenance and constant troubleshooting.

The external data landscape is so complex because today, no one singular standard or regulation governs external data. Automating the tedious transformation pieces of it integration gives your experts more space to understand and apply that data to your specific business needs.

Getting Started

Next Steps

If you're ready to begin automating your external data integration process, there are a few options for getting started.

Option one is to complete an external data assessment with an expert from Crux. During this process, we'll evaluate your organization's maturity level and provide you with concrete next steps for leveling up and maintaining your competitive advantage.

Option two is to begin exploring datasets to get started. For ESG specifically as an example, you can go two routes—focus on controversies or ratings. While there is no magic bullet for understanding ESG data, picking one of these options will give you a good starting point. Is it more beneficial to your company to screen out potential investments or focus on selecting those that rank the highest?

Option three is to dive into more resources and continue your exploration of external data resources before continuing forward. If you're looking for more on ESG specifically, here are some great places to start:

- This blog post & YouTube video on [10 Questions & Answers About ESG with Daniac Koedih, MSCI](#)
- This video on [The Past, Present & Future of ESG](#)—a previous event recording with a vendor management point of view
- [This article](#) on ESG Reporting Standards from Data Center Dynamics
- [This article](#) on the speed of Global ESG Standard-setting, or [Tips for CFOs Advancing ESG](#), both from CFO Dive
- [This piece](#) from PwC on thoughts from executives on ESG reporting

It's not a matter of if you should automate external data integration, but when. Everyone is swamped with data in today's market, especially as ESG data remains at the forefront of investing and trending topics. Whether you build automation from scratch or invest in an enterprise solution that does the work for you, external data isn't slowing down.