

A photograph of a diverse group of people in a meeting. In the foreground, a woman with curly hair and glasses, wearing a blue and white striped shirt, is smiling and looking towards the right. In the background, a man with a beard and short hair, wearing a dark blue sweater, is also smiling and looking towards the camera. They are seated around a dark table with a laptop and a pen visible.

# THE GLYNT GUIDE TO Accurate Sustainability Data

CONTACT: MARTHA AMRAM, PH.D  
MARTHA@GLYNT.AI



# The GLYNT Guide to Accurate Sustainability Data

In the era of voluntary climate reporting, spreadsheets were the tool of choice. Teams would collect the original data, type it into spreadsheets, review and correct and then post. In that era, accuracy was a key challenge. In fact, most companies spent more time and money on data quality checks than on data entry.

As we enter the era of mandatory reporting, audits and integrated financial-sustainability reporting, the concern about accuracy continues. The new systems must prove that they are easy to adopt, highly accurate and have a method to remain so.

GLYNT has been tackling this challenge for years. We built our system using best practices and technology solutions from AI, workflow automation and modern data infrastructure. We studied how high pressure high accuracy environments – such as stock trading – keep their data accurate. And we got audited ourselves. From these multiple perspectives, we have developed a system to deliver accurate data.

This guide provides a short overview of how GLYNT produces accurate data. We're happy to go into more detail. Customers benefit from GLYNT's hard work in three ways:

- **Accurate, Validated Data** – The data we send you has been validated many times within the GLYNT system
- **Certificate from GLYNT** – We've been audited and will share our certification. This reduces your audit burden and costs
- **Operational Savings** – GLYNT saves up to 80% over in-house systems because our costs of compliance are spread out over all customers

# What is Accurate Data?

Certain types of data errors were all too frequent in the era of voluntary reporting and sustainability teams built experience in what to look for and how to correct these errors. But in our new era, the definition of accuracy should be broader. Without this change, new errors will crop up as sustainability data is used in new ways. Here are three definitions of accuracy that meet this need.

**Accurate Extracted Data:** Most of the original data sources provided to GLYNT are PDFs, so the first question is *“If a 3 is printed on the PDF, is it captured as a 3?”* Experienced data teams know that 3 can be captured as an 8 on occasion, 0 can be captured as O and so on. So ensuring that data is captured as printed is the first measure of accuracy.

**Accurate Prepared Data:** Prepared data is a structured data file, in which all the extracted data has been aggregated and harmonized, regardless of the original source. A set of business logic is used, such as *“If energy use is printed for two meters but there is no bill-level total, sum the usage at the meter level to construct the bill-level total.”* In manual systems, humans apply these rules and they make errors. So the second measure of accuracy is about preparing data correctly.

**Accurate Report-Ready Data:** When data is reported for use by others, the standard criteria from finance is *“Is this data decision-useful?”* Sustainability audits also use this definition, and it more broadly captures what data users want to see. This broader criteria is an important step forward, as it recognizes that sustainability data now drives decisions, and bad data should not lead to decision errors.

- **Accuracy** – The data we send you has been validated many times within the GLYNT system
- **Completeness** – We’ve been audited and will share our certification. This reduces your audit burden and costs
- **In-Context** – GLYNT saves up to 80% over in-house systems because our costs of compliance are spread out over all customers
- **Unbiased** – GLYNT saves up to 80% over in-house systems because our costs of compliance are spread out over all customers

As you can see, the definition of decision-useful data incorporates the other two measures of accuracy, and extends it further. We built the GLYNT data preparation system to this highest standard.



# The Five Components of GLYNT's Data Preparation System

GLYNT had two goals when we set up our data preparation system. First, to deliver decision-useful data. We want to deliver sustainability data that is as rigorously produced as financial data. Second, we wanted to demonstrate the attributes of our system to auditors and customers. Here's a visual overview of our five components:

On the left are the enterprise systems where the original source data – aka Primary Data – resides. On the right are the enterprise systems that ingest the prepared sustainability data. This can include ESG, carbon accounting and sustainability reporting software. Once licensed and deployed, they are part of the enterprise systems.

So the key task is to pull source data from enterprise systems and to put prepared data into enterprise systems – while maintaining accuracy. Whether using an in-house system or GLYNT, the flow and the requirements are the same.

## The Five Components of the GLYNT Sustainability Data Preparation System



## The Five Components of GLYNT's Data Preparation System Cont.

Here is more detail on the five key components of the GLYNT data preparation system:

- 1) Validation at the Start.** As source data flows into GLYNT, we check for completeness, and duplicates. We check for data that is out-of-scope. And we check to ensure that each data file submitted has enough information to be processed. We'll alert the customer about omissions and issues. GLYNT accepts PDFs, scans, XLS, CSV, JSON and other file formats. Excellent data handling from the start sets the stage for accurate data results at the end.
- 2) Validation at the End.** Before the data is sent, we check to ensure it is accurate, complete, and in-context. For example, the definition of "energy cost" may vary. Did GLYNT prepare the energy cost data field in the prescribed manner? Are all data files submitted accounted for? These checks are implemented on every file sent.
- 3) Data Flows.** Once a file is submitted to GLYNT it is "fingerprinted" and stamped with a unique ID. Thereafter the file and all data spawned from that file are tracked, delivering data lineage and chain of custody results. GLYNT knows where every data file is in the system at every moment. For PDFs, which are the vast majority of data sources for sustainability, GLYNT uses Optical Character Recognition (OCR) technology to pull the data off the printed page. We check for accuracy at this moment, and if a data item fails our test, it is sent for a human review. GLYNT data is 98% accurate at the extraction stage. We monitor the corrections that the human makes, and improve our accuracy by automating fixes, preventing these errors in the future. We then add customer-provided data (such as site codes) and prepare the data into the final format. We have error trapping at each stage, see our discussion of KPIs below.
- 4) Monitoring.** Data preparation is a complicated system, with several points of internal coordination and moving parts. It takes a system to monitor the system! For example, GLYNT has monitoring on our data ingestion technologies: Are they up and running? Are the volumes flowing through as expected? This layer of technology gives GLYNT a real-time metric of the pulse of activity. A second area of monitoring is user access and change logs from user activity. This is key if humans are touching data. If you build an in-house system, you'll need the same thing. There is no shortcut.
- 5) Governance.** A key lesson from privacy and security compliance is that technology alone can't deliver the desired outcomes. It is always "People, Process and Technology." People have assigned tasks and a documented method of how to do them. But they need oversight, which is provided by a system of governance. GLYNT has standing oversight at the operations level, at the customer level, at the senior management level and at the board level. Each layer provides a different perspective and asks different questions of the system. Multi-layer governance is a method to challenge ourselves internally, on behalf of our customers.

# KPIs and More

A good data preparation system has clear goals, such as “Process 2 million data sources per day with 100% accuracy.” Key Performance Indicators (KPIs) measure progress to that goal. One set of metrics might be: Data sources processed per day; and available capacity per day.

Anyone who has been through SOX compliance in the US or COSO compliance outside of the US knows how helpful the layers of measurement and monitoring are. GLYNT has adopted these same methods in building our system. Working up from the bottom layer:

- **Monitoring** – The data preparation system must be instrumented so that a stream of activity data is sent to the central repository.
- **Controls** – Is the activity happening as expected? There are two ways to see this:
  - Preventative Controls: As the data is processed, put in a check on accuracy. One example: “*Field X is required. Was Field X extracted by the OCR?*” If it is not, an alert is triggered.
  - Detective Controls: Just after the data is processed, look back at the activity and determine if the outcome is as expected. An example: “*All required fields were checked and the number of alerts is low.*” If the alerts are above a certain level, an alert on the rate of alerts is triggered

We’re big fans of the two types of controls as they are two sets of eyes, different ways of thinking. It is these layers of checks – with different perspectives and multi-layer thinking – that deliver consistently accurate data.

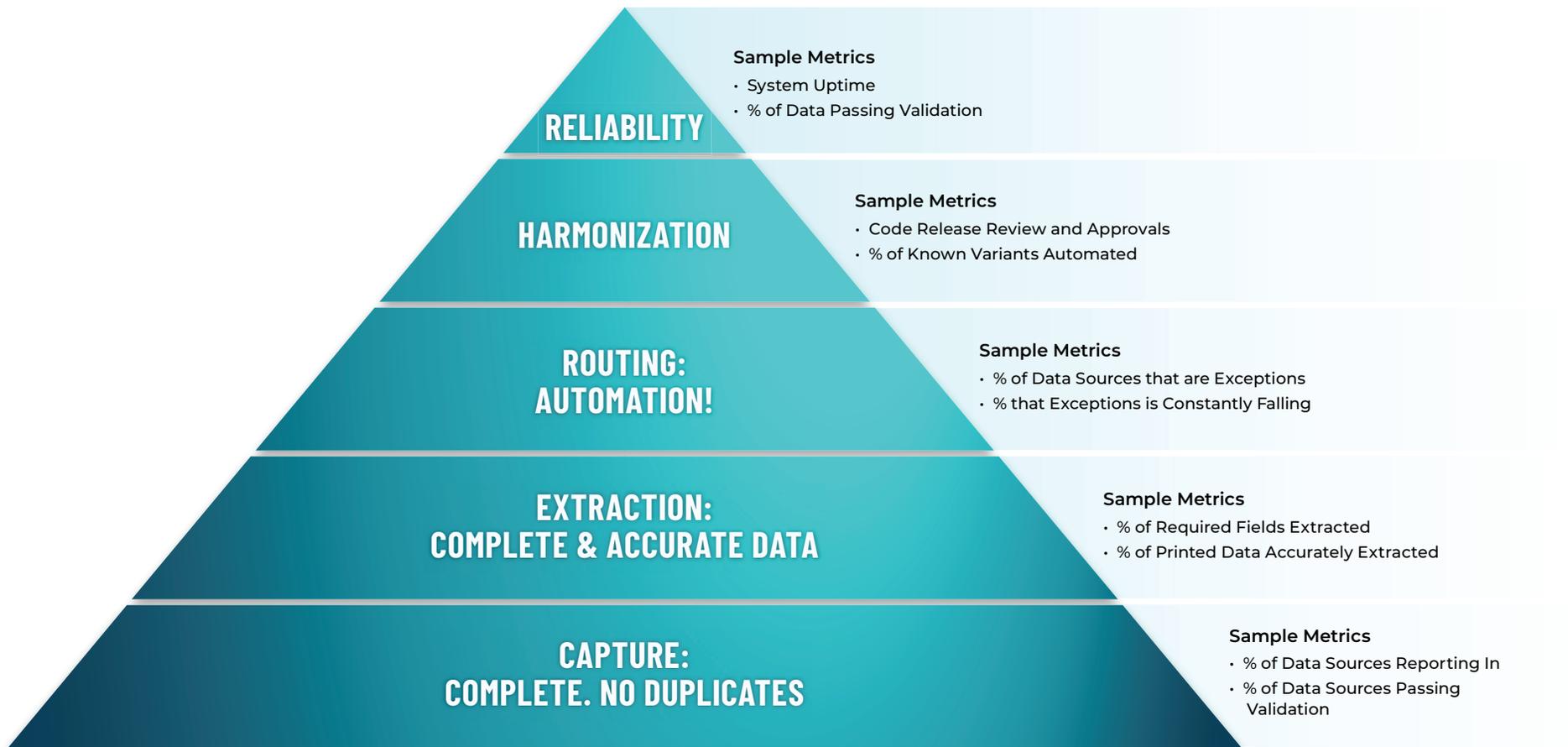
- **Key Performance Metrics (KPIs)** – The monitoring and control data feed into the KPIs. Following our example, we might have a KPI of “*Percent of all required fields successfully extracted.*” This is just one measure of performance that contributes to accurate and decision-useful sustainability data. GLYNT has built out a suite of these metrics and together they give us a real-time sense of how our system is performing on the metrics that reflect our goals.

As you can see, GLYNT has set up a sustainability data preparation system that is closely aligned with modern financial data standards. We have harvested the 20 plus years of lessons learned and best practices for SOX and COSO reporting. We have incorporated the audit criteria for sustainability reporting from the International Audit and Assurance Standards Board. And we dug in and executed. This takes years of work across people, process and technology. GLYNT data is accurate because we did this hard work.

# The Pyramid of Accuracy

Another way to view the system of data preparation is through accuracy measures at each level. We call this the Pyramid of Accuracy. It is not enough to check the data read by the OCR for accuracy, the data preparation system must start before that moment to ensure that the data is complete, and the data system must continue after that moment to check on accuracy in multiple ways.

## The Pyramid of Accuracy



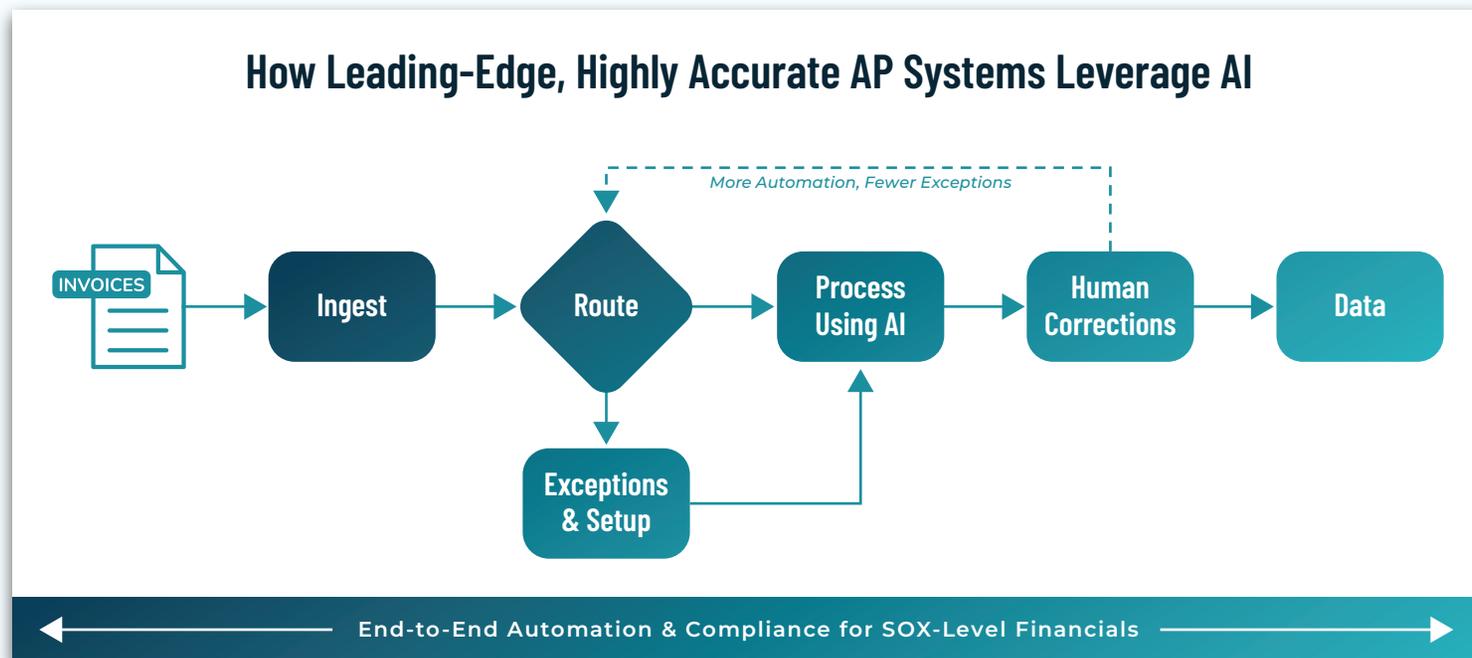
As the pyramid illustrates, when companies come off of manual systems or systems that use legacy OCR technology, they tend to focus on the second layer only. Best practices – and audit-readiness – requires multiple checks and multiple perspectives.

# AI, Accuracy and Privacy

The best-of-class Accounts Payable (AP) processing systems – the systems that process business invoices – use AI to produce more automated and accurate results. The push to automation from cost reductions: People are more expensive than technology and make more errors. So every time a human touches an invoice, it is not a good thing!

Automation also delivers much more accurate data. Humans doing data entry have accuracy rates in the 60–75% range. With data validation, double-teaming and so on, this rate can be brought up to 85–90%. But then a layer of quality control must be applied to identify and manage the remaining errors. In contrast, leading-edge AP systems are 95% accurate from technology and automation, and have a feedback loop that constantly improves accuracy.

GLYNT has brought this leading-edge AI to sustainability data. Here's a picture of the automated workflow.



## AI, Accuracy and Privacy Cont.

It's a straightforward flow, and the key is to correctly route incoming files to the AI and to constantly improve by harvesting the monitoring on human corrections. The Process Using AI box contains two key technologies: OCR and GLYNT's AI.

GLYNT's sustainability data preparations system has the exact same workflow. But, we don't use the same AI as leading AP systems for a simple reason: Lack of training data. In the world of business invoices, it takes 500,000 to a million invoices to train the AI. The large sample of data is used to show the AI what is "accurate and correct" data. Then the AI learns what to expect and can automatically find this data from the same or nearly the same invoices in the future.

To make the system work, a single large interconnected library of AI models is built. Every night or every week, the model is updated as new corrections data is available. It takes a lot of computing power, and accuracy improvements take a lot of new data and corrections. Leading-edge AP systems have harvested and operationalized AI tools from 2-3 years ago, not Gen AI. It is not yet clear how Gen AI can improve accuracy from invoice processing.

GLYNT's AI works a bit differently. No customer can or is willing to provide 500,000 sample utility bills. No customer wants their own data to be used for the benefit of other GLYNT customers. So GLYNT had to solve the training data problem in a different way. We built an AI system known as "Few Shot Machine Learning." It only takes 3 samples of an invoice, then GLYNT is automated and produces data at 98% accuracy. The human review is minimized and constantly decreasing.

This also means that each customer can provide all the training data GLYNT needs from the data they want processed. We segment our data flows, and each is customer-specific (GLYNT is a multi-tenant system). GLYNT can start with the data at hand, no need to gather extra data samples to train our AI.

And GLYNT's AI is no code and on demand. Our data operations and onboarding staff can configure the customer-specific AI as needed. If we have seen a data source before – say a specific utility and rate class – we are even more efficient in our set up. But we always retain the segmented customer data flows. No customer data is co-mingled.

GLYNT does not use Gen AI at this time. Testing by GLYNT and our customers has found that it is not useful for highly accurate data preparation. We'll investigate again, and keep our customers abreast of what we learn as Gen AI and Large Language Models (LLM) evolve.

Gen AI and LLM need boatloads of training data. Most enterprise databases are fairly small in this light. And they often contain incorrect data, which can make the Gen AI quite inaccurate (e.g. hallucinate). GLYNT's perspective that sustainability data is a key new flow of highly-accurate data that is operational in nature. As one customer said, "*Now I've got an ocean of data from GLYNT. I need help in using it!*" This is where emerging Gen AI applications will be useful. We wrote an article for Forbes about this, see the Resources section at the end.

# End-to-End Audit-Ready Sustainability Data

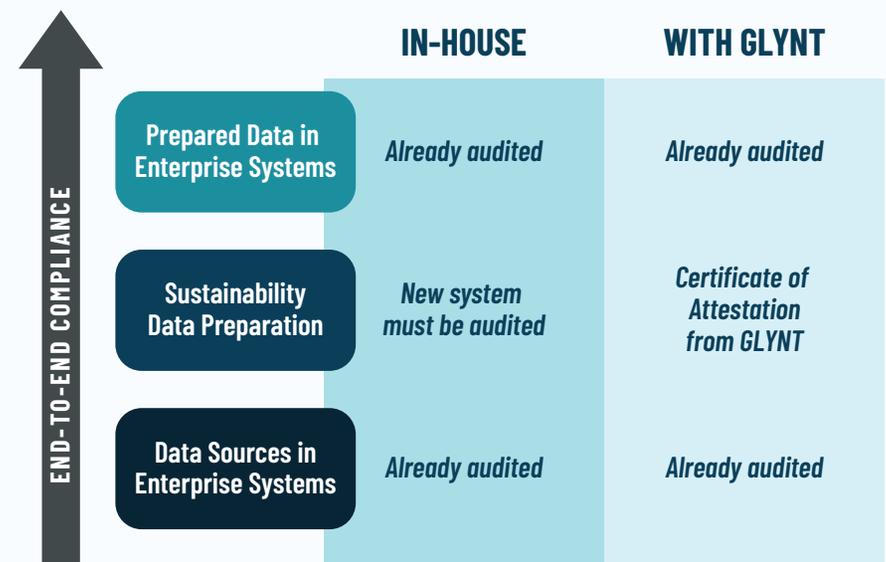
Audits are an external set of eyes on your data preparation method. Sustainability reporting has two types of external reviews: Reasonable Assurance, which is very much like a financial audit and Limited Assurance, which is a negotiated scaleback of the audit. While Reasonable Assurance is not required for a few years, market leaders recognize that half-way measures on audit-ready data have real cost, and that it is cheaper and less risky to get the new data flows audit-ready from the start.

Audits are about the system of data preparation, not just accuracy of data. And based on the layers of monitoring and controls detailed above, it is easy to see why. Unless there is a richly layered and thoughtfully constructed system, the data won't be reliable and decision-useful.

So whether you use an in-house system or GLYNT, you'll want to be able to show end-to-end compliance and audit-readiness. If you build an in-house system, expect it to be audited. If you use GLYNT, we'll provide a Certificate of Attestation. Then you don't need an audit of your data preparation methods.

If you simply upload a CSV into a carbon accounting, ESG or sustainability system, you won't have end-to-end compliance. You'll have compliance just from that point forward, assuming the software itself is compliant. This may not be sufficient for your customers, investors and the regulators. After all, the software simply reports the data. Data preparation is when the data is made accurate and reliable.

## GLYNT's Certificate of Attestation Makes Compliance Easy



The way we think about it here at GLYNT is as follows: No one processes their own payroll. They use ADP, Gusto and other compliant providers. Payroll data flows into monthly accounting closes and is accurate. It is compliant with local regulations, employee laws and so on. ADP undergoes a SOC 1 audit each year, having a third-party review of their data handling and flows. GLYNT is doing exactly the same for sustainability data. We're undergoing a SOC 1 audit in Q1'2024.

Once we're done, we can share our Certificate with you. Thus you can rely on GLYNT and avoid the cost of building an in-house sustainability data preparation system. Your auditors will appreciate our certificate, our documentation, and audit-ready files.

# The GLYNT Customer Experience

The GLYNT customer experience is simple: Provide GLYNT the original source data. GLYNT will send you the prepared data files. In, Out. That's it!

Operationally, this means every month your data files are received and your data is prepared in 5 days or less. New customers are onboarded and receive a full year of data in under 90 days. Current customers can add sites and data sources, maintaining full automation with GLYNT.

GLYNT does not have a team of hundreds working in a low-cost country correcting or entering data. We are a highly automated service and a company with a small headcount. We live and breathe technology and automation in a fully compliant, audit-ready sustainability data preparation system. We're nuts about accuracy and performance. We're working hard so you don't have to.

## RESOURCES

- [IAASB](#) – See the sustainability audit standards here
- [COSO](#) – Gain a perspective on compliance for sustainability from the standards setters
- [DCAM](#) – See the lessons learned from the financial sector on how to produce reliable data
- [Forbes](#) – We wrote about how Gen AI interacts with sustainability data
- [12 Steps for Assurance-Ready Sustainability Data](#) – GLYNT's Guide on what it takes to build an audit-ready, assurance-ready sustainability data preparation system



GLYNT is The Sustainability Data Company, producing investor-grade data for businesses around the world. Our audit-ready sustainability data enables accurate reporting, operational efficiencies and access to financial capital. With a purpose-built machine learning system, GLYNT is the automated solution for all types of water, waste, energy and emissions data. Speed work, lower costs, and power ESG, carbon accounting and other business systems with accurate, actual data from GLYNT. Learn more at [glynt.ai](https://glynt.ai)

