



SHIPPZO

VISIBILITY GUIDE

**Ensuring high data quality
to maximize ROI of real-
time transportation
visibility solutions**

INTRODUCTION

The ability to track an item as it travels throughout the world in real-time is becoming increasingly accessible with the proliferation of faster and more reliable connectivity, cost-effective IoT devices and more compatible systems and software platforms.

However, the question that supply chain management and transport teams ask themselves has shifted from 'where is my shipment' at the macro level to a more granular 'when will it arrive at the next stop?' and 'are there any risks of a delay?'. As such, the ability to predict shipment ETAs has become increasingly a must-have, critical capability for all shippers.

It goes without saying that ETA accuracy and reliability depends heavily on the quality of the data captured and the sophistication of the technology and computational methodology used.

This paper explores the considerations and methodology required to ensure visibility data leveraged by real-time transportation visibility platforms is of the highest possible quality, so that it can be used to conduct comprehensive analyses on shipment flows and generate valuable insights into a supply chain's operations and processes.

To find out more about techniques used to then turn the quality data into an accurate and reliable shipment ETA, please view our paper 'The business value of accurate and reliable ETAs'.

Contents

PART I		
The value of real-time visibility		3
PART II		
The components of data quality		4
PART III		
Addressing the causes of low data accuracy and reliability		7
PART IV		
Summary		11

About the Author

Tarik Agayr
Product Director, Data & AI/ML – Shippeo



Tarik is an expert at converting data into actionable insights and helps supply chain companies achieve significant ROI from their data. At Shippeo, he heads the Data & AI/ML function and leads a team of very talented Data Scientists and Machine Learning engineers. Their work is at the heart of the Shippeo platform delivering data products ranging from prediction algorithms, to anomaly detection, dashboards and decision support tools to our customers. Prior to Shippeo, Tarik was at Quintiq where he delivered supply chain planning and optimization solutions. Tarik has a masters degree in Operations Research and a Bachelor's degree in Engineering with a major in Applied Mathematics. He has been with Shippeo since 2018.

The value of real-time visibility

The fundamental purpose of a real-time transportation visibility (RTTV) platform is to capture operational data and present key stakeholders and trading partners with key insights around shipment operations in real time. The imperatives in front of these trading partners are significant:

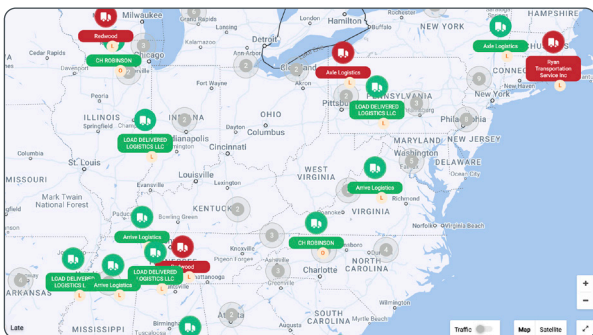
- 1 They need to anticipate and warn their customers of delays as much in advance as they can and do so in a manner that avoids resource-intensive back and forth communications.
- 2 They have to avoid stock-outs and production line halts: both of which are costly for any organization.
- 3 When disruptions happen, they need to re-adjust plans and operational processes quickly.
- 4 Shippers have to avoid congestion at docks throughout their supply chains - whether at a port or at a bay in a warehouse, by optimizing the assignment of their resources at the right bays or gates at the right time.
- 5 They always need a clear view on their costs. The faster they know this, the better it is for them and for their trading partners.
- 6 And last but not least, with sustainability such an important factor for supply chain stakeholders, every trading partner in the supply chain needs a granular understanding of the carbon footprint at a shipment level.

RTTV platforms have to provide solutions to these imperatives. They must collect real-time granular data from trading partners, trucks, ships, trains, etc. and use their AI/ML algorithms to generate key insights in a timely manner. As the old adage goes though, algorithmic outputs are only as good as the data you feed them. Which means, it all boils down to the data the platform is able to capture, collate, cleanse, complete, and contextualize in a timely manner. Laser-like focus on data quality is what differentiates a 'great' RTTV platform from a 'satisfactory' one.

Although other insights provided by RTTV platforms, such as those on costs, slots and carbon footprint, are equally important as ETA, for discussion purposes, we will focus on ETA in the rest of the paper to underscore the need for quality data in an RTTV platform.

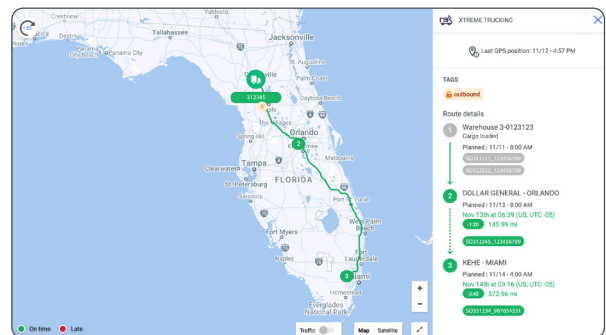
Shipeo maximizes the success of your RTTV project through three important commitments.

Shipeo is the only provider to offer a triple SLA on onboarding speed, tracking quality and ETA accuracy, to ensure a quick time-to-value for every customer. These SLAs all reflect the organization's focus on achieving high data quality levels for customers and help to ensure maximum project ROI.



Real-time visibility

Where is my shipment?



Predictive visibility

When will my shipment arrive at the next stop?
Are there any risks of delay?

The components of data quality

Quality data needs to come from quality sources in order to ensure accurate and reliable outputs, primarily ETA predictions.

'Data quality' is a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability, latency, and many other factors.

Three important measures that we will focus on in this paper are:



COMPLETENESS



LATENCY

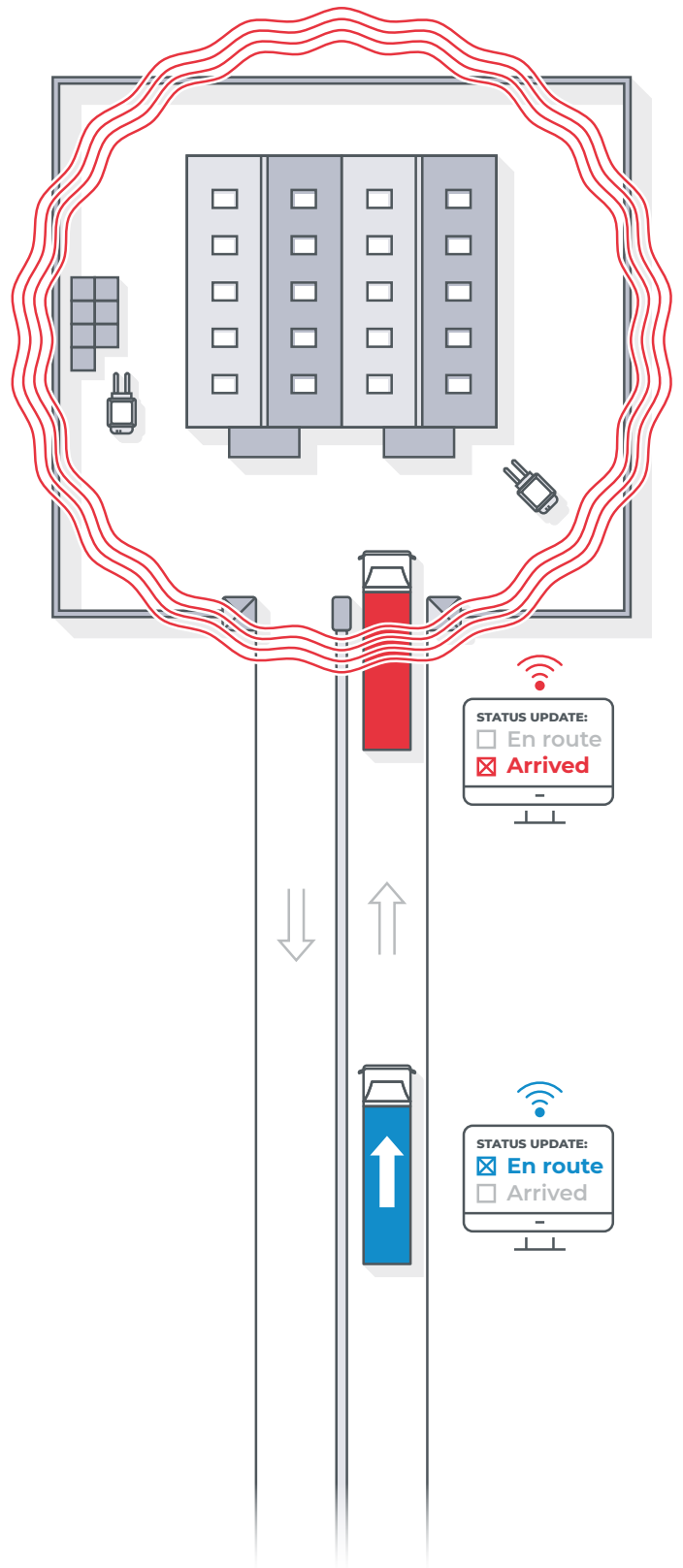


ACCURACY AND RELIABILITY

Machine Learning: why traditional algorithms won't work

Leveraging insight from historical data to calculate quality ETAs requires highly complex mathematical transformations. The sheer volume of available data makes it difficult to glean a consistent pattern for all situations and even if that is possible, the complex rules that make sense of those patterns and calculate predictions are difficult, if not impossible, to specify a priori. Even if they could be specified, we know that data changes all the time, conditions change and this makes those complex rules that govern these calculations change constantly. Which means maintaining those manually is an impossibility, and we need a learning system to ensure algorithms are constantly adapting and staying current.

Enter AI/ML. AI/ML models are mathematical models that are built by 'training' them on sample historical data, meaning that they are trained to reproduce patterns observed in historical data. More importantly, they are trained to learn from changing data - in other words, these models can make predictions without being explicitly programmed to do so. It goes without saying therefore that high quality data is needed for AI/ML techniques to unlock their full potential.



PART II

The components of data quality

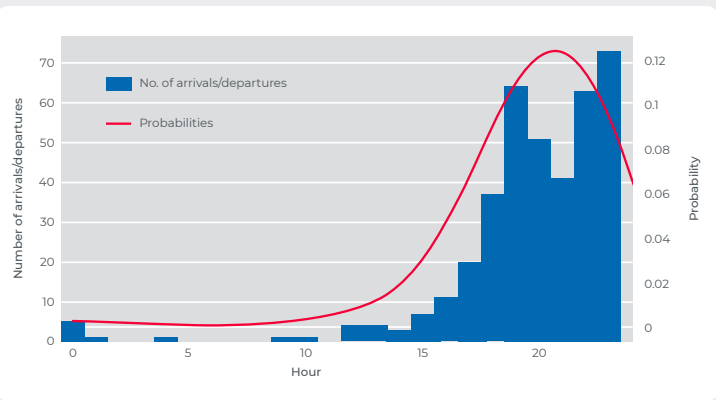
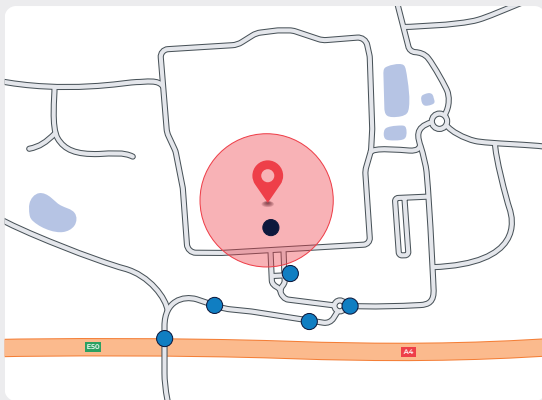


Completeness

Traditional approaches focus on getting the average case right, and for that you don't necessarily need a "full" data set. But shipment visibility doesn't work on averages; every shipment's ETA is as important as any other. More importantly, every shipment's journey is unique, so to cater to every edge case as much as possible, you need a "complete" data set. For this completeness, the platform needs to have a large network of data sources such as partner networks, carrier systems, telematic devices, driver mobiles, etc. to capture as many positions and events as possible, relentlessly.

Completeness is important because AI/ML algorithms need large datasets that need to be "complete" in terms of length (the number of table rows or 'records' and ideally no row is missing) and width (the number of table columns or 'features' and ideally no column entry is missing), especially when accurate prediction is a necessity for every single case.

A number of ML-based imputation techniques are used by Shippeo to improve data completeness. For instance, geofencing information is leveraged by these techniques to impute some of the missing data. Historical data is another source to impute data from.



Shippeo connects with 850 real-time data sources via partners, allowing more data to be leveraged.

Example 1: Geofencing:

Along with GPS positions, a geofence can be used to trigger arrival and departure events.

Example 2: Opening hours:

Sometimes these are unknown, but can be inferred.

Shippeo connects with **850** real-time data sources via partners, allowing more data to be leveraged.



PART II

The components of data quality

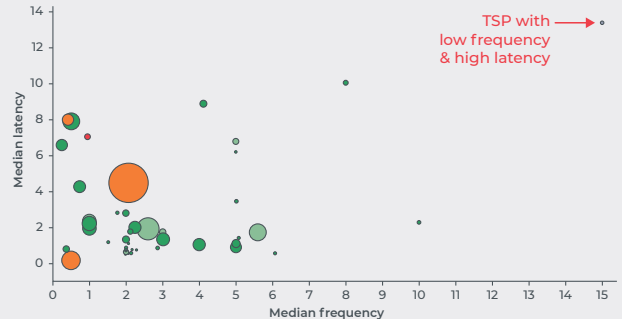


Latency

Latency, which is also referred to as timeliness, is associated with data availability. It measures the extent to which data that is needed is available at the right time. Latency is important at the time of prediction, because it determines how quickly ETA predictions are made, how up-to-date the information and valuable real-time insights are, and how up-to-date the data driving that prediction is.

For operational insights and ETAs to be accurate and reliable, the GPS data from telematics devices should be received regularly and with as small a delay as possible, making it important to measure the frequency and latency of GPS positions received. The chart below plots the median latency versus frequency for a subset of telematics service providers. This explores the time difference between the GPS position's actual timestamp and that of when it was received by Shippeo's servers in relation to the frequency of GPS positions being received.

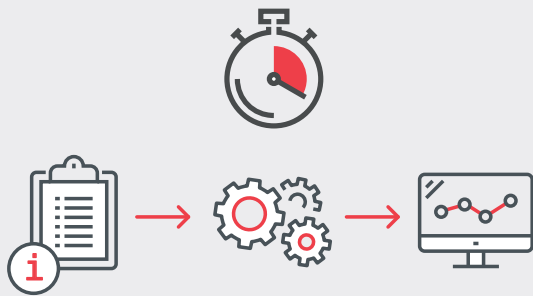
TELEMATIC PROVIDERS



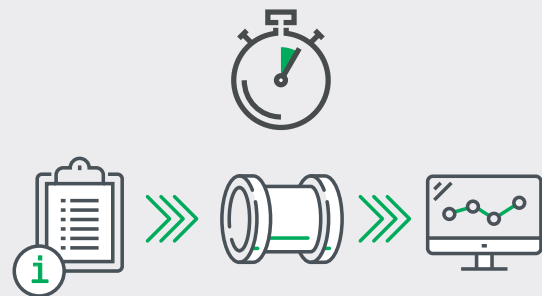
An illustration of the measurement of GPS frequency and latency by telematics service provider (TSP)

Shippeo relies on stream processing to ensure timely receipt and processing of data in order to reduce latency. Where streaming is not possible for any reason, the platform can switch over to batch processing.

BATCH PROCESSING VS. STREAM PROCESSING



Every 20 minutes



Near real-time



Accuracy & reliability

Accuracy & reliability are the two most important data quality indicators when it comes to maximizing the ROI of a RTTV solution. Reliability refers to the trustworthiness of the data, while accuracy refers to how accurately the data captures information or events as they occur in real time. Because ETA predictions and

insights rely heavily on events and GPS positions as input data (amongst other things), it's important to make sure the data leveraged is accurate and reliable. The old adage is true here: 'garbage' input data provided to AI/ML algorithms will result in 'garbage' predictions.

Addressing the causes of low data accuracy and reliability

The thoroughness of carrier onboarding, carrier buy-in, partner system integrations and overall RTTV platform implementation is key to obtaining high quality data.

In addition, the main solutions for improving data quality involve addressing GPS data accuracy, geocoding accuracy and user behaviors.

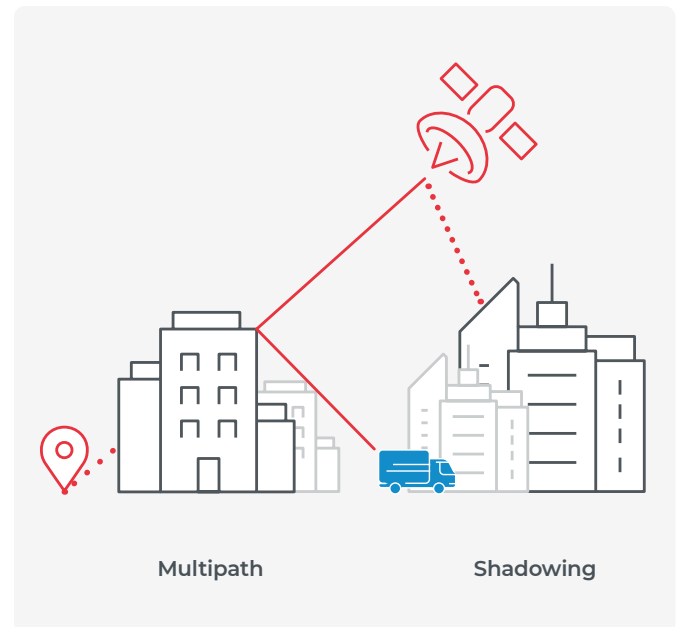
GPS data accuracy

The challenges of working with raw GPS data

Capturing GPS data frequently and stringing the GPS points to map an itinerary of a trip is relatively easy, assuming telematic devices are transmitting. However, predicting driver breaks in road transportation is one of the most challenging steps in providing an accurate ETA. To improve the accuracy of stop detection (e.g., breaks, night rests), it is important to make accurate detections for when drivers are not driving en route. A naive approach would be to use the relative speed between consecutive raw GPS positions. However, this becomes increasingly inaccurate as the frequency of positions received becomes smaller.

Besides, there are a number of factors that impact GPS positioning accuracy:

- Electromagnetic activity, humidity and pressure fluctuations in the atmosphere
- Distribution of satellites
- Street canyons and multipath effect (creating noise because signals are reflected or blocked by tall buildings)
- Time to first fix (when a GPS receiver is first switched on and sends inaccurate position data while it is still getting its bearings)
- Receiver issues such as an unstable clock, manual clock modification, digit truncation, etc.

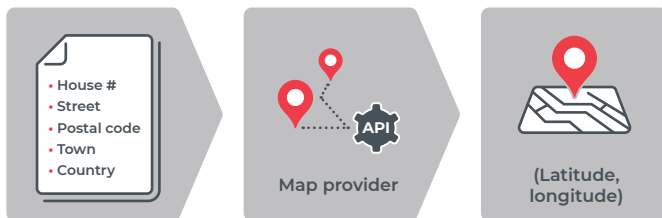


Shippo uses well-known methods such as filtering and map matching alongside other advanced proprietary statistical techniques to produce a smoothed and cleansed GPS signal. These methods take uncertainty into account, and hence minimize the impacts of inaccurate GPS position data.

Geocoding accuracy

Geocoding is the computational process of transforming a physical address description to a location on the Earth's surface (as a set of numerical coordinates). Needless to say, a lot of things can go wrong with geocoding.

Address information is often extracted from transport orders, sent via API from Transport Management Solutions. The information set contains information about the loading and delivery sites, which are geocoded before being instantiated and used so that shipments can be tracked across a journey from start to finish.

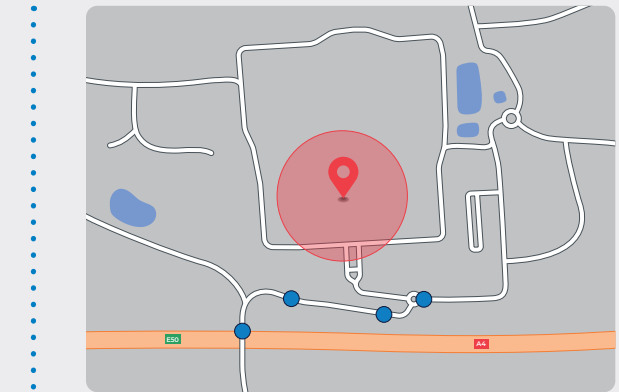


However, sometimes address details from the source systems contain errors. Town names get mixed up, digits in ZIP/postal codes or even street numbers go missing. Furthermore, the delivery/loading site could be geocoded incorrectly. Each error in itself can lead to tracking inaccuracies; combined they can lead to serious data inaccuracies. Shippeo's AI/ML algorithms attempt to mitigate such errors by flagging addresses that are potentially geocoded at the wrong location for instance, and offer proactive ways to correct locations to improve future accuracy.

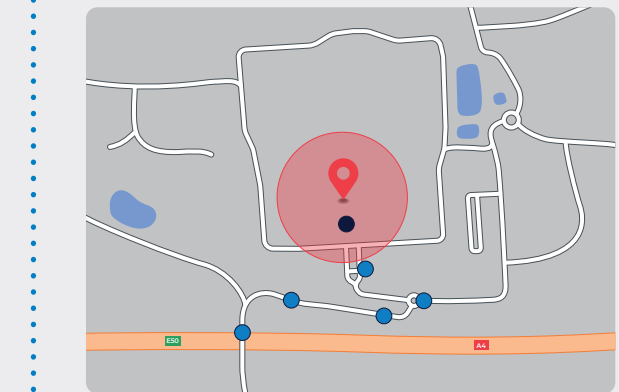
CURRENT ORDER STATUS: Left loading site

Automatic GPS tracking couldn't be achieved in this example because the delivery site is recorded as being at the wrong location. The order status never automatically changes to *delivered*, because the arrival is not captured."

WHAT IS GEOFENCING? (SIMPLIFIED)



✓ The truck is approaching the delivery site.



✓ The order is considered as 'arrived at delivery site'. The black dot corresponds to the first GPS point inside the geofence. The arrival timestamp is that of the GPS position.



✓ The truck remains on site for a while, and then leaves the delivery site and then drives off. The second black dot corresponds to the first GPS position outside the geofence area.

User behaviors affecting tracking rates

Insights on operational performance derived from perfect flow tracking, with GPS pings from start to finish, are more reliable and accurate. ETA prediction models in particular yield good results when they are trained on datasets that contain such perfectly tracked flows. These flows need to be available in sufficient amounts to produce meaningful data.

However, some of these flows can't be used at 'training time' because the shipment is not perfectly tracked. In fact, many anomalies can arise throughout the lifecycle of the shipment, preventing it from being 'automatically' tracked by a GPS from loading to delivery.

These different anomalies occur due to many reasons, including:

- Late order creation
- Lack of adherence in the confirmation or resource assignment process
- Poor GPS signal
- Address & resource management issues

Enter human-in-the-loop. Some data quality issues can only be fixed by users because they need expert judgment. For example, is data missing because a shipment never eventuated (so it is right that it is missing) or is it because of a data capture problem (in which case it needs fixing). These fixes can only be implemented if such issues get spotted and proactively pushed to users for immediate solving and if users are equipped by their RTTV provider with decision support tools that prescribe the next best set of actions to help them address the different root causes through intuitive workflows and user experiences.

Leveraging technology to eliminate the root causes of low GPS tracking rates

Improving your tracking rate is not an overnight process. In addition to other initiatives described above to improve tracking rates, Shippeo has also created a decision support tool, Smart Tracking Analyzer™ (STA), that not only performs in-depth root cause analysis for all tracking rate issues, but also goes a step further and helps users to actually understand the problems via an intuitive dashboard and then suggests remedies to fix the root causes.

Typical questions answered by STA are:

- What can I do to improve my tracking rate?
- Why are the untracked orders not tracked properly?
- What prevented the orders from being tracked?
- What are the main root causes for tracking inaccuracy?
- Where should I start if I want to improve my tracking accuracy?

PART III

Addressing the causes of low data accuracy and reliability

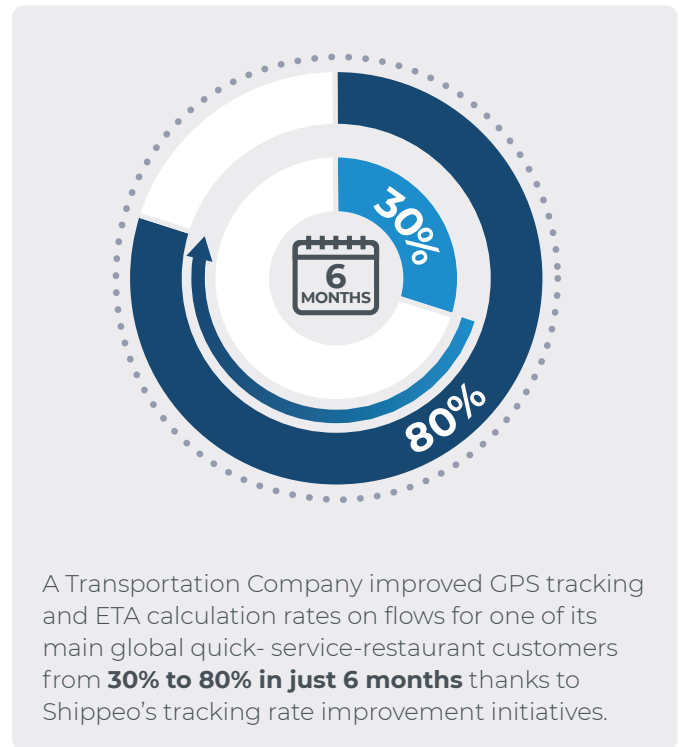
In practice, STA detects 99.8% of the root causes of tracking rate problems. It helps users with a clear overview of the root causes behind the lack of tracking for their orders at a level of granularity of their choosing.

From the STA, it is possible for a shipper or a carrier to generate a report that shows metrics such as:

- their global tracking rate
- the top 10 carriers causing tracking issues
- the top 3 main non-tracking issues
- the root causes, if fixed, would have the biggest impact on improving tracking rates
- the carrier with the most non-tracking issues of a given type, in volume or as a proportion of all carriers.
- which carriers are improving their tracking rates fastest

With such information, Shippeo's users typically devise a 3 to 5 point action plan to address the highest priority issues in order to achieve up to a 20 point increase in tracking rate in as little as a month.

To find out more about how this category-leading tool works, along with Shippeo's other tracking rate improvement initiatives such as carrier tracking rate newsletters, you can organize a demo with a Shippeo expert.



Summary

As many will know, AI/ML models can be very powerful to predict accurate ETA in freight transportation. These methods have proven to give results with performance levels never achieved before with simplistic ETA prediction techniques.

The most important prerequisite to achieving high performance of these models is to feed them with high quality data. For this reason, Shippeo has invested heavily in research and development and has recruited

a team of talented data scientists and machine learning engineers in order to build advanced solutions to consistently generate high quality data for shippers and carriers. That said, an AI/ML solution can only go so far and sometimes the last mile in getting to the highest quality data needs user input and intervention. That is why Shippeo also provides users with the right tools and intuitive workflows on our RTTV platform to proactively fix data quality issues where appropriate.

Real-time transportation visibility platform data quality best practices

Only invest in a RTTV solution that...



Clearly defines data quality measures such as completeness, latency, accuracy and reliability and offers appropriate measurement tools.



Consistently and comprehensively addresses common GPS data quality issues through AI/ML algorithms, data filtering and correction techniques.



Ensures a full-service approach is taken when onboarding carriers, providing users with an excellent level of support and training in local languages.



Identifies user-related root causes of low tracking rates and prescribes users with actions to take to fix problems and improve tracking.



Uses AI/ML algorithms to prompt users (whether shippers or carriers) to quickly correct inaccurate inputs or wrong actions.



Provides intuitive user experience and workflows to help remedy any data quality situation.



Benefits from provider SLAs committing to fast, high quality carrier onboarding, high tracking rates and high ETA accuracy rates.



GET IN TOUCH WITH ONE OF OUR EXPERTS TODAY

Hundreds of Shippians are present in cities across the globe.

Email contact@shippeo.com

KEY



Shippeo offices



Shippeo team presence



Shippeo, a global leader and European specialist in real-time transportation visibility, helps major shippers and logistics service providers leverage transportation to deliver exceptional customer service and achieve operational excellence. Their Multimodal Visibility Network connects FTL, LTL, parcel, and container transport and integrates 850+ TMS, telematics and ELD systems using a unique API. The Shippeo platform provides instant access to real-time delivery tracking, automates customer processes and offers unmatched ETA accuracy thanks to a proprietary and industry-leading algorithm developed in-house. Over 100 customers, including global brands like Coca-Cola HBC, Carrefour, Schneider Electric, Total, Faurecia, Saint-Gobain and Eckes Granini, trust Shippeo to track more than 28 million shipments per year across 75 countries.

Learn more at www.shippeo.com

