



collibra®

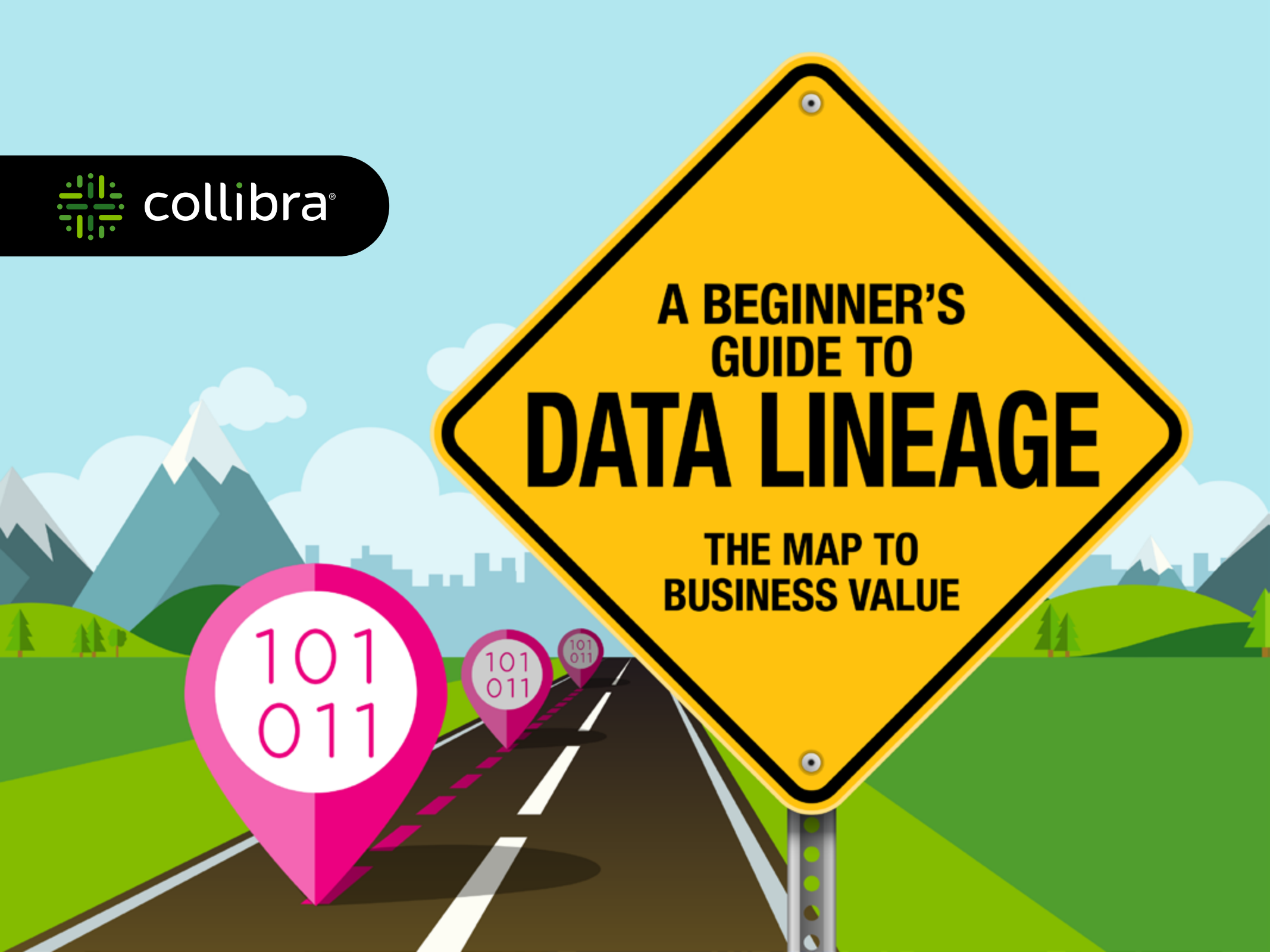
A BEGINNER'S GUIDE TO DATA LINEAGE

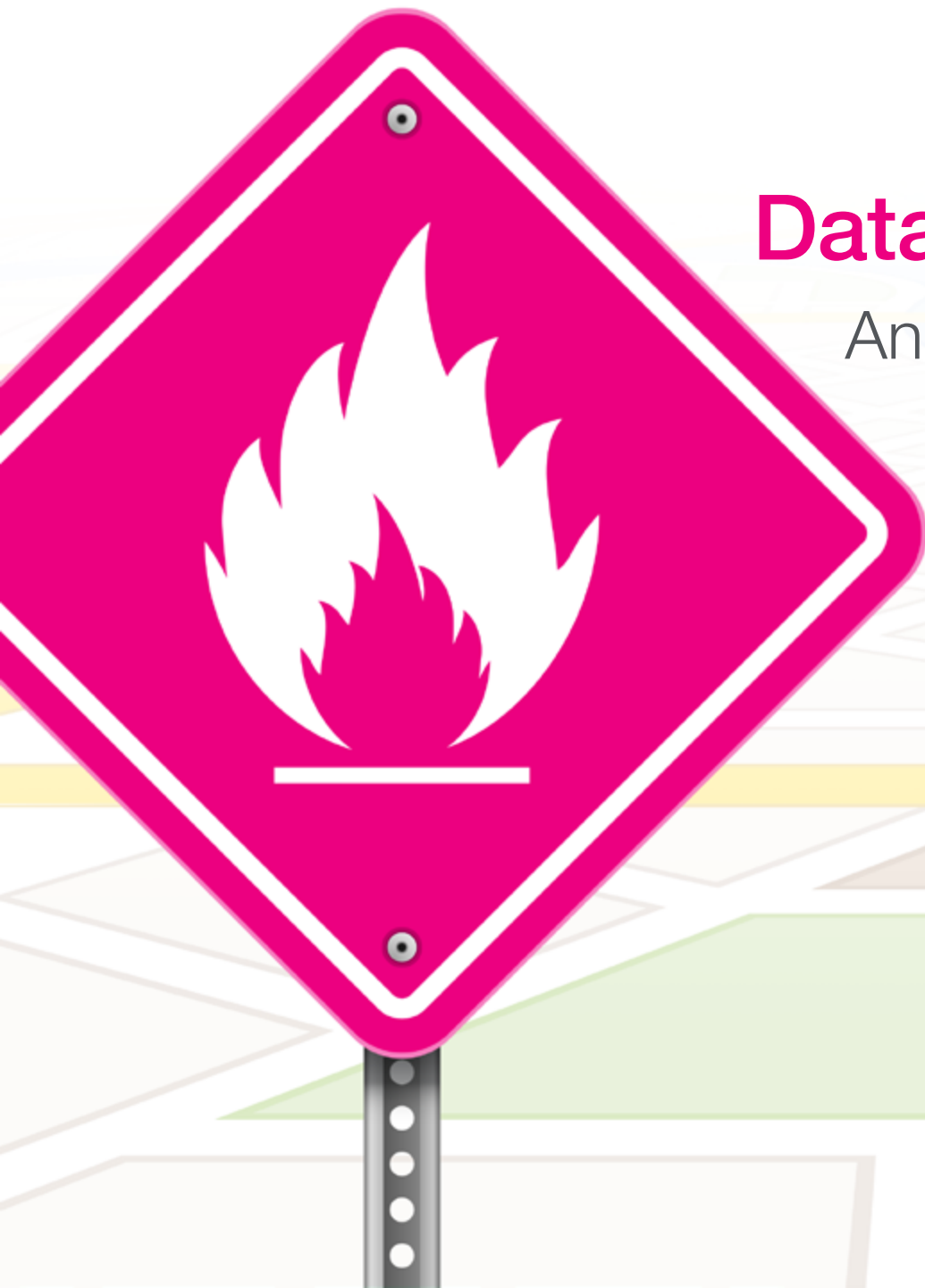
THE MAP TO
BUSINESS VALUE

101
011

101
011

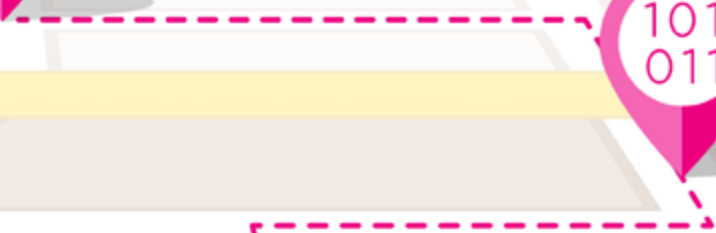
101
011





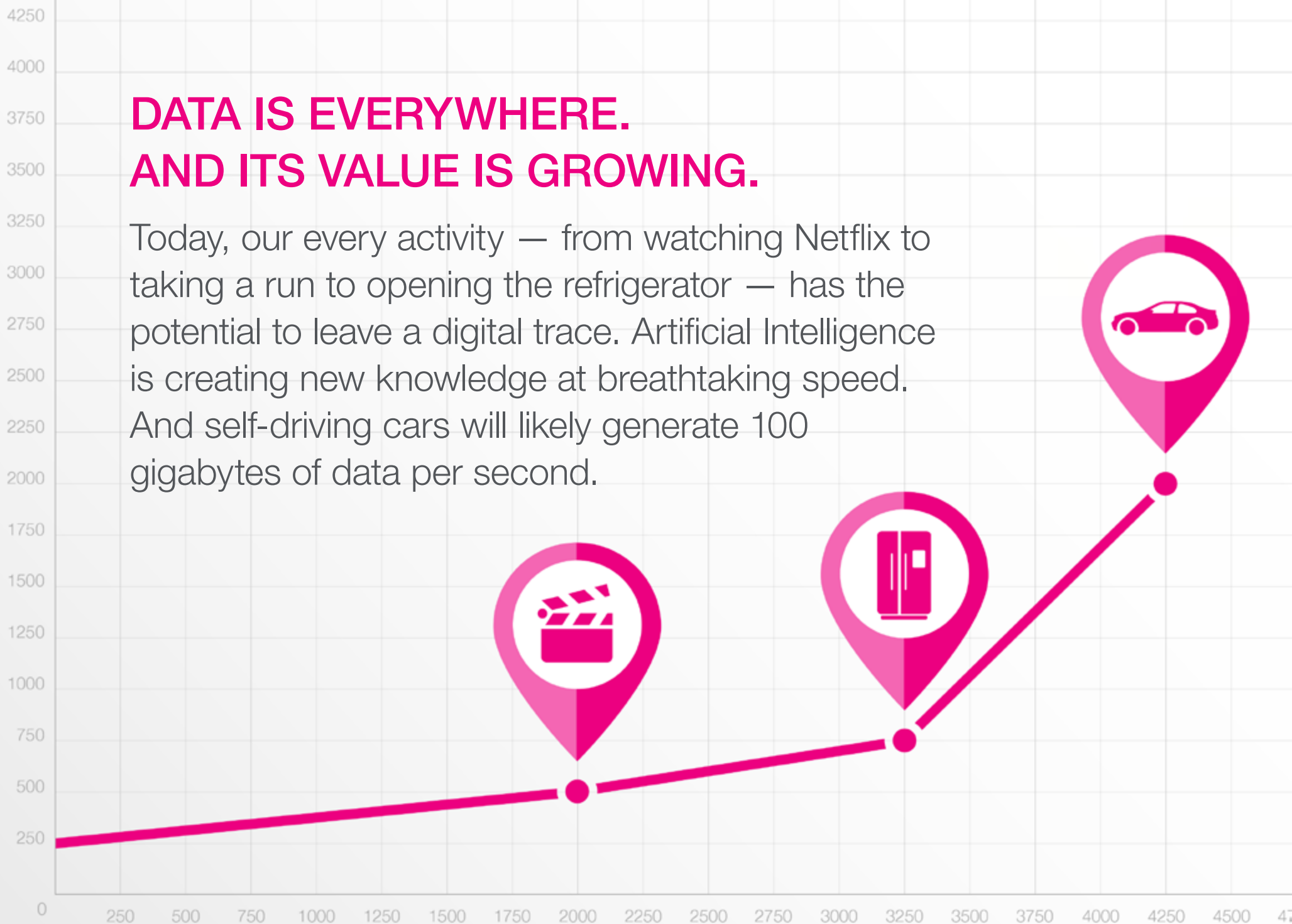
Data lineage is hot.

And that's no surprise.



DATA IS EVERYWHERE. AND ITS VALUE IS GROWING.

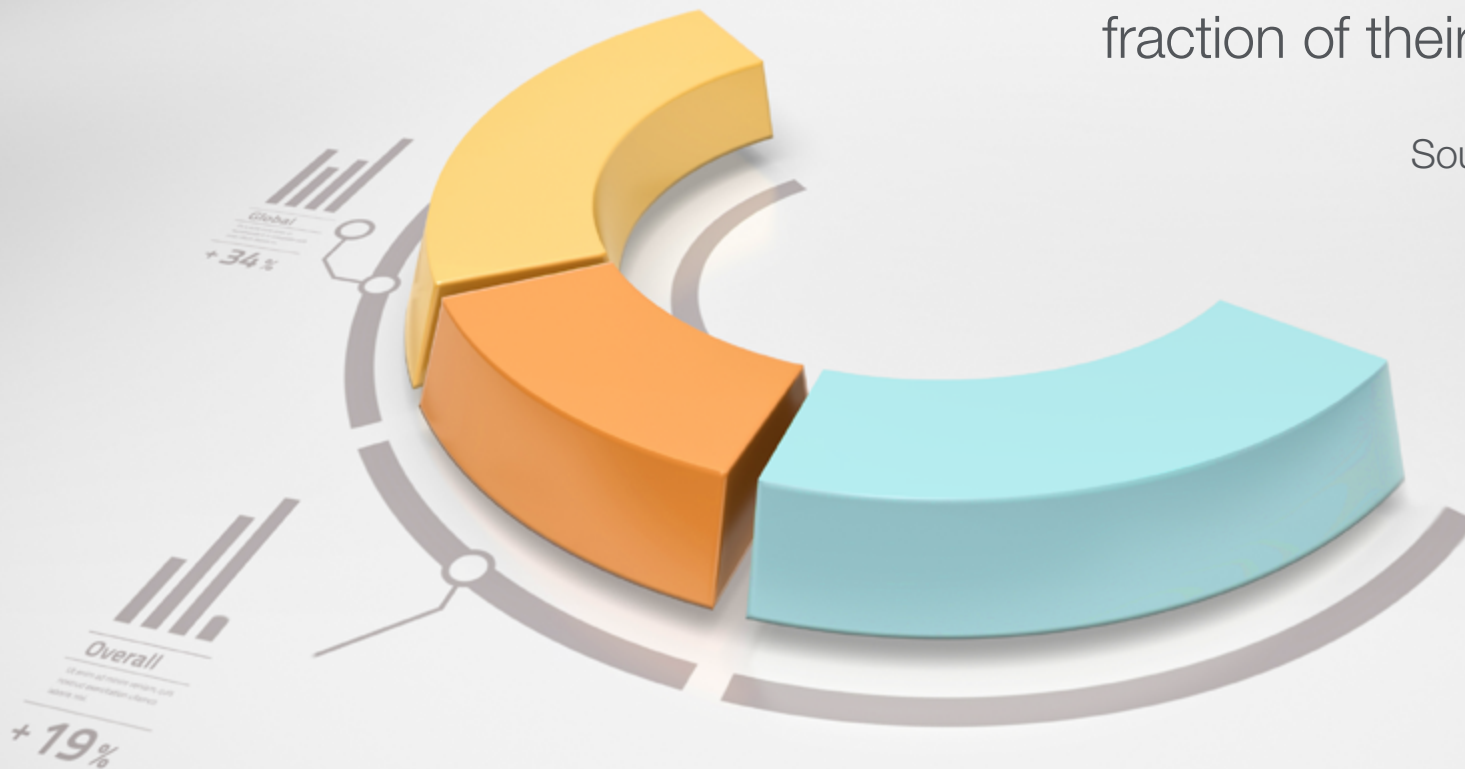
Today, our every activity — from watching Netflix to taking a run to opening the refrigerator — has the potential to leave a digital trace. Artificial Intelligence is creating new knowledge at breathtaking speed. And self-driving cars will likely generate 100 gigabytes of data per second.

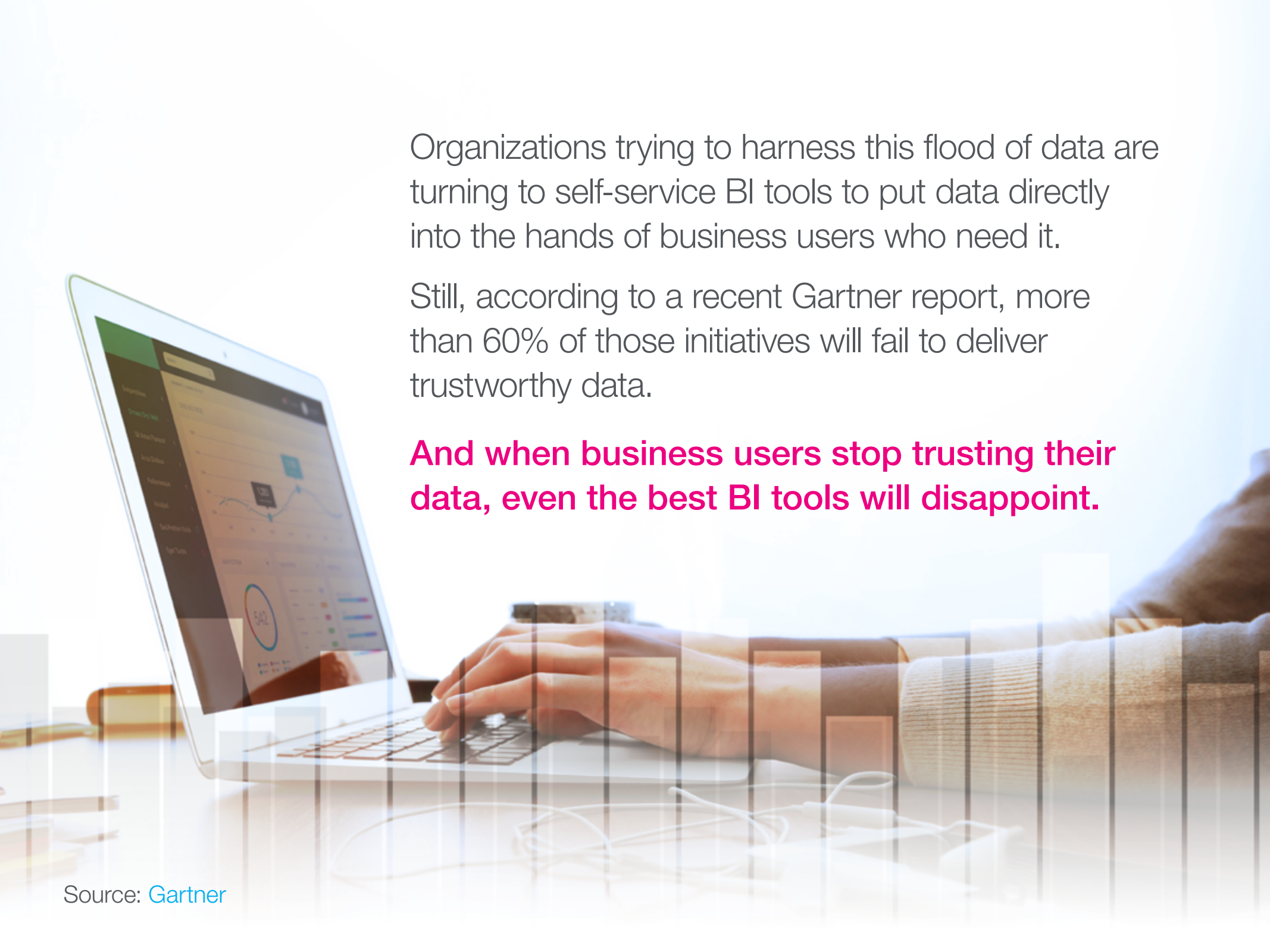


But while data is expanding, our capacity to derive value from it is not keeping pace.

A McKinsey study suggests that most companies are capturing only a fraction of their data's value.

Source: [Mckinsey.com](https://www.mckinsey.com)





Organizations trying to harness this flood of data are turning to self-service BI tools to put data directly into the hands of business users who need it.

Still, according to a recent Gartner report, more than 60% of those initiatives will fail to deliver trustworthy data.

And when business users stop trusting their data, even the best BI tools will disappoint.

Mastering the data deluge to extract value from your data begins with understanding where your data is coming from, where it's been, how it's being used, and who's using it.

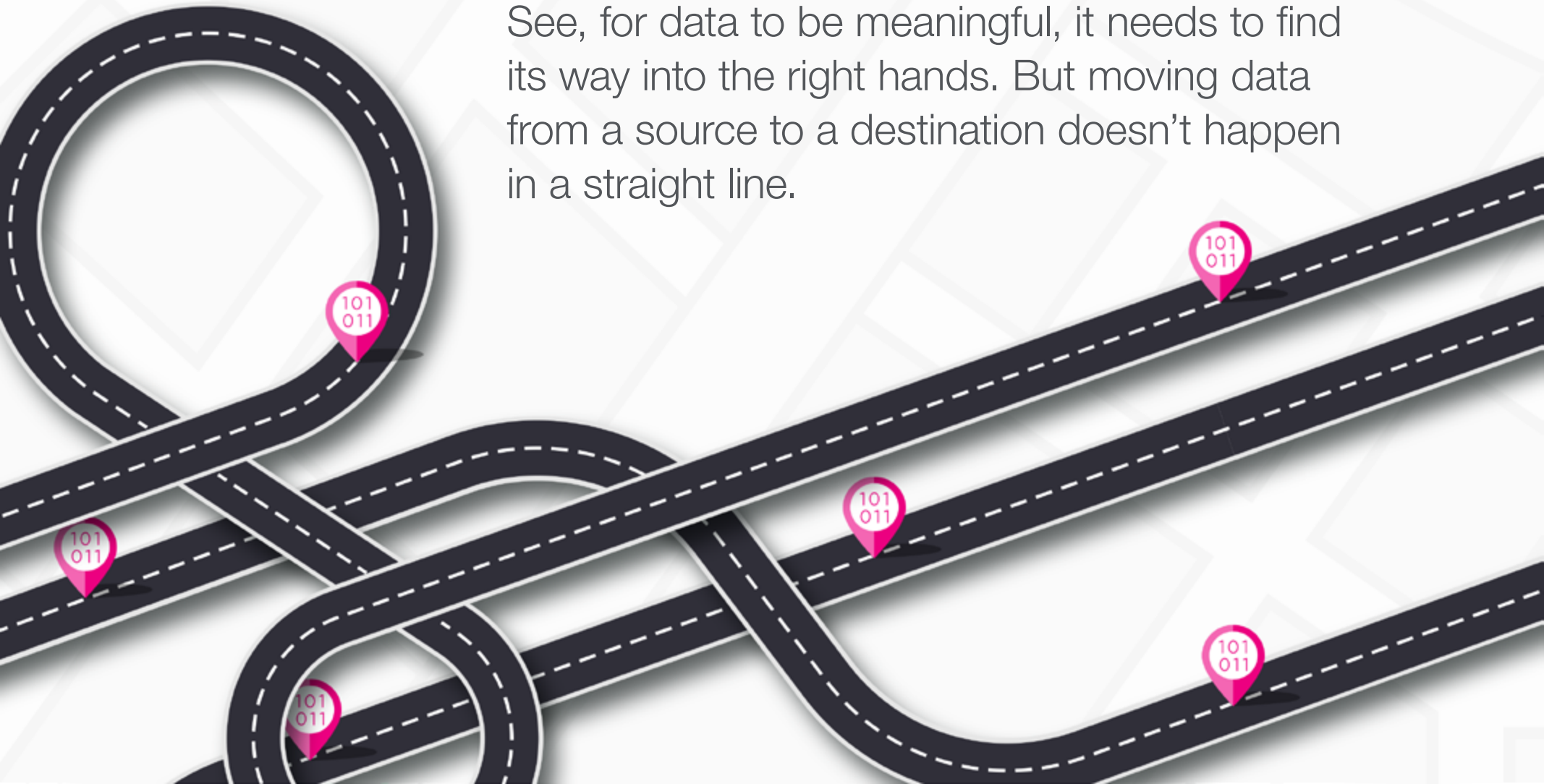
SIMPLY PUT, THE DATA'S LINEAGE



SO, WHAT'S DATA LINEAGE ALL ABOUT?

It's about finding data at any point on its journey.

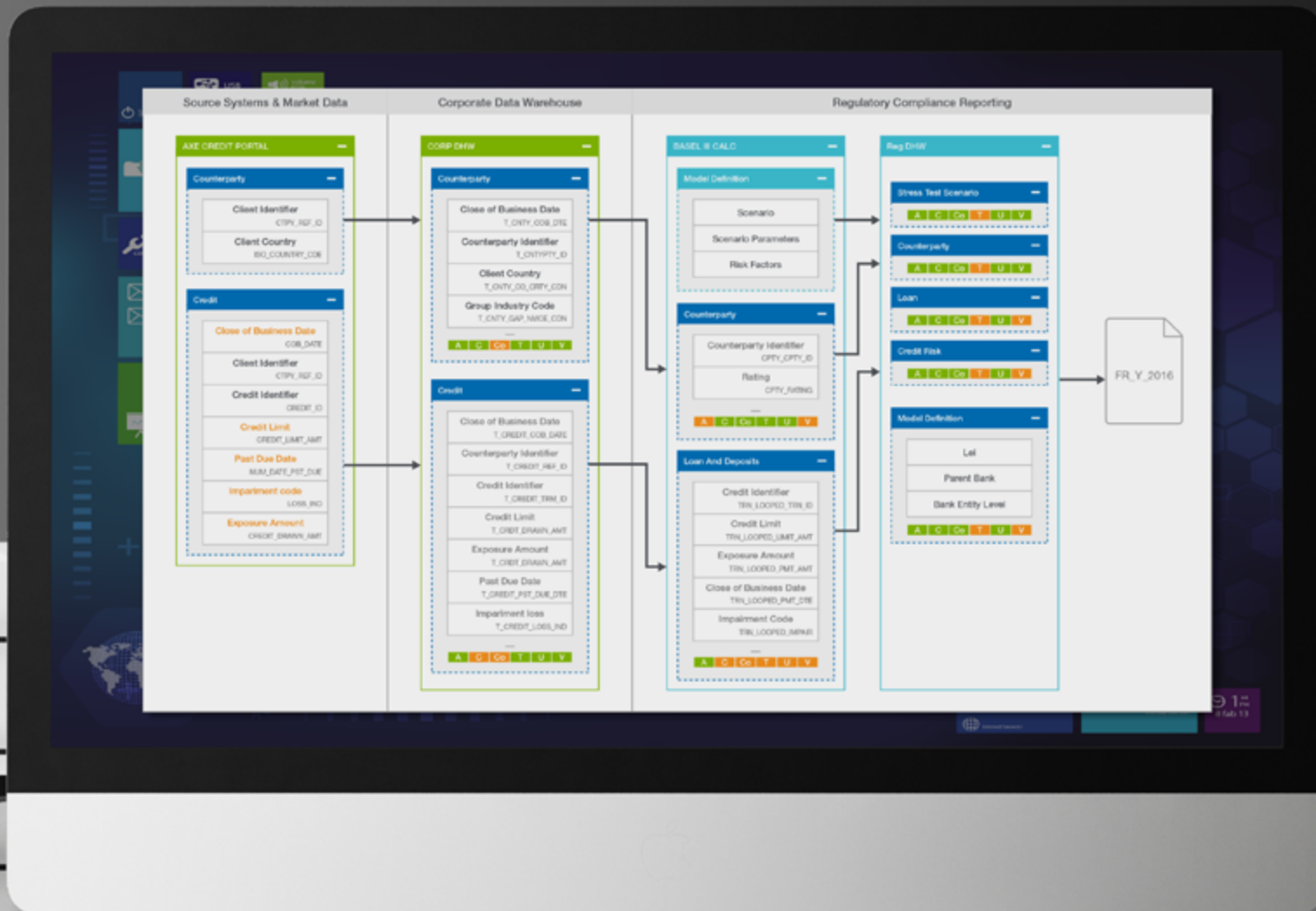
See, for data to be meaningful, it needs to find its way into the right hands. But moving data from a source to a destination doesn't happen in a straight line.



Data pulled from various sources sometimes ends up in a staging area — a data warehouse or a data lake. And before it's delivered to the business user, data is typically aggregated, manipulated, and transformed — often with ETL tools, sometimes with spreadsheets, or ad hoc SQL or Python scripts.



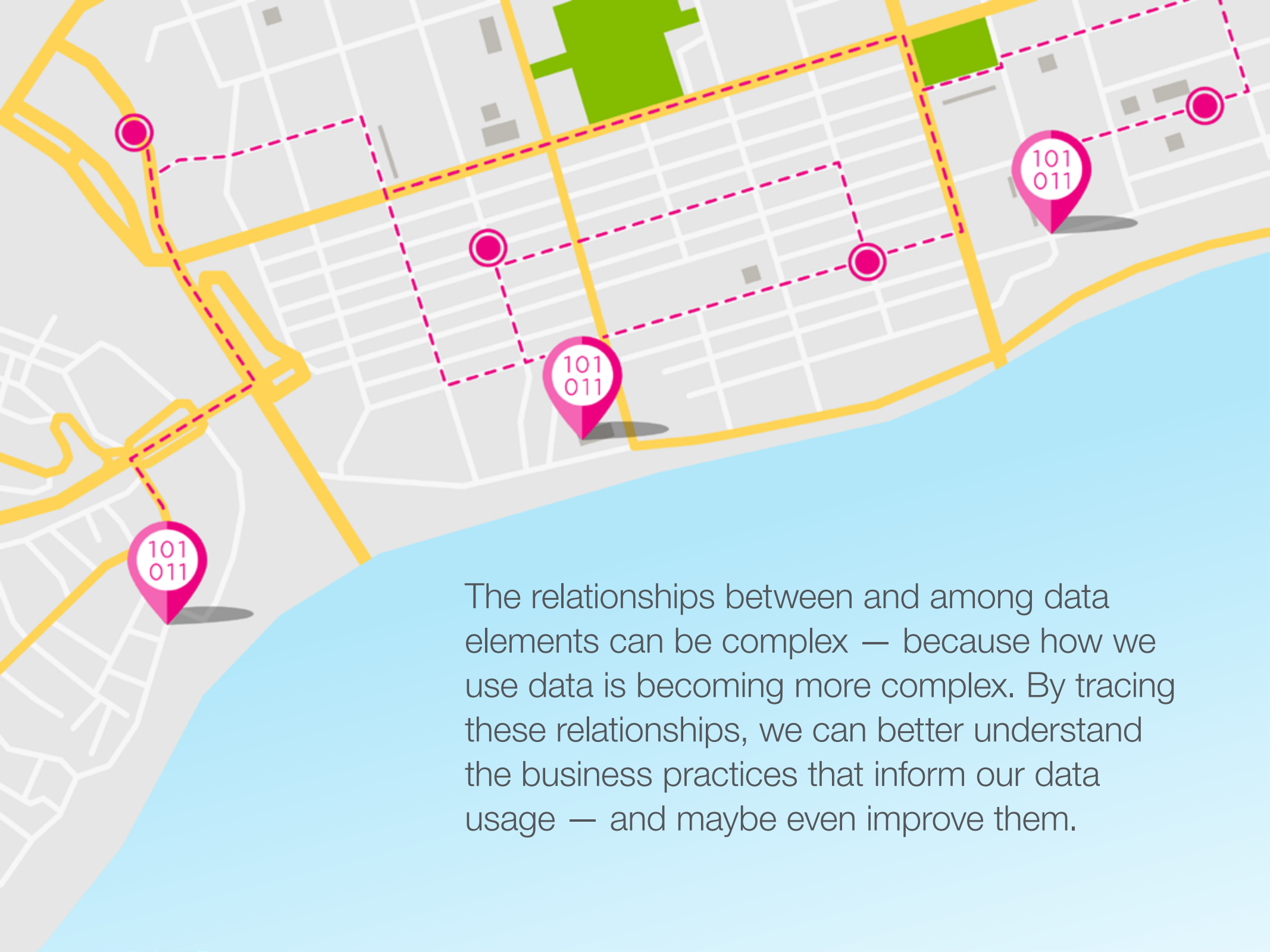
It's a data wrangling nightmare. But the best data lineage diagrams document these journeys to help data users better understand processes and their dependencies.



IT'S ABOUT UNDERSTANDING YOUR DATA AND ITS RELATIONSHIPS

Who created the data I'm looking at? Is it good data? Who touches that data? And when was it last touched? What decisions are being made using this data? And who's making them? If I change something, will I break something else? What impact will that change have on my reports or queries? What's the cost of making a data change?





The relationships between and among data elements can be complex — because how we use data is becoming more complex. By tracing these relationships, we can better understand the business practices that inform our data usage — and maybe even improve them.

IT'S ABOUT TRUSTING THE QUALITY OF THAT DATA

When data moves, or when it is manipulated, or when it is created for one purpose and used for another, things can go wrong.

And data quality can suffer. Data lineage provides a roadmap to data consistency, completeness, timeliness, and conformity at every point in the data journey — allowing anyone touching that data to assess its quality or identify quality gaps.




Establishing data lineage is the practice documenting, visualizing, and tracing the interdependencies of your data.

But fundamentally it's about something much simpler: it's about finding, understanding, and trusting your data.

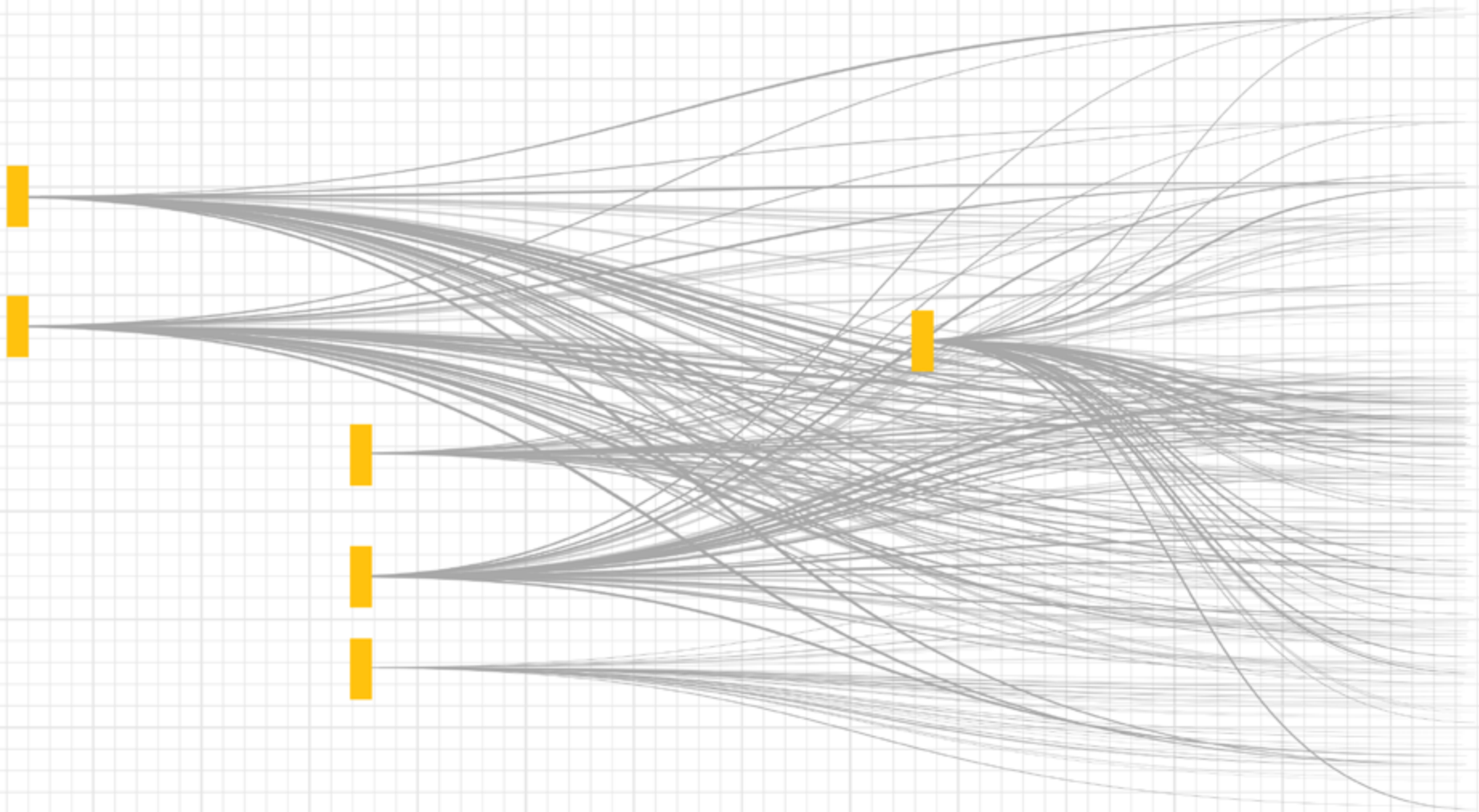
In other words, it's about driving data value.



A person is shown from the chest down, wearing a grey long-sleeved shirt. They are sitting at a desk, holding a white smartphone in their left hand and using their right hand to interact with the screen. To their right, a silver laptop is open on the desk. The background is a bright, out-of-focus window. The image has a semi-transparent grid overlay.

So, a clear picture of your data's lineage — its journey across your organization — is a valuable tool for every data citizen working to move the business forward.

**SOUNDS GREAT, RIGHT? BUT THERE'S A BIG PROBLEM.
AND IT LOOKS SOMETHING LIKE THIS:**



With data — and data sources — growing exponentially, mapping how data moves across your enterprise can be an exercise in futility.

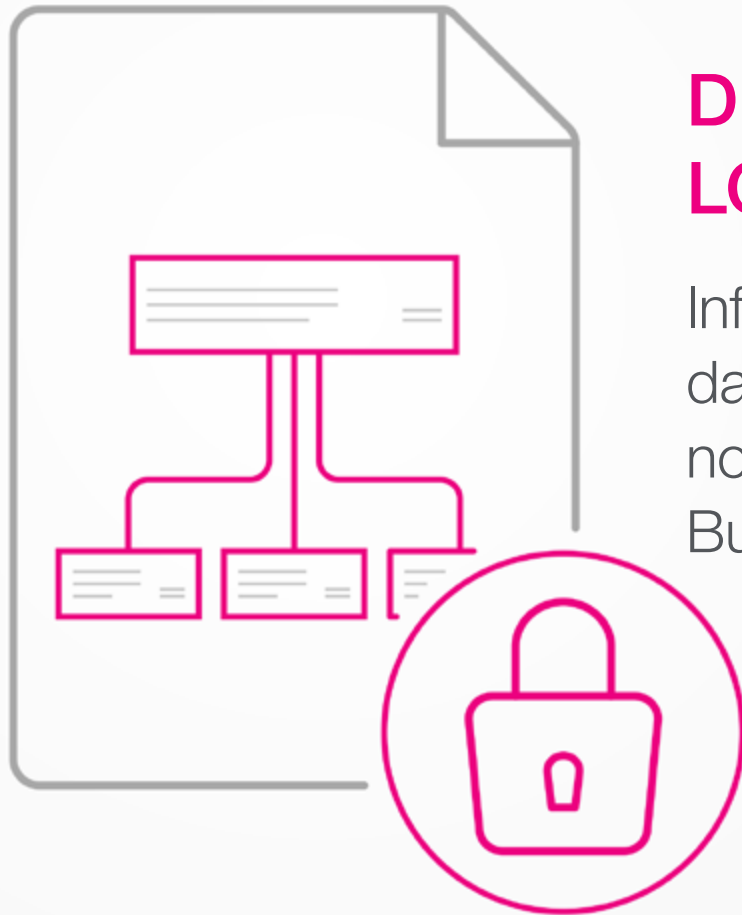
And most businesses are still struggling to untangle the data lineage knot. For very good reasons.





DIAGRAMS ARE TOO COMPLEX.

And too detailed for most people in the organization to comprehend — even a skilled information architect. That makes them meaningless for the business user who needs them to understand the data.



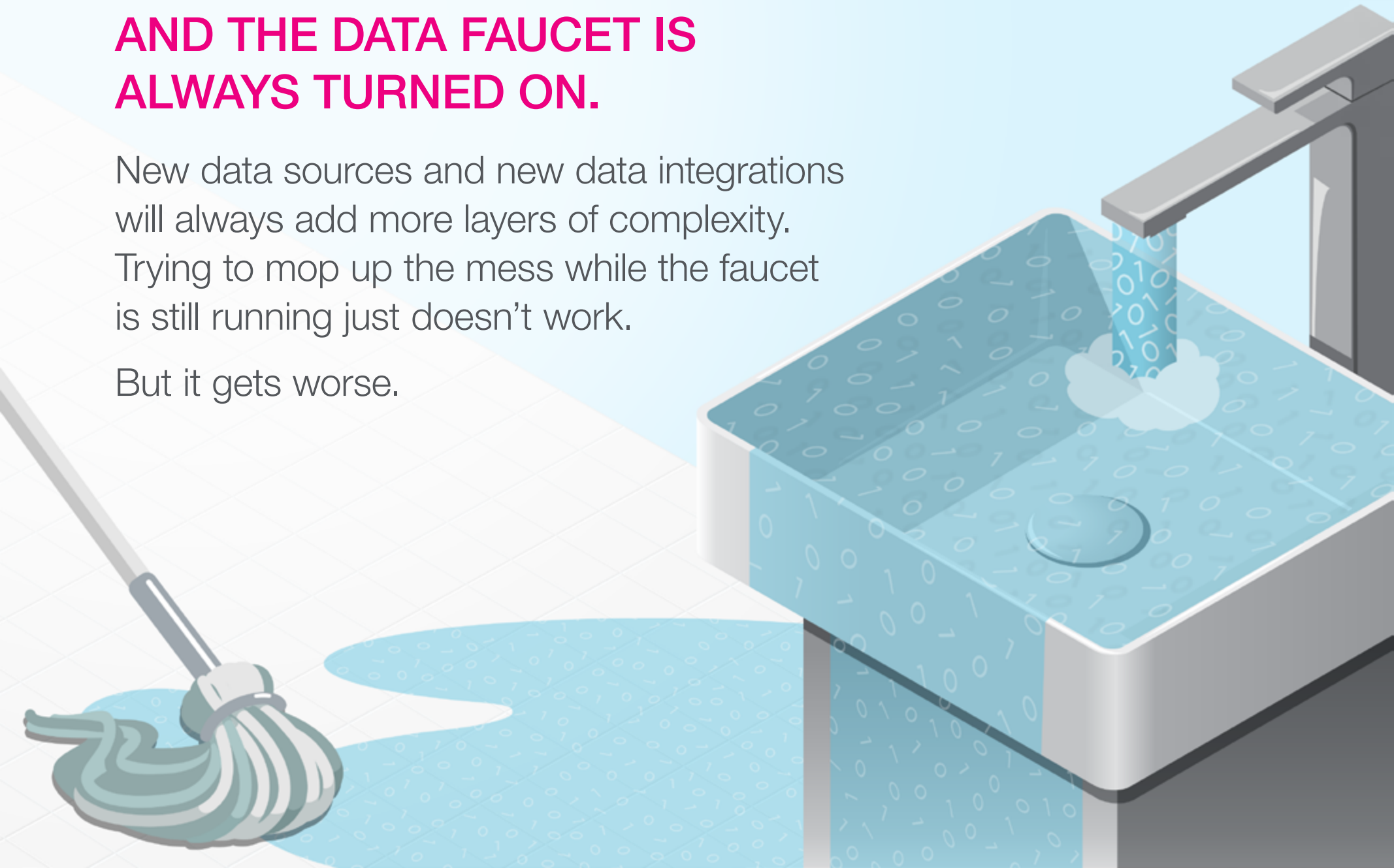
DIAGRAMS ARE LOCKED UP OR LOCKED DOWN.

Information architects spend hours mapping data lineages. Then they save them to a server no one can find and move on to other projects. But even when those diagrams get circulated, they are usually static, read-only documents that don't give business users a way into the data they need.

AND THE DATA FAUCET IS ALWAYS TURNED ON.

New data sources and new data integrations will always add more layers of complexity. Trying to mop up the mess while the faucet is still running just doesn't work.

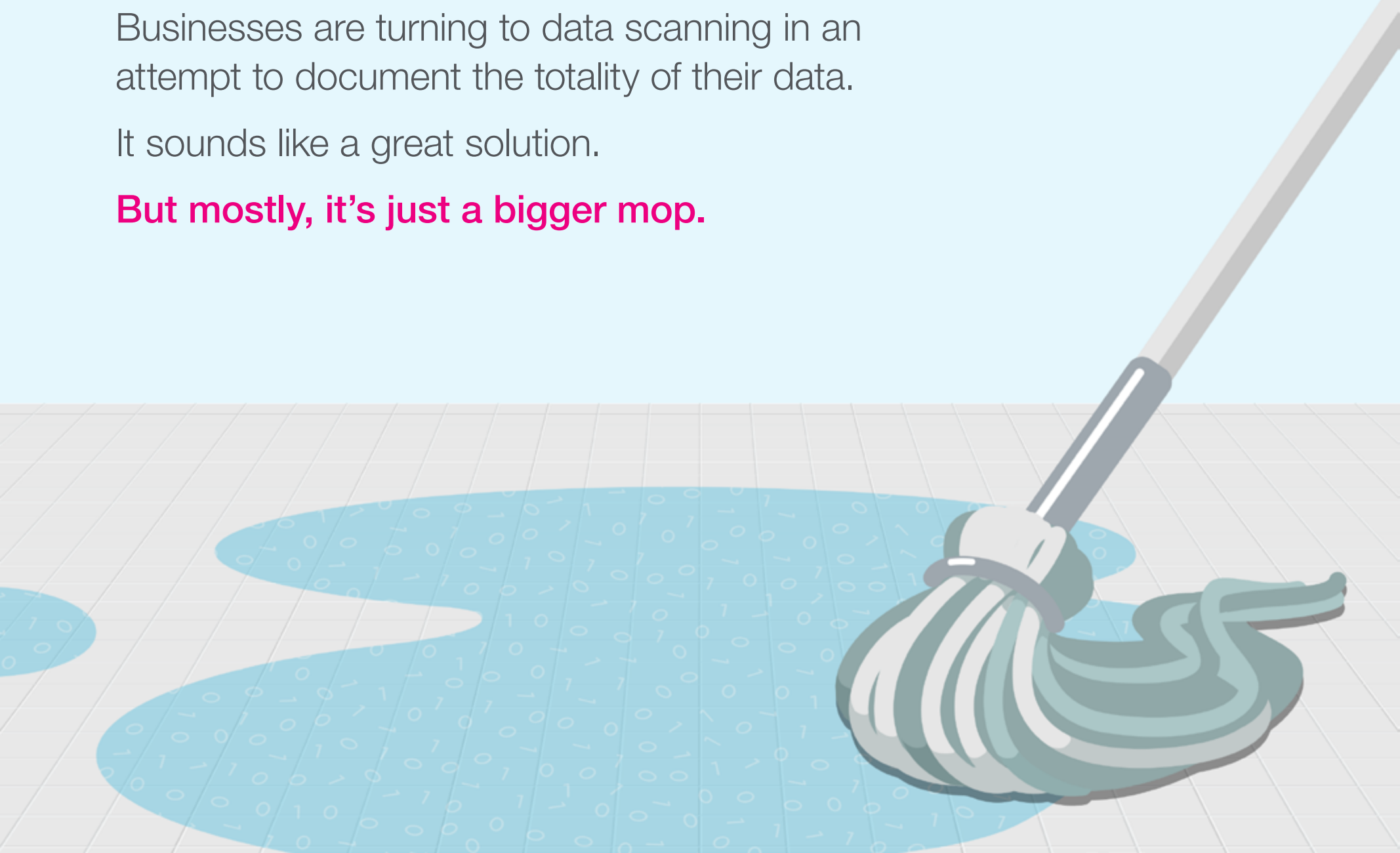
But it gets worse.



Businesses are turning to data scanning in an attempt to document the totality of their data.

It sounds like a great solution.

But mostly, it's just a bigger mop.





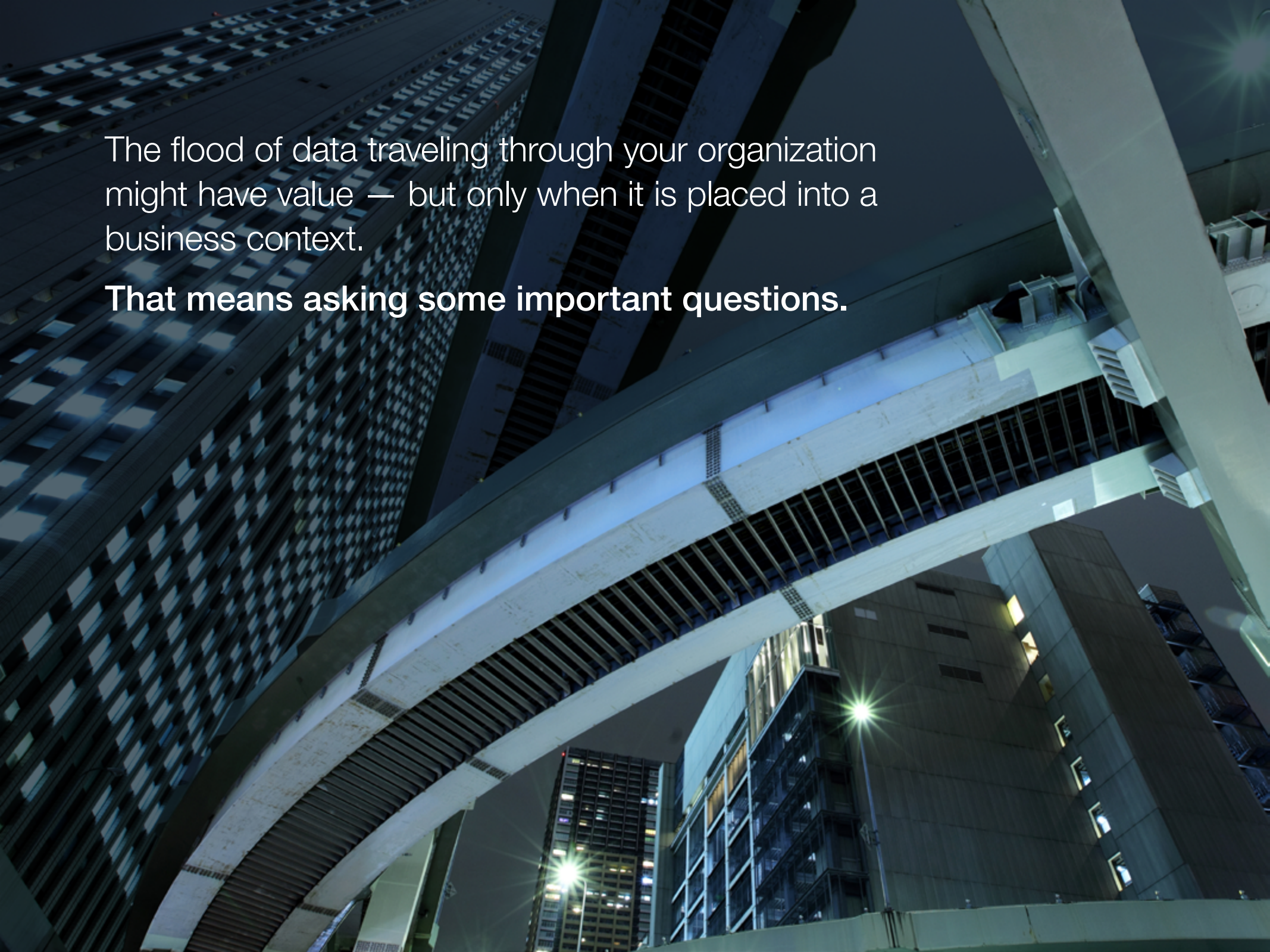
Data scanning delivers even bigger, more complex, hyper-detailed diagrams that require armies of interpreters.

And because it's a backward-looking process, data scanning can only ever capture where your data has been. It can't provide the tools you need to reach a future state.



Does that really help your business?

We don't think so. And that's why we believe it's
time for a paradigm shift.



The flood of data traveling through your organization might have value — but only when it is placed into a business context.

That means asking some important questions.



QUESTIONS LIKE:

What data is relevant to the business?

What will business users need to understand?

How will they find the information they need?

Will they trust it when they find it?

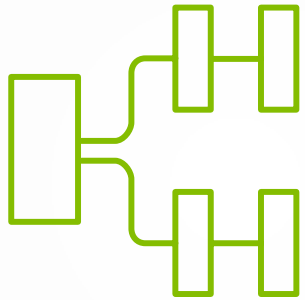
Those questions shouldn't be difficult to answer.

And when data lineage is driven by business needs, they aren't.

Because everyone knows who to turn to for answers.



And when data lineage becomes a usable tool, rather than a static archive, it provides business users with new ways to do things. They can:



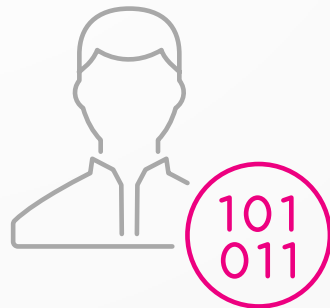
Initiate workflows



Interact with data




Take a snapshot at any point in the data flow



Identify data owners



Crowdsource new knowledge



Driven by business needs, data lineage becomes a governed asset—a starting point for collaborative conversation.

Because instead of wasting time pouring over painfully complex diagrams, your business users can see clearly where their data comes from, which business process it relates to, and how it changes over time.

101
011

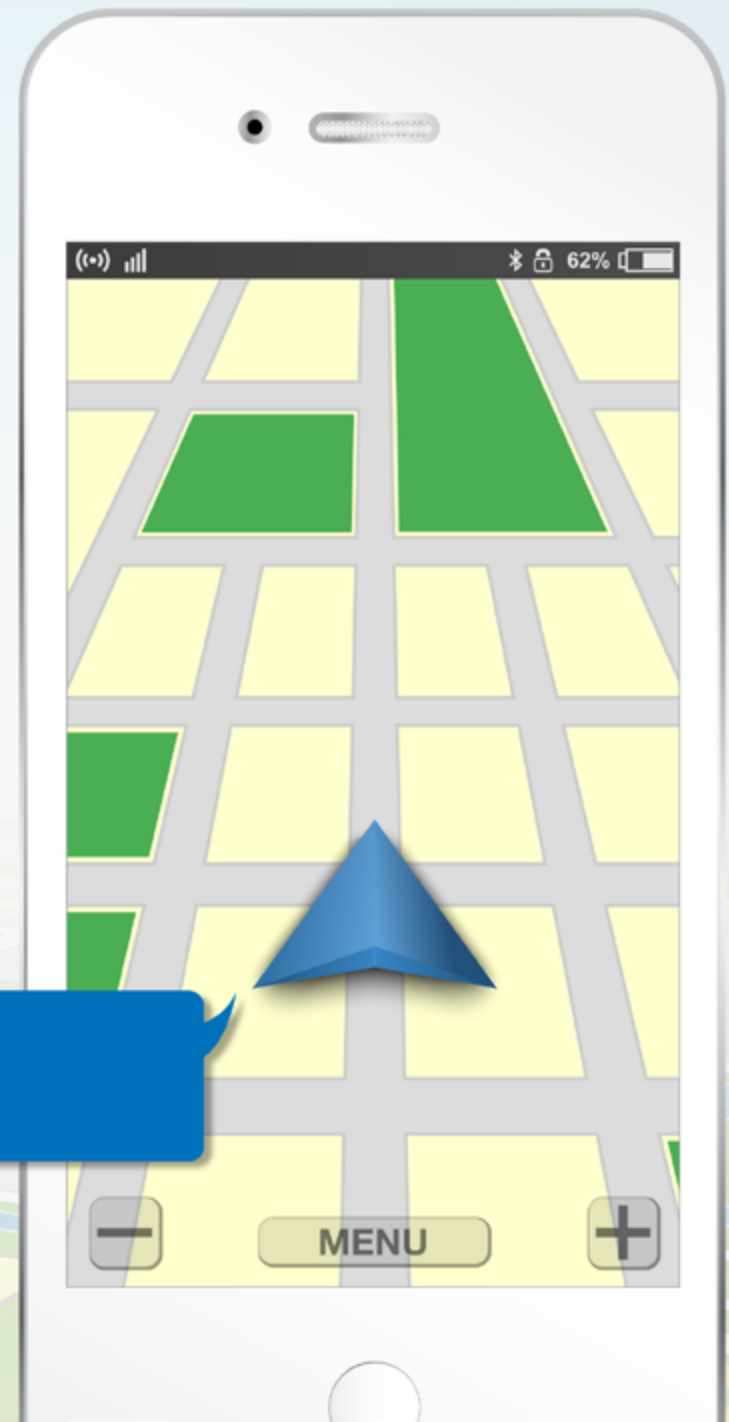


Instead of missing what the data might be telling you, your data citizens can rely on data lineage maps to identify the data that matters, understand its meaning, and trust its quality.

And instead of launching yet another data project to see where you've been, you can start thinking about where your organization should be going.

**That's the promise of governed data.
And we can help you get there.**

Where to next?





collibra®

©2017 Collibra

collibra.com

info@collibra.com

Follow Us

