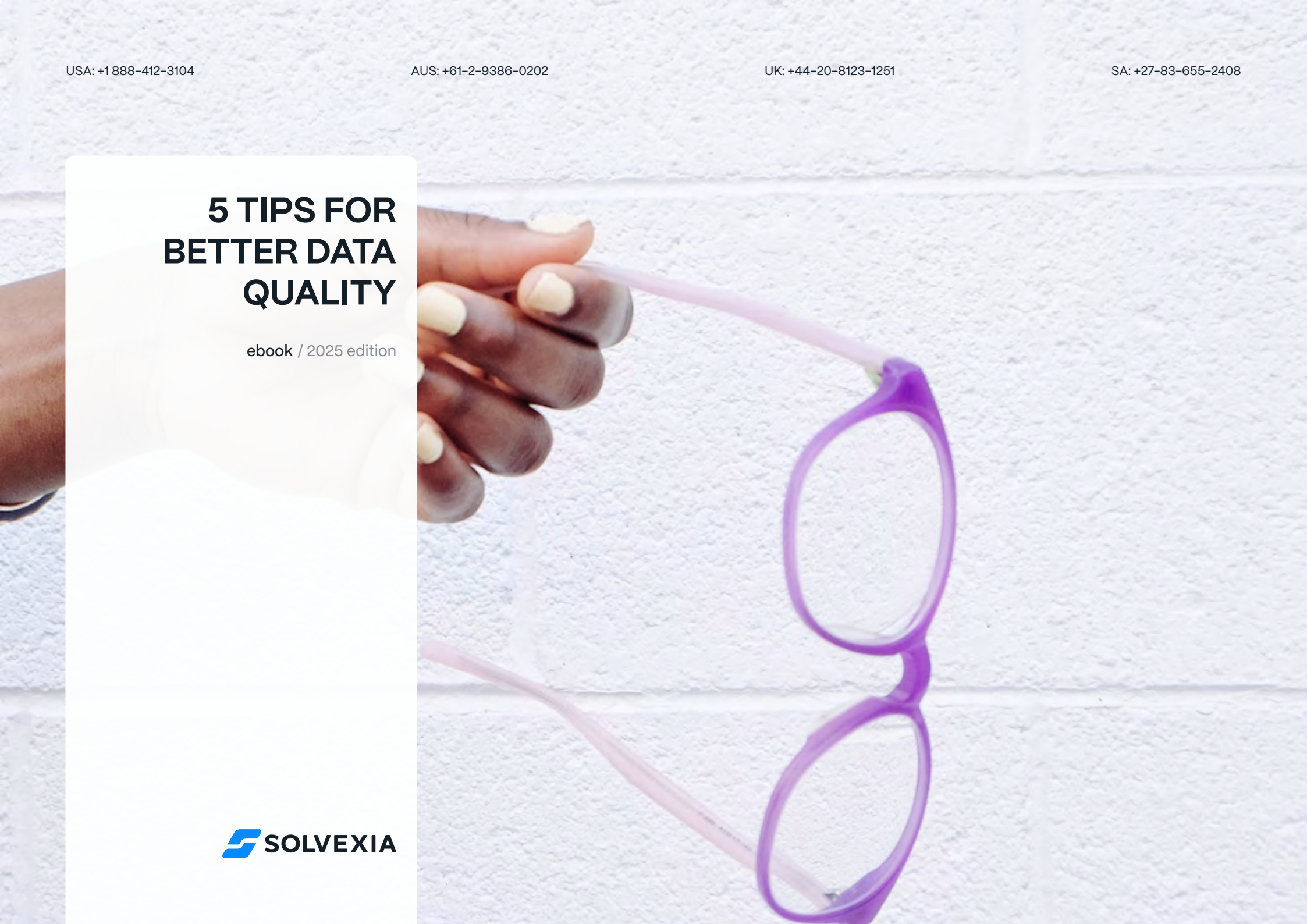
USA: +1 888-412-3104
AUS: +61-2-9386-0202
UK: +44-20-8123-1251
SA: +27-83-655-2408

5 TIPS FOR BETTER DATA QUALITY

ebook / 2025 edition



SOLVEXIA

GARBAGE IN, GARBAGE OUT

The objective of any BI or Analytics initiative is to equip your company with better decision-making capabilities. However, this is only possible if what you are presenting is trustworthy and credible.

In a perfect world, all the data we get would be clean, consistent and credible. Unfortunately, however, we need to deal with the reality of disconnected and siloed data, spread across hundreds of applications. We rely on data that is sometimes hard to or even totally out of our reach. And then, once we manage to get access to all our data, we grapple with the issue of bad data.

It is estimated that bad data costs the US economy $3 trillion a year and that up to 30% of operating expenses are wasted due to bad data.

So how do you tackle this solution?

Start by thinking about the journey that your data takes to get from a source to a dashboard. There are things you can do at each point in this journey to help improve the quality of your data:

Tip 1: Collect good data by getting as close to the source as possible.

Tip 2: Trust the data you collect by setting clear expectations for quality.

Tip 3: Find and fix issues in your data as early as possible. Use the concept of the 1–10–100 rule for data quality.

Tip 4: Continuously improve the quality of your data by measuring for quality and act on recurring issues.

Tip 5: Have a controlled and auditable way to make adjustments and changes to your data.

01

Get closer to your source data

How credible is the data that you are using to create your reports and analytics? Who is supplying the data? Is it coming straight out of a database or a source system or are you relying on data that has been manually prepared by somebody else in a spreadsheet?

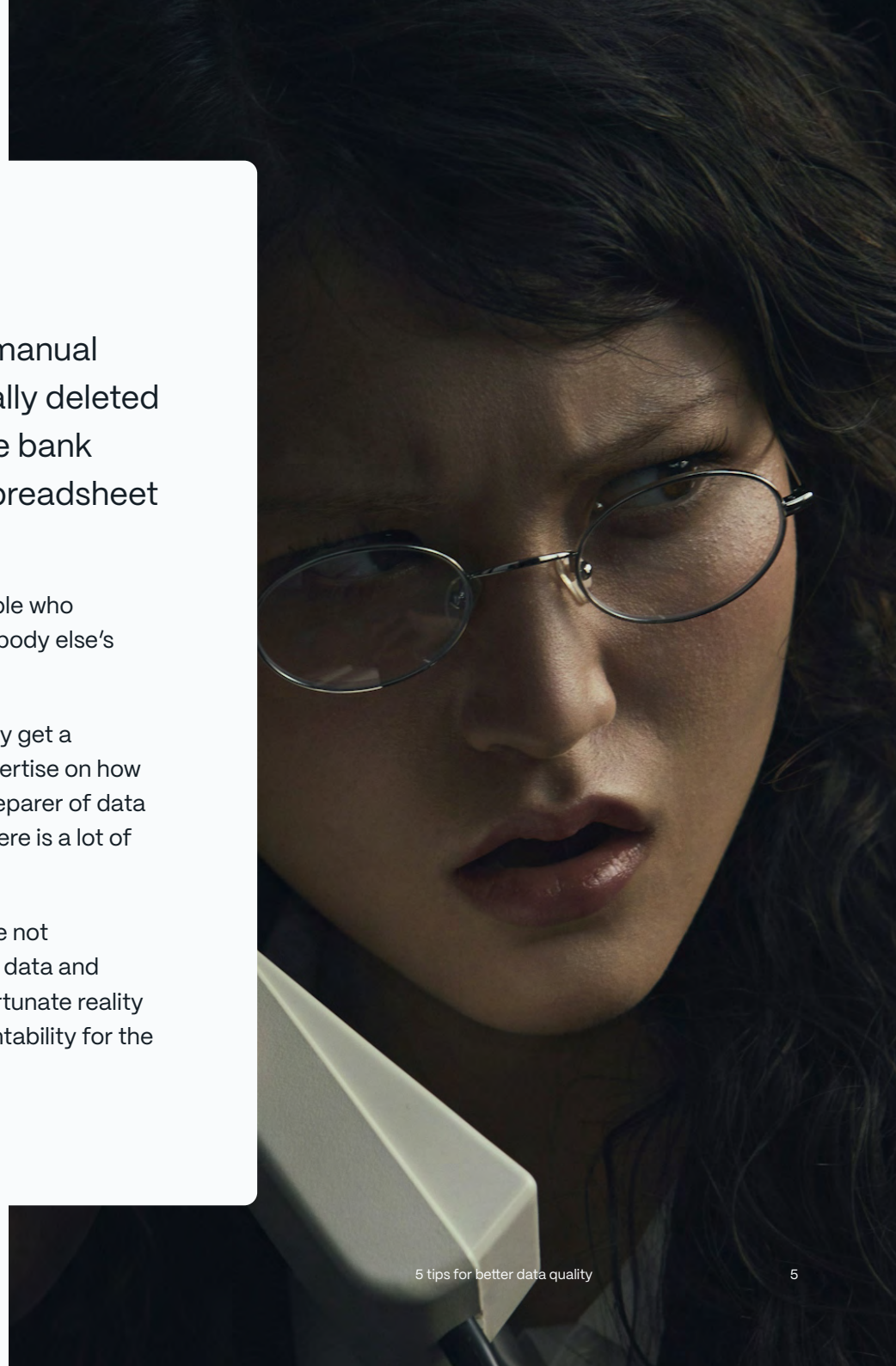
As a general rule, the less manual hands involved in your source data, the better. Wherever possible, you should seek data that is produced in an automated way.

The reason why this matters is simple. You don't want to be starting with erroneous data. Studies have shown that 90% of manual spreadsheets have errors. A person at Barclays Bank accidentally deleted cells instead of hiding them in their workbook, the error cost the bank millions. Similarly, JP Morgan made a cut and paste error in a spreadsheet that broke one of their risk models resulting in a $6 billion fine.

There is usually a separation between the producers and managers of data and the people who consume that data and turn it into information. If part of that data supply relies on somebody else's manual work, problems can arise.

First of all, the person supplying the data can make an error – that's the obvious risk. They get a formula wrong, they miss a step, etc. You also have the issue of staff turnover, so the expertise on how your data is prepared may degrade over time. You also need to think about where the preparer of data is getting their source data from? They could be using erroneous data themselves. So there is a lot of unpredictability in this scenario.

The implication for you is this concept of the "hidden enemy". These upstream issues are not necessarily going to be apparent to you. The risk that you could be working with this bad data and reporting to your boss for months on-end before you find out there is an issue. The unfortunate reality in this scenario is that your stakeholders are going to look to you for answers and accountability for the accuracy of your reports. So you really want to avoid this from happening.



Tips for getting closer to your source data

Seek out your data sources

With reporting and analytics, it's really common to inherit data sources from your predecessor or colleagues and just take them as a given. Your first action must be to question these data sources. If you have got data being manually prepared by someone else, put on your auditor's hat and go to talk to them. Quiz them about what they do and really try to get a sense of the risks to the quality of your data.

Ask data providers where they get their data from

It might be the case that you don't actually need their spreadsheet at all and you can go around them to the source system or data warehouse itself. But you won't know unless you take the time to really understand what they are doing to prepare your data.

If you can't go around them to the source data, put a system in place to lower your risk of getting bad data. Encourage that person to take the quality of their data seriously – for example, by educating themselves just like you are doing right now.

02

Set data quality expectations

Following on from our first tip, you need to set expectations on the quality of your data. This is important for the scenario where your data is being manually prepared. However, it's equally important for when you rely on data that comes from a system or data warehouse.

How frequently do you need data, is it every month, week or day? How old do you expect the data to be – i.e. when was the last time that the data was refreshed? How granular does the data need to be in order for you to get your work done? What format do you expect the data to be in?

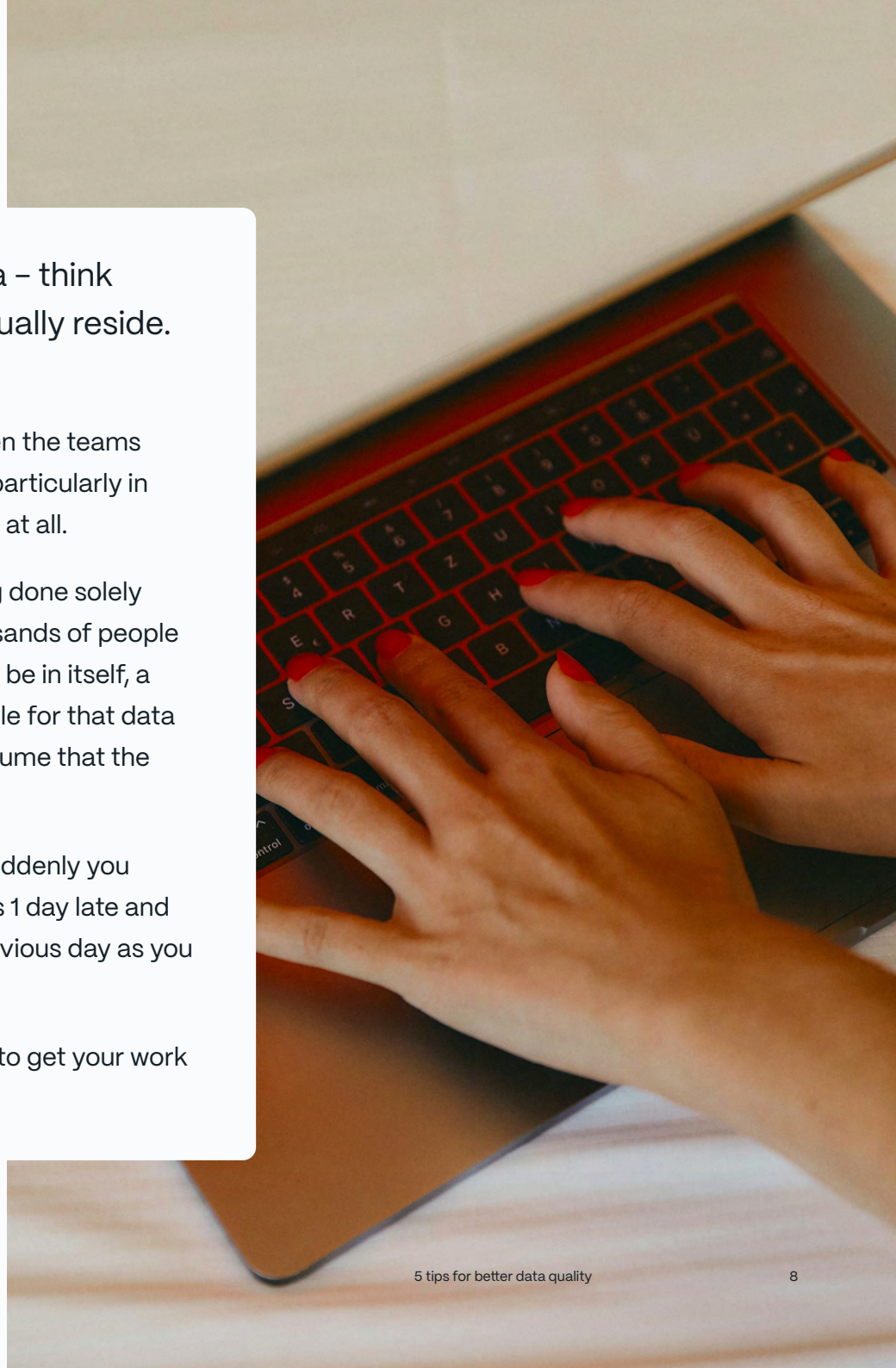
The average small business has around 14 different silos of data – think about these as 14 places where the data you may need can actually reside. Step into a large enterprise, and that jumps to 500 data silos.

To make use of these silos we rely on communication and understanding between the teams and people that run those silos. Unfortunately, however, communication is hard particularly in large companies where different teams may not directly interact with each other at all.

The first thing to consider is that the majority of data being prepared is not being done solely for your consumption. A data warehouse could be supporting the needs of thousands of people across your company. That spreadsheet being prepared manually could actually be in itself, a report that goes to a management team. Simply expecting the person responsible for that data to know or even care about your specific needs isn't reasonable. You have to assume that the provider of your data needs to be educated on what you are using their data for.

Imagine you have received an Excel workbook for the last 12 months and then suddenly you receive a CSV file, that has a slightly different column format. Or, the data arrives 1 day late and only has data to the end of last month, instead of it being refreshed up to the previous day as you expect and need.

Each of these issues becomes a disruption for you. Suddenly, you're scrambling to get your work done to meet your own deadlines, introducing more error risks and so on.



Tips for setting data quality expectations

You need to communicate with the suppliers of your data. If they don't know you, how can you expect them to care about what you need from their data? Specify your expectations for:

- **Accuracy.** How well does this data reflect reality?
- **Completeness.** Is it comprehensive? Does it give you everything you need to get your reporting done?
- **Consistency.** Are you being given data that matches what is being shared with others?
- **Timeliness.** Is the data up to date?
- **Validity.** Does the data follow a specific format and your business rules?
- **Uniqueness.** Is the data free from duplicate records? Is there something in the data that identifies a unique record?

Write these expectations down and share them. What you are looking to establish here is a set of guiding rules that both you and the supplier of your data understand.

03

Catch issues early

If you think about the journey that your data takes from its original source, all the way through to the final output, being a report or dashboard, there are lots of touch points along the way for things to go wrong. It's really important that you catch these issues as early in the supply chain as possible because the earlier you find them, the cheaper and faster it will be to resolve.

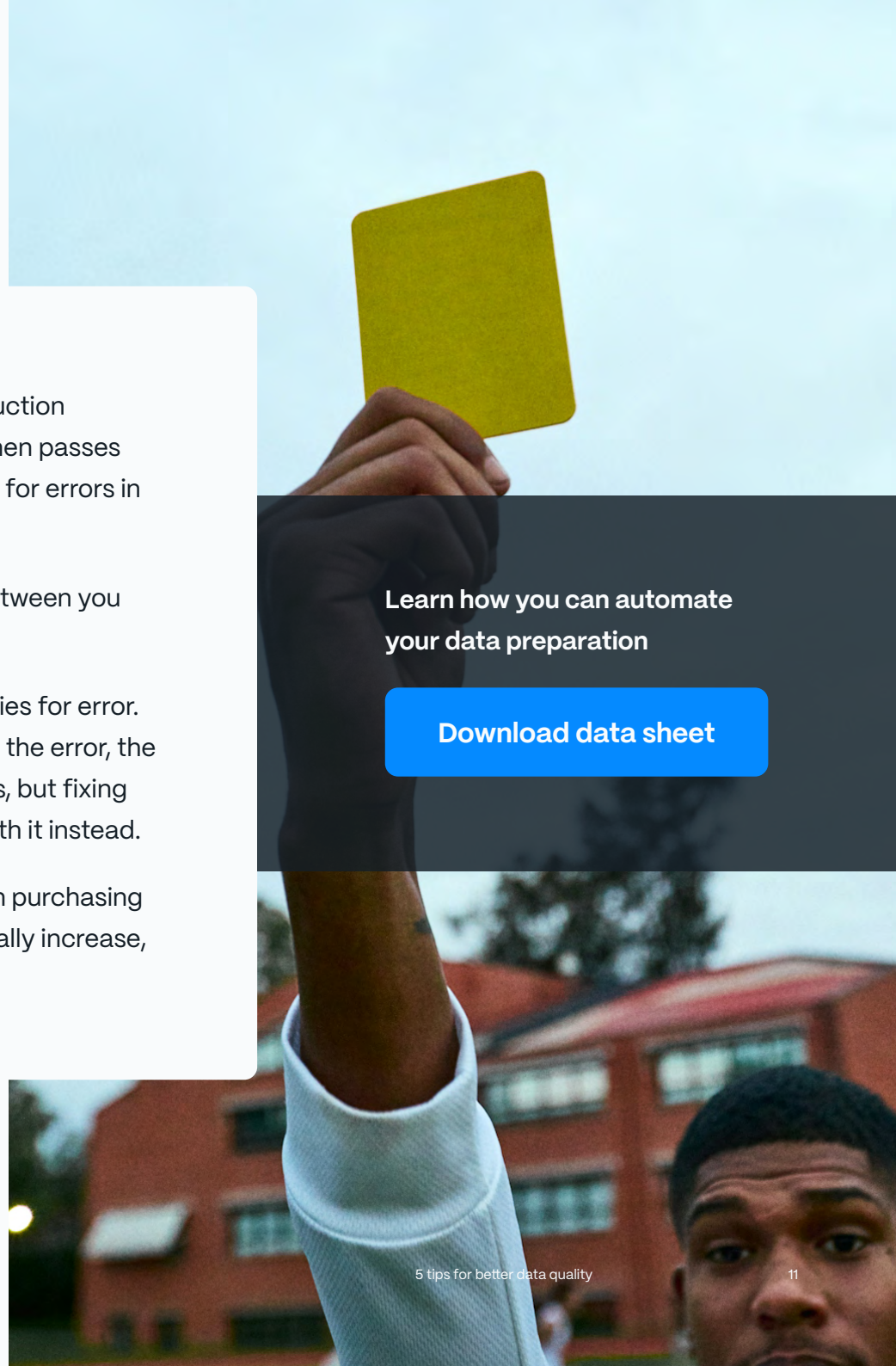
This is where the 1–10–100 rule provides us with a really good and simple principle to follow. So let's break this rule down and talk about how it will help you.

The first thing to note is that errors can and do occur at every point in your production line. The average error rate for data entry into systems is around 1%. Your data then passes through your systems and perhaps a data warehouse. Well, the industry average for errors in data warehouses is another 1 – 3%.

The data then somehow makes its way to you. If there is a third party that sits between you and that data, that's more opportunity for error (as we talked about previously).

And then finally you take that data and work with it, introducing more opportunities for error. What the 1–10–100 rule is arguing is that the earlier in this journey you find and fix the error, the cheaper it will be for your company in terms of remediation. It may seem obvious, but fixing errors is not fun and it's really tempting to let the next person in the chain deal with it instead.

A simple analogy is that of a customer who incorrectly enters their address when purchasing an item from an online store. If not corrected, the cost of the error will exponentially increase, until eventual failure, when the package is sent to the wrong address!



Tips for implementing 1–10–100 rule

The "1" relates to error prevention. So these are activities that can prevent errors when data is originated. You may need to work with your IT to define better controls and measures for example by automating the entry or checking off that data.

The "10" relates to remediation. Set up robust validations for data quality in your reporting process. This is where data automation is truly an asset for Finance because it means you can do what might take you 4 hours of data validation in 2 minutes.

And while it is hard and boring to perform these checks, the thing to remember is failing to do these checks means you push the errors to the "100" stage, which is a failure and really, this is the worst-case outcome for everyone involved.

04

Measure and act on data quality

So we've established the need to set expectations on the quality of your data and the importance of finding and fixing issues as early as you can in your data journey. That's all good and well, but none of it really matters unless you put all of this into action. That brings us to a critical, one–two punch. First, you need to measure for data quality and second, you need to act, that is, fix the issues that occur.

If you're thinking, why me? Why do I have this overwhelming burden of finding and fixing data, you're definitely not alone. According to Forrester, nearly 1/3 of people say they spend almost half their time vetting and validating data before it can be used for their analytics.

Tips for measuring data quality

Refer back to the 6 dimensions of data quality including accuracy, completeness, consistency, timeliness, validity and uniqueness. The goal is to put metrics and rules in place that help you pass data through these 6 dimensions and in doing so produce exceptions that you need to fix.

Create an assessment checklist for each of the six dimensions. Examples:

- **Completeness.** A measure for this might be the absence of blank rows or key columns being empty. This is something that can easily be measured by counting the records that have this issue and dividing it by the total record count.
- **Uniqueness.** This might involve checks that can be performed against other trusted data sets. For example, the number of products in your source file, does it genuinely match the list of unique products being sold by your company?
- **Timeliness.** This can be measured if your data contains some kind of date/time field. You can easily compare this to the current day to get a sense of how timely the data actually is.
- **Validity.** This can be checked in a variety of ways. Does the data have the expected number of columns? Do the column headings match what you expect? Does a date field contain only dates etc.?
- **Accuracy.** This is where your expertise and judgment really help. So can you compare data to a "real world" truth? Historical data and trends can help – for example, to identify a sudden and odd spike in sales for a region which may, in turn, uncover an issue.
- **Consistency.** This can be achieved by comparing data across different files and sources. For example, does the average age of your life insurance policyholder match the recorded average age based on what is in your policy administration system?

Define the measures for each dimension that is relevant to you. You then need to put processes in place to actually do the checking against these metrics which in turn will produce errors and exceptions for you to correct.

05

Allow for change in a controlled way

As explained in this eBook, you start with expectations, then you put a system in place to measure against your expectations and that leaves you with fixes and changes you need to make to your data.

Some of these fixes may be able to get pushed back to the source system. However, many fixes will need to be done by the business. This in itself introduces more error risk, which needs to be mitigated.

In a perfect world, all your fixes will be systemized, but this isn't necessarily feasible either due to the lack of agility. So the key is a balance between agility and control, which you can do by understanding and avoiding some of the common mistakes people are likely to make.

Tips for controlled and agile change

Never modify original or source data. Imagine you have a table of data in a spreadsheet and there are errors in one of the columns. Your first instinct might be to type over those cells containing errors. Instead, create a safe and controlled space for making adjustments to the data through new columns. That way, if anybody else looks at your data, they can immediately get a sense of what you got as an input and what you produced as an output from your work.

Clearly label your adjustments and changes. Columns need to have headers. A common issue we encounter is the use of color in a workbook to denote the changes that have been made with no key or reference to what those colors mean. This is actually counterproductive for the people who are trying to make sense of your work.

Document the changes you are making. Add commentaries and explanations as to why you are changing the data and ensure that you have an audit log of every change that you've made. Lots of data automation tools, like Solvexia, will automatically record this for you.

Standardize your workflow for how changes are made. This includes how you organize your files and data, how you go about making changes and documenting each change. The goal needs to be on repeatability and consistency.

You also need to consider that some changes are ad-hoc and may never need to be done again while other changes may keep coming up over and over again. It's those recurring fixes that need to be addressed in a more systemized way either by having data changed back at the source where possible or by using automation.

WRAP UP

Bad quality data is making it hard to win when it comes to BI and analytics projects. The reality is that siloed and disconnected data is not going away anytime soon, but there are 5 simple ways to improve the quality of your data.

First, give yourself the best start possible by getting as close to your source data as possible.

Follow the 1–10–100 rule of data quality and really focus on finding errors as early in the data pipeline as possible.

Have a simple assessment checklist to compare your data against the expectations that you have set.

Finally, when it comes to changing your data, ensure you have a clear separation between source data and manual adjustments.

SOLVEXIA

Learn how Solvexia can improve your data quality

Download data sheet

SOLVEXIA

Solvexia is a low-code platform that automates spreadsheet-driven data preparation and manual processes.

It enables finance to free up their time by running processes 100x faster with 98% fewer errors, allowing businesses to reallocate essential resources to value-added work.

Finance and accounting teams choose Solvexia to unify data from systems and spreadsheets and automate their processes without involvement from IT. Its drag-and-drop interface makes it fast to deploy and easy to learn.

Solvexia combines and manipulates data, performs calculations and creates interactive reports, analytics and smart dashboards for richer insights for your stakeholders.

You can leverage Solvexia to automate hundreds of processes for your company, including reconciliations, revenue and expense reporting, regulatory compliance, rebate management and much more.

100x
faster processes

Positive ROI within 6–12 months

Modern cloud architecture

Works with existing systems and spreadsheets

Deploy processes in **minutes**, not months

98%
fewer errors

SOLVEXIA

Ready to overcome your data quality challenges?

Book a demo to see what Solvexia can do for you.

Book a demo today