

ACHIEVING REGULATORY COMPLIANCE WITH DATA LINEAGE SOLUTIONS

An Industry Perspective Report brought to you by
ASG Technologies and FIMA

This report is based on recommendations made by the industry experts who participated in ASG's live webinar in the Summer of 2016.



SETTING THE SCENE

In the Summer of 2016, ASG and the FIMA conference series partnered to produce a webinar focused on addressing the real questions that arise when working on data lineage projects. *Why is lineage important? How are people approaching the creation of a data lineage analysis? What are reasons it should potentially be automated? What does automated analysis bring in terms of accelerated compliance?*

This report uncovers the answers to these questions, as identified by the several industry experts listed below, and includes their recommendations on building out your own data lineage projects.

PANELISTS



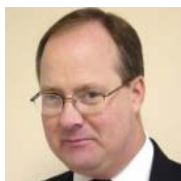
THOMAS SERVEN

Vice-President of Enterprise Data Governance
State Street



FRED ROOS

Director of ICAAP Transformation
Santander Holdings US



IAN ROWLANDS

Vice-President of Product Marketing
ASG Technologies



What does the current data and regulatory environment look like? ▲



IAN ROWLANDS: There are several factors today adding to a heightened focus on data lineage. One of the most prominent is change in the regulatory environment. In the past, you may have only needed to respond to one new regulation a year, and been audited at most once a quarter. It would have been likely that your IT

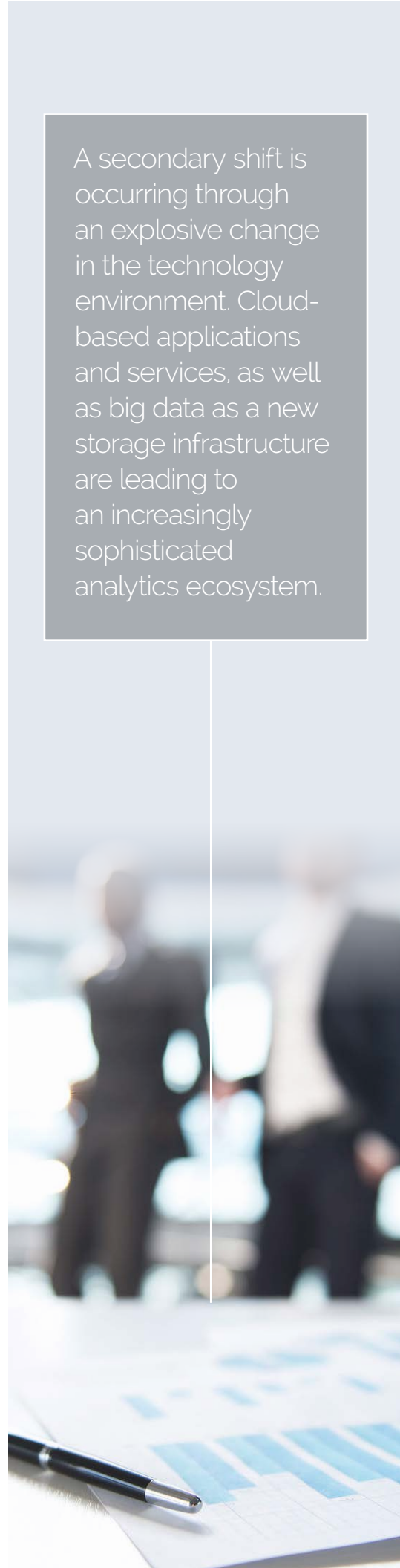
team knew more about the information that they were being asked for than the regulators themselves.

It's a lot harder now. There are overlapping regulations, and their number is constantly growing. Audits now occur more frequently as well, and you can even be asked to respond to challenges on demand. Regulators and the auditors are getting smarter, and they are helping us to do a better job, but at the same time that means that the challenge of meeting requirements has gone up.

A secondary shift is occurring through an explosive change in the technology environment. Cloud-based applications and services, as well as big data as a new storage infrastructure are leading to an increasingly sophisticated analytics ecosystem. This is another two-edged sword; while we can do many more interesting things with the data that we have, keeping track of the data and making sure that information is credible is increasingly difficult.

We must keep pace with conflicting pressures; the pressure to respond more quickly and more accurately to compliance demands, and on the other hand, a much more complex data environment making it harder to do the job. All of this in a context where there's increasing executive pressure. While to a certain extent, that pressure has always been there, it is interesting how often we're now hearing the request to shift from defensive data intelligence to offensive data intelligence. Executives don't just want to be told how better data management can keep them out of trouble. Now they want to know how spending money on it actually improves the bottom line.

A secondary shift is occurring through an explosive change in the technology environment. Cloud-based applications and services, as well as big data as a new storage infrastructure are leading to an increasingly sophisticated analytics ecosystem.



Why is lineage important? ▲



THOMAS SERVEN: I think there are four significant factors that lineage contributes to. The most important is the **establishment and proper use of trusted sources**. Most data governance programs have a goal of establishing a single version of the data "truth" and encouraging people to use that single version. Lineage allows

us to be able to understand the data lifecycle better, helps us to determine where trusted sources can benefit, and how we're doing in connection with their adoption.

Another multifaceted area is **data quality, with one of the most recognizable aspects being root cause analysis**. When a data quality issue occurs, it's often the case that the issue started several steps up in the lifecycle. In order to do a proper root cause analysis, having a lineage solution is an incredibly helpful tool.

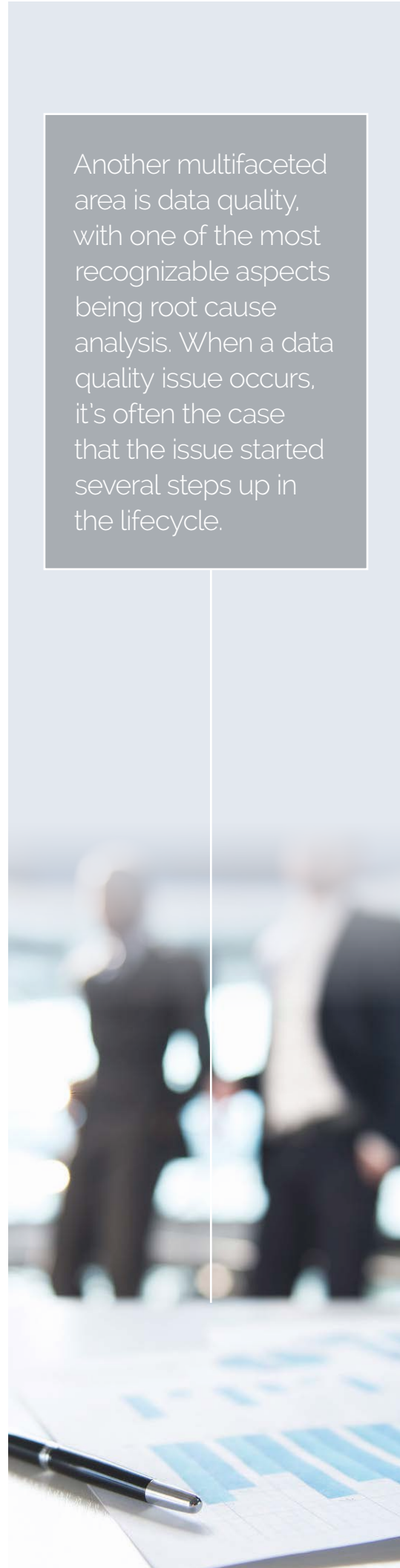
Lineage is also important in the **establishment of data ownership**. As data goes through changes of custody, each owning entity can be identified by looking at data lineage, from the point of creation to the point of final consumption.

A last point that doesn't get enough consideration is around **reproducibility or data reuse**. This is the idea that data may have restrictions on it for certain purposes. Those restrictions may be contractual; they may be based on a market data provider or other privacy considerations. Understanding the origin of that data helps us to understand what some of those reproducibility and reuse issues may be.

“

"Most data governance programs have a goal of establishing a single version of the data "truth" and encouraging people to use that single version. Lineage allows us to be able to understand the data lifecycle better, helps us to determine where trusted sources can benefit, and how we're doing in connection with their adoption."

Another multifaceted area is data quality, with one of the most recognizable aspects being root cause analysis. When a data quality issue occurs, it's often the case that the issue started several steps up in the lifecycle.



What are some differences in a manual versus an automated approach to data lineage? ▲



FRED ROOS: Better data management practices should start from a pilot program. Begin with one particular area within an institution that is well-understood, assess what the current maturity level is from a data perspective using a data maturity model, then assess where all the data comes from using a manual perspective. Once you have a

baseline in terms of manual lineage you can start thinking about automated lineage as the next step, as you're able to look at what your manual lineage was able to produce as a baseline for comparison.

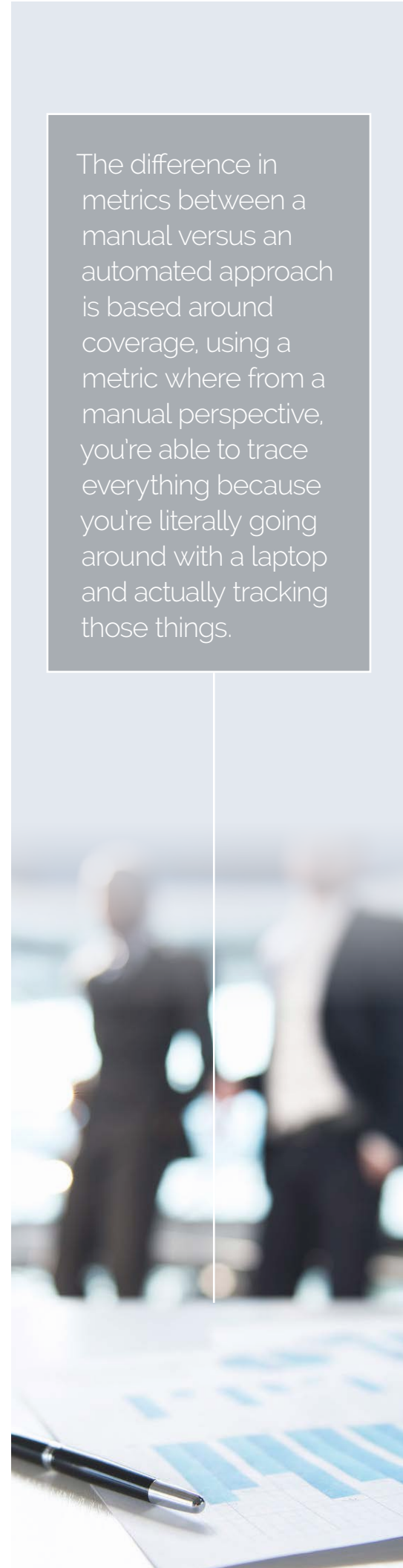
The difference in metrics between a manual versus an automated approach is based around coverage, using a metric where from a manual perspective, you're able to trace everything because you're literally going around with a laptop and actually tracking those things.

Automated lineage is interesting because you're able to see what a physical tool can track, what it can't track and why you can't track it. *Was it because something was written in a different language that your automated lineage tools weren't able to pick up? Was it because you now found a weakness in terms of how something is tracked?* In this way you can uncover valuable insights.

“

“Once you have a baseline in terms of manual lineage you can start thinking about automated lineage as the next step, as you're able to look at what your manual lineage was able to produce as a baseline for comparison.”

The difference in metrics between a manual versus an automated approach is based around coverage, using a metric where from a manual perspective, you're able to trace everything because you're literally going around with a laptop and actually tracking those things.



How are you currently managing your data lineage activities? ▲



THOMAS SERVEN: There are a couple of important questions to consider when beginning a lineage collection process. A crucial consideration is that these initiatives can become very large, very fast. There is an almost endless amount of lineage information to capture in a large organization. Understanding, the scope of

the effort, what the drivers are, whether it's regulatory in nature or being driven by a business need is an important first step. Other considerations are whether you are talking about lineage for data elements, for information assets, reports, extracts, or databases because lineage could be at the element level or at the database or extract level as well as a combination of the two. In our case, it is.

Structured and unstructured data demand different approaches. Structured data lineage is fairly precise. You can take about five attributes to that structured data, from a so-called five-part key and be able to uniquely identify that specific data element. Obviously, within unstructured data, it's a very different exercise. *Are you going to identify critical data elements as part of this process? And if so, what is a critical data element?* We've had robust conversations about how to identify and define critical data elements. Ultimately, it's in the eye of the consumer.

Another piece of this process that doesn't get enough discussion is making sure that whatever you're capturing can be integrated into your business metadata as well. If you have a lot of isolated information about where data came from and how it may have been transformed, but you can't apply things to it like the legal entity or the business unit or the asset type that it refers to, then you can lose the ability to do a lot of important risk data aggregation reporting. You need to be able to take lineage information and integrate it into the rest of your metadata and to your repository.

In terms of automated versus manual lineage data, we currently use a combination of the two. We have some tool sets that inherently have metadata associated with lineage built into them, so we have database appliances, we have extract, transform and load (ETL) tools that have that information readily available so we can harvest that metadata and bring it into our metadata repository fairly seamlessly. In other cases, that information has to be documented manually. You need to be realistic about what can actually be brought in on a manual versus automated basis, striking the right balance between the two.

Structured and unstructured data demand different approaches. Structured data lineage is fairly precise. You can take about five attributes to that structured data, from a so-called five-part key and be able to uniquely identify that specific data element. Obviously, within unstructured data, it's a very different exercise.



Why should we automate lineage analysis? ▲



THOMAS SERVEN: One simple answer is scalability. If you look at the vast amount of data for which it might be desirable to compile and store lineage, it quickly becomes a daunting task to be able to manually capture and maintain all this information going forward. It's important to think about how a process can scale over time

and the role that automation is going to play in that, bearing in mind that a certain portion is always going to be manual.

Automation also frees up time to spend looking at exceptions and verification. Automation doesn't necessarily mean that something is going to be perfect the first time; you're still going to need human beings to be able to monitor that process and verify that it's correct. The more human beings you can free up to do a verification process, exception management, and work on the manual side of things, the better, and that creates a scaling effect.

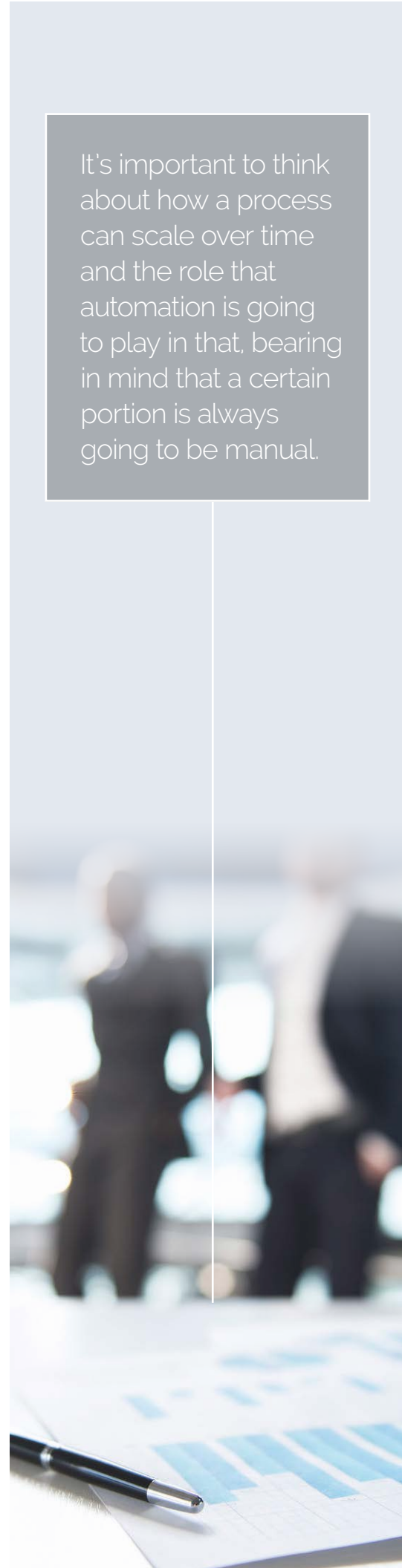
The other benefit of automation is that once you have something running effectively and you start to trust it, the reliability is likely going to increase over what human beings would do as well. There is a popular viewpoint that good automation will in fact increase accuracy and play well to regulators.



FRED ROOS: As part of the development process, you're able to understand where all of your information is coming from. Automation becomes a powerful tool for your data stewards and data owners to be able to verify the origins of your data.

In terms of future development of source systems and additional applications, automation provides your IT folks a lot more visibility into what the implications of change might be from both a downstream and upstream perspective on regulatory reporting, stress-testing, and business continuity perspective if you go into replacing an existing data mart or warehouse.

It's important to think about how a process can scale over time and the role that automation is going to play in that, bearing in mind that a certain portion is always going to be manual.





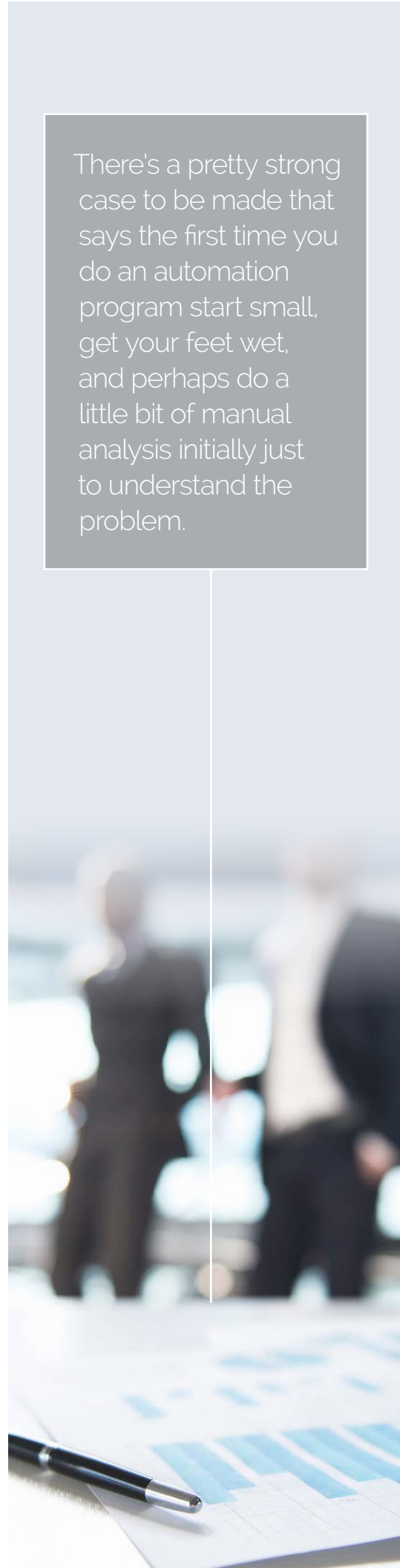
IAN ROWLANDS: In the past I've worked with a client who was not initially doing any automated lineage analysis and would make changes to reports, later finding out that the changes they were making were impacting downstream systems and generally leading to rework and repetitive expense. When they actually moved to lineage analysis upfront, it served as a downstream impact analysis as well as a lineage analysis and they eliminated that repeating expense.

There's a pretty strong case to be made that says the first time you do an automation program start small, get your feet wet, and perhaps do a little bit of manual analysis initially just to understand the problem. After you lay the groundwork, then you can industrialize and get to a point where you're automating what can be automated, with an exception management process in place that enrolls data stewards in resolving issues that arise.



"Automation becomes a powerful tool for your data stewards and data owners to be able to verify the origins of your data."

There's a pretty strong case to be made that says the first time you do an automation program start small, get your feet wet, and perhaps do a little bit of manual analysis initially just to understand the problem.



How mature is data lineage analysis today? ▲



FRED ROOS: The maturity of data lineage analysis will vary according to the institution. The scale issue goes from one to five, five being fully optimizing your current processes. I think there are very few firms out there which are five. Very well-managed institutions are typically around three or four, and I think most are still either at one or two.

Maturity is a function of the regulatory scrutiny that a particular institution has, how close their data is to the new generation of a particular firm, and lastly the stability involved with a particular institution. For example, a very large financial institution that had gone through several mergers and acquisitions during a general financial crisis would find that their data lineage is much harder, and the process to do it might not be as mature. Other organizations which are relatively stable are plugging in a more mature data lineage analysis framework.



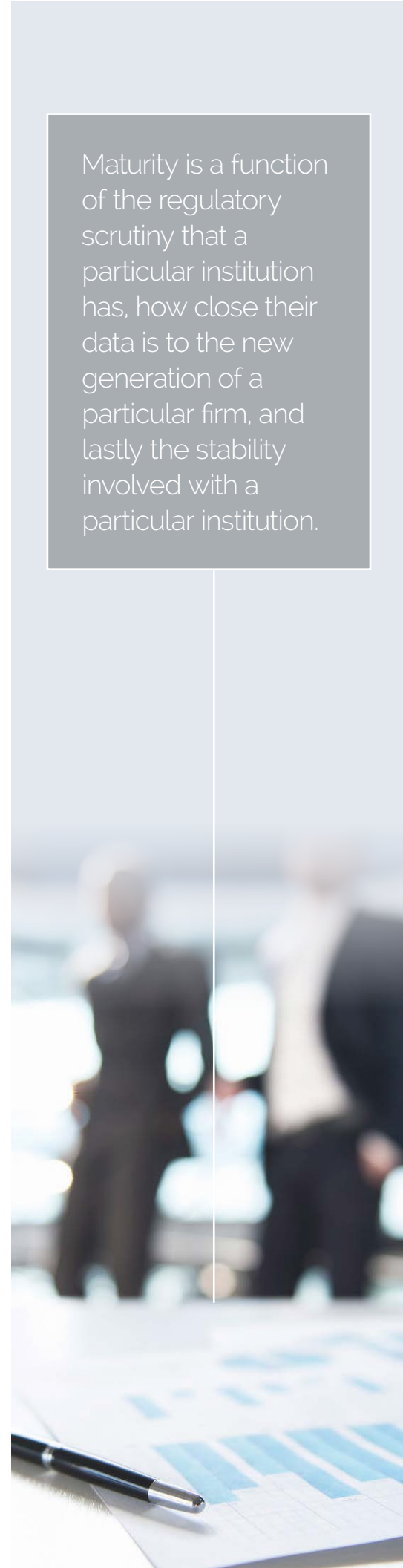
THOMAS SERVEN: Movement towards lineage automation is a maturing process for most financial services organizations. This has been the result of relatively recent regulatory phenomena that have pushed folks in this direction. I think everybody is learning about the best practices in the States and sharing information so the goals at

the highest level are to make sure the right lineage metadata is being collected, that it's in an automated tool that's going to be available to all stakeholders and able to be integrated into other processes like risk concentration reporting, and ultimately having collection become automated as well.

Maturity is a function of the regulatory scrutiny that a particular institution has, how close their data is to the new generation of a particular firm, and lastly the stability involved with a particular institution.



"Maturity is a function of the regulatory scrutiny that a particular institution has, how close their data is to the new generation of a particular firm, and lastly the stability involved with a particular institution."



KEY RECOMMENDATIONS

Use the following key recommendations as your roadmap to a successful data lineage program.



DEMONSTRATE SUCCESS EARLY BY STARTING WITH A RELATIVELY SMALL AND WELL-DEFINED SCOPE that has a relatively large impact in the institution; demonstrate the pilot is scalable for other data sources or other areas of your business. Then make the case with senior executives that it's not just a compliance project; it's really a whole new way to look at data and where data comes from and manage it within the institution.



THINK ABOUT HOW THAT LINEAGE NEEDS TO BE INTEGRATED WITH THE REST OF YOUR BUSINESS METADATA to give you the ability to be able to slice and dice for organization, legal entity, and other factors.



UNDERSTAND THE ROLE OF TECHNOLOGY AND CREATE A TECHNOLOGY ROADMAP but be careful not to create roadblock areas where you're wholly dependent upon the implementation of technology.



DETERMINE HOW THAT LINEAGE PROCESS IS GOING TO BE BLENDED OR INTEGRATED INTO THE SDLC OF YOUR ORGANIZATION on a go forward basis. Those developing software and creating databases need to be very aware of how lineage is going to be captured and maintained on a go forward basis to be successful.



ABOUT ASG TECHNOLOGIES



technologies

ASG Technologies provides peace of mind to each of our customers. With proven solutions for information access, management and control, ASG enables greater insight to drive intelligent decisions from legacy to leading edge environments.*

As a company, ASG is dedicated to delighting every enterprise that partners with us.

With the right software, your entire company becomes more competitive. We've seen over and over how companies with a smarter IT footprint have improved not only IT efficiency, but also business productivity and cost savings.

In fact, more than 70 percent of global Fortune 500 companies trust us to optimize their existing IT investments. For nearly 30 years, we have been fortunate enough to work with some of the world's most successful companies, such as American Express, Coca-Cola, Daimler, General Electric, HSBC, IBM, Lockheed Martin, Merrill Lynch, Procter & Gamble, Sony, Verizon and Wells Fargo.

As your solutions partner, we predict and deliver on your individual needs to help address change in your business processes. Our flexible solutions offer speed, scalability and the best price for performance in platform-agnostic packaging. Regardless of your infrastructure, ASG can provide you with peace of mind.


Learn more at www.ASG.com

ABOUT WBR DIGITAL



WBR Digital's team of content specialists, marketers, and advisors believe in the power of demand

generation with a creative twist. With senior executives from medium-sized businesses and Fortune 1,000 companies attending more than 100 WBR events each year, we are uniquely positioned to energize your organization's marketing campaigns with a full array of marketing and bespoke content services.

 WBR is proud to be virtually paperless. Together, we're saving trees, helping the environment and are looking to improve every single day.

