

Procurement Leaders

Intense Focus on Data Quality:

Why Predictive Procurement Orchestration Holds The Key To Trustworthy Procurement Data

Written by Bertrand Maltaverne, Senior Analyst

Spend Matters™

INTRODUCTION

Imagine that one day your bank statement shows that you have received a large sum of money, money that you were not expecting. You will probably ask yourself: what is this money? Where is it coming from? What do I do with it?

These are probably the same questions that Sho Taguchi, a citizen of the municipality of Abu in Western Japan, asked himself in May 2002 when he received a transfer of 46.3m yen (approx. \$360,000). Over the course of two weeks, Taguchi withdrew the money – to gamble.

The truth is that Taguchi was unemployed and, therefore, belongs to one of the 463 low-income households in Abu slated to receive 100,000 yen each as Covid-19 support money.

But, instead of wiring 463 times 100,000 yen, a municipality official wired the total amount to the first name on the list of 463 names. That name was Taguchi's.

After realizing the mistake, the municipality asked Taguchi to return the money which he said he would do, after initially refusing. And, the other 462 households were sent their 100,000 yen as was originally intended. So, the relief fund ended up costing Abu's municipality twice the amount it had planned for a routine, budgeted expenditure of unemployment support.

Contents

Introduction 1

Procurement has a data problem 2

Procurement's data problem is nothing new, but it has now become more urgent 3

Untrustworthy procurement data has costly business consequences 3

Data proliferation makes the problem more painful and costly 5

A new approach is required needed to orchestrate trustworthy procurement data 6

The traditional approvals-based approach to procurement data validation does not work 6

The "live data validation" approach: Introducing Predictive Procurement Orchestration 7

Conclusion 8

This story is anecdotal, but it nevertheless raises important questions about people, process, technology and data:

- Why did the official not realize their mistake?
- Why wasn't the system able to detect the error when the data for the transfer was entered?
- Why did the mistake stay unnoticed for so long?

These questions are the same questions that often come up in the context of procurement data. Too often, procurement leadership only discovers a costly error in hindsight, after it's already too late to take a corrective action. And this challenge of backwards-looking manual validation is at the heart of why procurement leaders have now become intensely focused on the topic of data quality. In this white paper, we'll examine the underlying challenges that procurement teams face when dealing with data, the traditional approach to data validation and a new technology solution called Predictive Procurement Orchestration (PPO) that offers a different approach to addressing the problem of untrustworthy procurement data filled with potentially costly errors.

TRUSTWORTHY DATA IS A PROBLEM FOR PROCUREMENT TEAMS

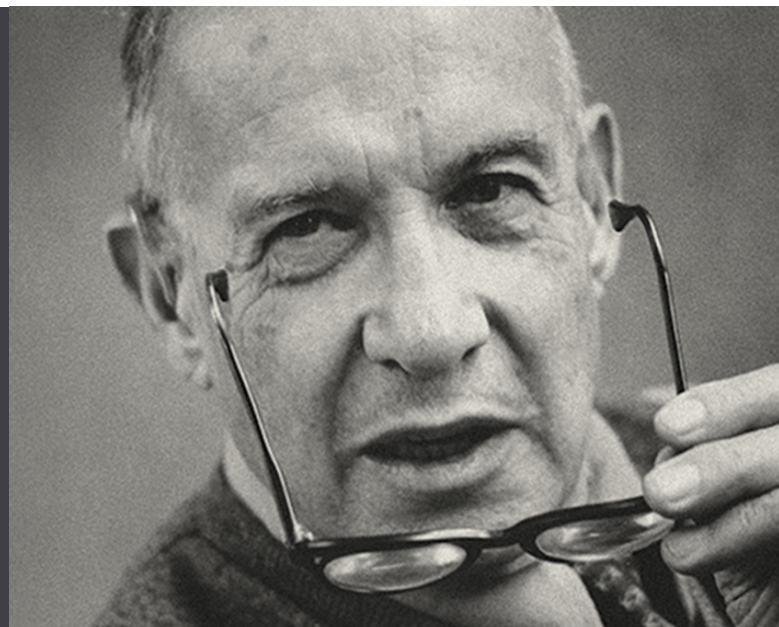
Good procurement work is heavily dependent on access to quality data. Procurement professionals are knowledge workers. Data about suppliers, spend and commercial terms is fundamental to procurement's operational excellence – it is not possible to track and report the quantitative financial value of procurement's impact without baseline cost and spend data for every procurement cycle. Data quality in procurement is therefore often described as "foundational", a hard baseline requirement for the function delivering success.

Unfortunately, many procurement professionals face persistent data quality issues. To understand these issues, let's first define the key terms. What are the key attributes of procurement data quality?

Knowledge worker:

The term was coined by Peter Drucker in 1959, in his book *Landmarks of Tomorrow*, and he expanded on it in his bestseller, *The Effective Executive* in 1966.

Drucker had become convinced that knowledge was a more crucial economic resource than land, labor or financial assets. He also declared that increasing the productivity of knowledge workers was "the most important contribution management needs to make in the 21st century."



Data quality is often judged on six main performance criteria:

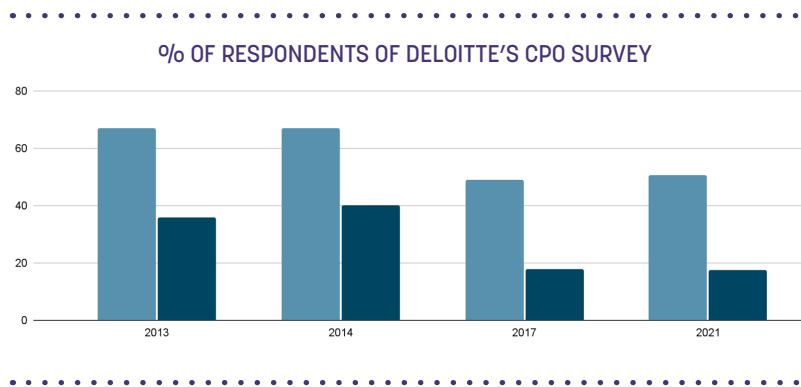
- **Availability:** Does the organization have data to begin with?
- **Validity:** Are the data values valid?
- **Consistency:** Are the values the same across the board regardless of their location? And are they collected in a consistent way?
- **Integrity:** Are the relationships between the data elements and their respective data sets correct and complete?
- **Accuracy:** Is the data a truthful representation of the objects it is supposed to model?
- **Relevance:** Does the data in question support the objective?¹

Procurement data seldom meets all six criteria. Procurement teams reliant on manual data validation workflows are perennially unsatisfied by their ability to avoid costly errors. This is a massive problem for any Procurement Analyst, Procurement Enablement, Procurement Excellence, Procurement Operations function, and it puts hard caps on the overall productivity and value attainment of the procurement team. Spend Matters routinely find that data availability and accuracy are the two most costly hindrances to optimal performance for procurement organizations. As the quantity of data grows, the challenge of relevancy (especially timeliness) has become a much greater challenge, especially determining whether a given record should be incorporated into a business discussion and decisionmaking process.

Procurement's data problem is nothing new, but it has now become more urgent

Procurement's data quality problem is long-standing and persistent. Year after year, when procurement professionals are asked about what hinders their ability to perform, we repeatedly mention data quality and availability as one of our top concerns.

For example, Deloitte's CPO survey shows that a majority of organizations still have data quality issues. CPOs consistently report that solutions to data quality issues remain slow or elusive:



Untrustworthy data has costly business consequences

DFREY5: I feel really bad for the person that fat fingered a \$900mm erroneous payment. Not a great career move ...

JRABINOWIT12: certainly looks like they'll be looking for new people for their Ops group

DFREY5: How was work today honey? It was ok, except I accidentally sent \$900mm out to people who weren't supposed to have it

DFREY5: Downside of work from home. maybe the dog hit the keyboard

- Chats between two employees at HPS Investments, the recipient of what has been called "the biggest blunders in banking history", disclosed as part Citibank's lawsuit in federal court. Reported by [CNN](#), Feb 21, 2022

¹ **Please note that there are several versions of "The Six Dimensions of Data Quality". In certain versions, "Integrity" is swapped with "Completeness", "Relevancy" is swapped with "Timeliness", with essentially the same or similar meanings. Certain lists omit "Availability" and include "Uniqueness" in order to express the importance of cleaning a dataset for duplicate records. Certain lists break out "Consistency" by "Structural" and "Semantic" Consistency. Certain lists include "Lineage" or "Custody" as well as "Currency" and even "Reasonableness" as a catch-all. Perhaps in a fitting example of the problem that these criteria seek to remedy, there is no commonly accepted or canonical "Six Dimensions" used across the data science and machine learning community. For further reading see the following Science Direct query for "Data Quality Dimensions": <https://www.sciencedirect.com/topics/computer-science/data-quality-dimension>





Data quality costs	induced by bad data	<p>Direct costs</p> <p>Indirect costs</p>	<p>Verification costs</p> <p>Compensation costs</p> <p>Costs/missed savings based on wrong/untimely decisions or actions</p> <p>Sunk investment costs</p> <p>“Worm in the apple” (knowing that some data may not be right destroys confidence/trust)</p>
	to manage data quality	<p>Prevention Costs</p> <p>Detection Costs</p> <p>Repair Costs</p>	<p>Training costs</p> <p>Monitoring costs</p> <p>Standard development and deployment costs</p> <p>Analysis costs</p> <p>Reporting costs</p> <p>Repair/re-entry implementation costs</p>

To understand the magnitude of these costs, research conducted by Gartner and Dun & Bradstreet showed that if the cost of a single record is typically \$1, then, on average, the resolution costs about \$10 per record, and the cost of correcting bad data is \$100 per record.

There are 10x factors at play in the costs of untrustworthy data. Some of these costs represent a significant budget that adds to the costs of collecting and embedding data within business processes. The magnitude of these costs represent a distinct challenge for procurement organizations specifically, a challenge that is unique from people, process or technology.

There is also the silent and unmeasurable costs of doubt. As the famous Michael Scott quote goes: “You miss 100% of the shots you don’t take,” and there are few problems that cause more untaken shots by procurement teams than untrustworthy data. When a procurement team does not trust the data, it doesn’t just lead to missed opportunities.

Data-related spending breaks down into four areas.

				
	1. Data sourcing	2. Data architecture	3. Data governance	4. Data consumption
Description	Cost associated with procuring data from customers, ¹ 3rd-party vendors, etc	Cost associated with data infrastructure (procuring software, hardware) and data engineering (building and maintaining infrastructure)	Cost of data-quality monitoring, remediation, and maintaining data-governance artifacts (eg, data dictionary, data lineage)	Cost associated with data analysis and report generation (including spending on data access and cleanup)
Components	3rd-party data	Labor, infrastructure, and software	Labor, software	Labor, software
Typical owner of spend	Head of business unit	CIO	Chief data officer	Head of function or business unit
Typical spend, % of IT spend	5–25 ²	8–15	2.5–7.5	5–10
Example for a midsize financial institution,³ \$ million	70–100	90–120	20–50	60–90

¹Excludes internal data-capture processes.
²Industries that don't directly touch consumers (eg, consumer packaged goods) spend a higher share (>20%) on data sourcing.
³For midsize organizations with revenues of \$5 billion to \$10 billion and operating expenses of \$4 billion to \$6 billion. Absolute values vary by industry and size of the organization; eg, absolute spend is, on average, higher for the telecommunications industry.

McKinsey & Company

It also creates bottlenecks and mis-communications that can damage customer and supplier relationships. The simple fact is that when procurement teams don't have confidence in their data – whether that data originates from suppliers or stakeholders – it becomes much harder to efficiently and accurately make value-creating decisions.

The root cause of untrustworthy procurement data is both obvious and ubiquitous: free-text data entry checked using a laborious, manual error-prone validation process. In fact, most of the procurement leaders interviewed about the topic of data quality for this article stated the data in their supplier quotes received no validation step at all prior to analysis, meaning that costly errors were only identified painfully late in the process.

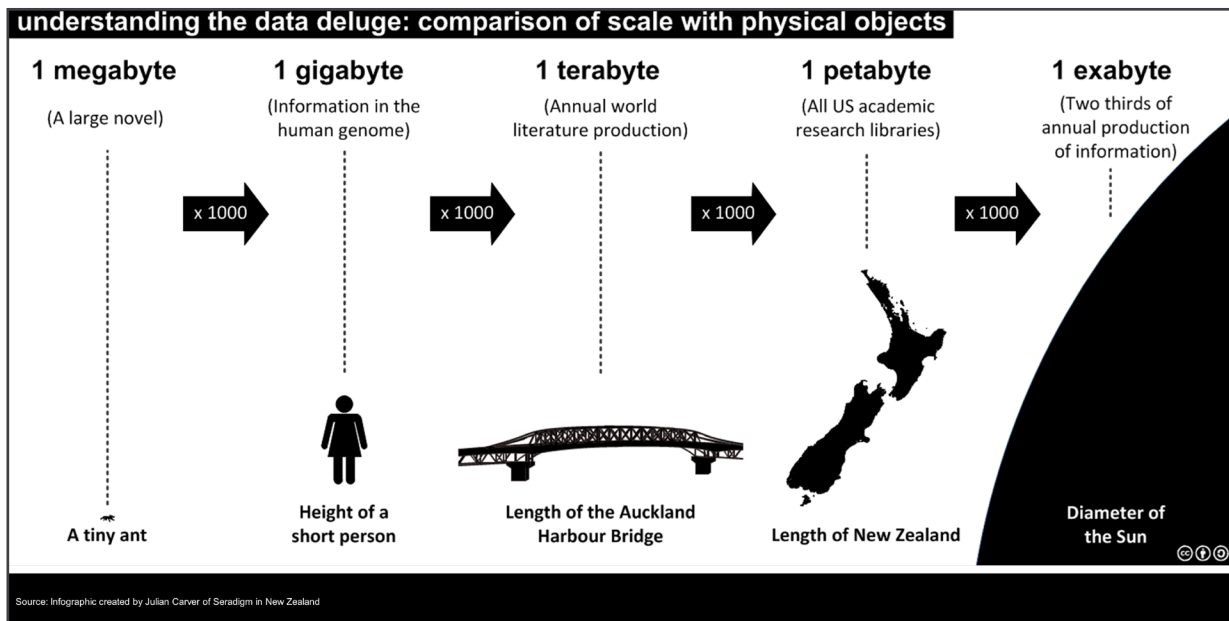
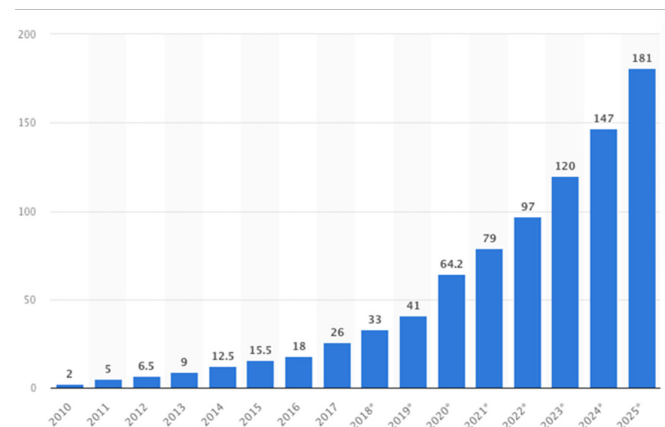
Data proliferation just makes it worse

The progress induced by the digitization of data and the digitalization of procurement processes exacerbates the issue because the amount of data that is produced has dramatically increased. For many teams, this explosion of data has completely overwhelmed the manual validation processes that have traditionally been used for ensuring data quality.

Estimates show that the volume of data that will be produced this year (2022) will be around 100 zettabytes. To put this rather abstract number into perspective, it is useful to consider that one zettabyte is equal to 1021 (1,000,000,000,000,000,000) bytes.

To make it even more concrete: if an ant represents one megabyte, then an exabyte is the size of the sun:

VOLUME OF DATA/INFORMATION (IN ZETTABYTES) CREATED, CAPTURED, COPIED, AND CONSUMED WORLDWIDE FROM 2010 TO 2025. Source: Statista



SOURCE: Infographic created by Julian Carver of Seradigm in New Zealand

And, a zettabyte is 1,000 exabytes. It is not just the mass of information that creates a challenge. Data changes frequently and its “use-by date” expires very quickly. As many procurement processes move from consuming static internal databases to consuming “live” or continuously updated third party data, the problem of ensuring that data is accurate, valid, consistent, integral, and relevant is only increasing by the hour.

A NEW APPROACH IS REQUIRED TO ORCHESTRATE TRUSTWORTHY PROCUREMENT DATA

“INSANITY IS DOING THE SAME THING OVER AND OVER AND EXPECTING DIFFERENT RESULTS.” - ALBERT EINSTEIN

The traditional approach to manual data validation in procurement has never worked.

Traditional methods of managing procurement data have never worked very well because procurement data includes both internal data (from stakeholders and ERP systems) and external data (from suppliers), creating translational challenges across item and service codes, descriptions and other numerical identifiers.

In fact, traditional manual methods for validating and improving procurement data quality only apply to a small subset of the data that most procurement teams need in order to be effective. Here are a few of the drivers for endemic and persistent data quality issues in procurement:

- **Data entry tasks are often fragmented, making the data prone to errors.** With numerous internal and external stakeholders involved in creating and curating procurement data, the result is often a high variance on data quality. This is especially true for commodity and GL code classifications.
- **Data entry and data consumption are often siloed, blocking validation.** Siloes between systems and processes make cross-referencing the same value in multiple tables super challenging. When individuals

who don't work in procurement enter in data (e.g. in a free-text Purchase Requisition) they aren't motivated to follow processes and guidelines regarding data entry.

- **Data quality controls that rely on human validation are not scalable.** Instead, manual validation makes the number of hours in a day a hard cap on procurement data quality. With the amount of procurement data growing exponentially, the number of errors and bottlenecks grow right along with it, rendering the traditional approach to mastering data harmful.

Let's examine the relationship between manual entry and quality.

First, among all the data that procurement needs to operate, a lot originates outside of procurement's control and relies on manual entries. Multiple stakeholders inside and outside of the company manually enter data that procurement must then validate in order to leverage. This phenomenon has been exacerbated by the widespread implementation of “self-service” supplier portals and stakeholder intake ticketing systems and case management portals. While such “self-service” approaches to procurement systems put the tasks of filing supplier profiles, purchase requests, etc. back onto the business units and departments or suppliers themselves with the laudable goal of increasing efficiency and productivity, they create downstream challenges in data quality. These challenges simply stem from the fact that self-service systems dramatically increase the number of people in an organization who touch a procurement process and feed it with manually entered data.

Low intention is when people have a low motivation to follow processes or guidelines and are ready to take shortcuts.

It is not because they do not care; It is more a question of attention and priorities.

Data that is manually entered in by people who do not work in procurement, and who lack sufficient motivation or incentive to enter in high-quality data are unlikely to provide the data foundation that procurement leaders want and need. A typical example is the selection of the relevant

Procurement teams aware of these data quality issue often resort to problematic remedies:

- **Centralize data in one place.** However, centralization of data in a “data lake” does not improve the underlying data quality. At best, it helps bring the scale of issues to light.
- **Set up intermittent data cleansing cycles.** While this can create temporary improvements it is very expensive and data quality degrades between campaigns.
- **Define controls and workflows using approvals, often involving multiple stakeholders.** Data controls that add stakeholders add to the “low motivation, low intent” root cause, while also adding workload and creating painful process bottlenecks.

These traditional approach for data quality issues often fail to address the underlying causes of untrustworthy procurement data, while creating new unforeseen problematic consequences. Perhaps most significantly, given the dramatic increase in data these approaches cannot **scale**.

A false peak is a well known phenomenon in mountaineering and is a peak that appears to be the pinnacle of the mountain but upon reaching, it turns out the summit is higher.

False peaks can have significant and discouraging effects on climbers’ psychological states by inducing feelings of lost hope or even failure: expectation and excitement build, sinking feeling, enthusiasm wanes, energy sinks, depression hits.

The continuous and exponential growth in data that needs to be considered and managed has exacerbated the problems mentioned — more errors, more bottlenecks, etc. — which is incompatible with organizations’ agility and resilience objectives.

Finally, these approaches can have a disastrous impact on the procurement team’s morale. Whenever a team thinks it has finally

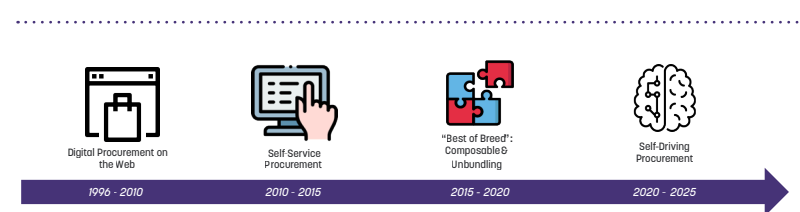
reached the summit of its data quality climb, a new set of data elements and/or sources appears. Data quality improvements resemble an endless climb: each time we think we’ve reached the summit, a higher peak appear

Introducing Predictive Procurement Orchestration

“If I had asked people what they wanted, they would have said faster horses.”

- Henry Ford

Predictive Procurement is the latest evolution of digital procurement. It follows the digitization and digitalization eras and represents the true digital transformation of procurement as it leverages technology to do things that were previously impossible.



The reality is that procurement is moving from a world of backwards-looking static reports to a world of continuous always-on data feeds that update in real time without human intervention. As more enterprise data becomes subject to continuous updates, this creates the opportunity to move from a intermittent “data-cleaning” approach to data quality to a continuous improvement approach to “live” data quality. Here’s where

- **Predictive procurement embeds “always-on” validation loops using external sources to check internal records and generates prioritized lists of exceptions.** Validation loops run continuously to cross-reference data elements between tables in disparate systems. Arkestro predicts and checks the value using external data sources and continuously delivers recommendations to fix.
- **Predictive procurement embeds version control and anomaly detection.** Arkestro embeds version control within the existing solutions and processes, showing trends, spikes and outliers in a particular value over time. By identifying and ranking exceptions, Arkestro enables real-time anomaly detection before an incorrect purchase or quote is approved and impacts the P&L.

Predictive Procurement Orchestration (PPO) leverages technology to do what traditional methods of procurement data management have so far been unable to do.

PPO is characterized by three main features:

1. **A no-login, seamless user experience across multiple systems:** Users access data via an embedded agent designed to orchestrate data that applies to the same entity but lives siloed and unsynced across multiple systems. End-users are not required to log in to an application for the agent to run.
2. **Pre-embedded and always-on data validation:** An embedded platform uses live data to predict inputs and optimal attributes before any process begins. Live data is extracted from email-only or text-only workflows and is updated automatically and checked for accuracy, field-mapping and expiration.
3. **Alerts, notifications, and delegated actions-by-default:** An embedded platform, operating without human intervention, self-triggers recommendations to human users and then learns from the acceptance rate of those recommendations, thereby improving that acceptance rate over time.

The whole premise of Predictive Procurement Orchestration is to change to a new approach to data quality: a self-improving and self-healing data process that applies an “always-on” approach to procurement data monitoring. PPO enables organizations to leverage the power of predictive models for real-time error handling by simulating an action (such as a data entry) before it happens. The same methodology can be used to rank exceptions and anomalies to improve a process in real time as the process runs.

CONCLUSION

Most organizations believe that their procurement system centers on people, processes, and technology. What’s missing in this picture? Data. The data that flows through technology and processes and is then interpreted by people to make decisions is a distinct domain. Data is often an integral part of the organization that’s overlooked. Despite our best efforts to invest in systems and tools designed to capture, clean, and consume data, untrustworthy data still routinely costs organizations tens of millions of dollars of value in unrecognized opportunities.

With a growing amount of data and an increasing need for agility and resilience, procurement teams cannot rely on traditional approaches to manual data validation. Predictive Procurement Orchestration (PPO) offers a unique and novel approach to enabling a foundation of trustworthy procurement data. PPO is always-on and self-driving: it simulates an action like a data entry before it actually happens, predicts a range of potential values that it ranks and filters to exclude anomalies and exceptions to ultimately recommend or execute the optimal action. Thus, we see that PPO has the potential to offer a scalable solution to modern procurement team’s core challenge of achieving trustworthy data quality **at scale**. As procurement leaders remain intensely focused on the challenge of improving data quality, we expect that interest in this solution space will only increase.