

E-BOOK

23 DATA & ANALYTICS '23 PREDICTIONS FOR 23

Data insights from industry leaders and influencers



“The cornerstone of data-driven enterprises is clean, integrated data.”

Dr. Michael Stonebraker

Turing Award Winner 2014
Co-founder, Tamr Inc.

TABLE OF CONTENTS

Foreword from Andy Palmer	3
Foreword from Michael Stonebraker	4
Executive Summary	5
Data Culture	6
Data Integration	18
Data Governance	24
Data Architecture	28

Foreword from Andy Palmer

Digital transformations are more important than ever before. But as countless enterprise leaders have learned, when they do not build a data-driven culture centered on managing data as an asset, their efforts are hardly transformative. We've seen companies make significant strides over the past decade, but many still have a long road ahead. The proverbial "long pole in the tent" for becoming data-driven - having clean, curated, and continuously-updated data for many users to consume - is a problem that is as old as time in the enterprise. We've been through decades of attempts to deliver on this dream: data warehouses in the 1990s, data lakes in the 2000s, and most recently, cloud data platforms.

After years of prediction, the shift to cloud is here. Every organization I know is migrating their infrastructure to the cloud, some rapidly, some gradual. And it's clear that as enterprises adopt cloud data platforms, they'll continue to live in a hybrid state of cloud and on-prem for at least the next two decades. The growing pains from these migrations can be substantial, but so is the change. The shift from on-prem to cloud is no less significant than the shift from minicomputers to personal computers in the 1980s and 1990s. The cloud provides opportunity. But the reality is that

a number of legacy, on-prem requirements will persist for the coming decades. Hybrid, while not the preference, will be the reality for the foreseeable future. Organizations - and vendors - who embrace this reality, are the ones that will succeed.

While cloud data platforms seem like just the latest in many generations of shiny new data technologies, the true bottleneck that will prevent success is, once again, the problem that we've had since we started automating business processes back in the 1970s: the resolution of the idiosyncrasies of the data from many operational systems into tables of clean, curated, continuously-updated data that is organized as required for consumption. Solving this age-old problem requires modern technology that uses the power of probabilistic modeling and leverages human expertise very efficiently. Over the coming year, many of the cloud data platforms will embrace next-generation data mastering as a keystone of the modern enterprise cloud data platform.

We're also seeing the boundaries between structured and unstructured data disappear. To borrow the words of Tamr Chief Product Officer, Anthony Deighton, "there is no such

thing as unstructured data. Only data that is yet to be structured." Anthony is exactly right and increasingly, we're seeing tooling and infrastructure orienting around this fact.

Finally, the distinction between internal and external data is much less pronounced than it was in the past. In reality, much of the data in a company's enterprise systems is incorrect. That's why data enrichment is becoming increasingly more important - and necessary - to deliver clean, curated data. Companies are finally realizing that they must use external data to enrich and complete their internal data.

As we head into 2023, I'll leave you with two final thoughts: anything that feels like a panacea is probably wrong and your biggest barrier to becoming data-driven isn't technology. It's people. We've reached the point where technology is no longer the issue. It's a person's ability to consume, organize, and understand the data available to them that is the most significant barrier we face.



Andy Palmer

Co-founder, Chairman & CEO, Tamr Inc.

Former Head of Software & Data Engineering, Novartis Institute for Biomedical Research

Former CIO, Infinity Pharmaceuticals

Foreword from Michael Stonebraker

The cornerstone of data-driven enterprises is clean, integrated data. It's just that simple. And as we head into 2023, here is my advice to you.

1. Move data, starting with decision support data, to the cloud as aggressively as possible

Starting with decision support is easier than starting with online transaction processing, which is why I suggest you start there. But instead of taking the traditional “lift and shift” approach, I strongly suggest that you incorporate data integration as part of your data movement to the cloud. Restructure, reformat, and integrate your data along the way so it is prepared for consumers of data within the enterprise.

2. Create an enterprise-wide strategy for managing the idiosyncrasy created by data silos

All enterprises today have many data silos and they are not going away. But creating the infrastructure required to eliminate the idiosyncrasy created by the silos generates huge potential business value. That's why businesses need to develop an enterprise-wide strategy for mastering data across data silos. Currently, this occurs on an ad hoc basis

for individual projects, with the emergence of “self service data prep” 10 years ago serving as a reflection of this problem. But when you implement a systematic way of mastering your data across your entire organization, you'll realize greater benefits.

3. Put a Chief Data Integration Officer in charge of creating enterprise best practices

I believe this is the most important thing an enterprise can do. By documenting a consistent set of best practices for how an organization integrates and masters its data, organizations can apply learnings from the past to future projects.

4. Make cleaning data everyone's job

If clean, integrated data is the goal, then cleaning data must be everyone's responsibility. Whoever creates the data should feel obligated to make the data as clean as possible at the point of data entry and should curate data at the point of data creation. This is the aspirational goal, and will help organizations move towards a state where their data is cleaner and more integrated than before. Additionally, as much as Google Knowledge Graph, Apple

Knowledge Graph, and other consumer data organizations curate the world's consumer data, the internal data organization inside of each company will become responsible for engaging creators and consumers of data in the curation process.

By applying this advice, you'll accelerate your ability to become data-driven and provide the clean, integrated data your organization needs to succeed.



Michael Stonebraker

Adjunct Professor at MIT
Turing Award Winner 2014
Founder, Postgres project at University of California, Berkeley
Founder, Ingres
Co-founder, Tamr Inc.

Executive Summary

The Greek philosopher Heraclitus once said “everything flows and nothing stands still.” Nothing could be more true when it comes to data.

The volume of data is increasing, as are the number sources that generate it. Data organizations are changing to meet the growing demands of their business partners. Investments in data and data-related initiatives are rising. New approaches to managing data at scale are emerging. And data technology is evolving to better support the new regulations, new strategies, and new priorities of modern organizations focused on digital transformation.

Perceptions about data are changing too. Executives are putting their money where their mouth is in order to ensure their business becomes truly data- and analytics-driven. Business users are becoming more comfortable with data and are using it to drive better decisions.

Data leaders are realizing that they need to expand their role to focus not just on implementing data technology, but on solving business problems. And that the value of data is most commonly realized in four key ways:

- 1. Cost savings** that optimize direct and/or indirect spend
- 2. Increased growth**, either by selling something new to existing customers, or selling existing products to new customers

3. Reduced risk across corporate, financial, customer, or product

4. Increased operational efficiency through the creation of more quantitative context for key operational processes and decisions

We spoke with data leaders and industry influencers who shared their perspectives on what we can expect in 2023. And while their insights were diverse - and sometimes contrarian - four consistent topics emerged:

- **Data culture** is changing quickly within organizations
- **Data integration** is both a key challenge and the keystone for realizing value from data
- **Data governance** is shifting its focus from source-based to consumption-based governance
- **Data architecture** is rapidly embracing the idiosyncratic reality and large number of raw data sources, both internal and external

Within these topics are 23 predictions for the year ahead. Let’s take a closer look.

Data Culture

“Is data in the room?” It’s a simple question, but for many organizations, it’s a difficult one to answer. Because the truth is, despite years of effort and investment, many organizations still do not have a data-driven culture.

True data-driven organizations do things differently. Their culture mandates that data is everyone's job - not just the responsibility of the Chief Data Officer (CDO) or data team. They manage their data as a product and core asset of the company. They ensure that data meets the needs of the business. And that business decision-making starts with quantitative methods that are supported by clean, high-quality data.

Asking if “data is in the room” is an easy way to remind our business users to include data in their decision-making. It’s not, to me, just a technology problem anymore, but a business problem with technology components.”

André Vargas, Chief Data Officer, Creative Artists Agency (CAA)

Data-driven companies also invest in their data organizations, with more than **70% of organizations** reporting that they have a CDO in place today. But as data evolves, so too are the roles and organizational structures that support it. In some organizations, roles are converging. In others, new roles are emerging to address specific needs of the business. Still others struggle to find the resources they need to support their data initiatives. And of course, the tradeoff between centralizing resources into one center of excellence versus

distributing data professionals into the front lines of a business remains a point of debate.

Becoming data-driven requires a cultural change as well as a mindset shift, both in focus and perception. When people understand the importance of using data and insights in their daily activities, that’s when they’ll realize the true value.

“For organizations to become data-driven, data needs to be embedded in everybody’s job - and everybody needs to think of data as a fundamental asset of the organization in the way that they think of cash and cash flow across the organization.”

Randy Bean, Founder and CEO, NewVantage Partners, a Wavestone Company

As we head into 2023, we believe that cultural changes are afoot. Organizations will finally begin to commit to becoming data-driven. New roles will emerge while existing roles will evolve and expand. And data products will rise as a primary focus to enable organizations to realize data as a core asset.

Prediction #1

Executives will move from lip service to a true commitment to data.

While some organizations were built with data at the heart of their business, others need to evolve in order to make data a core part of their culture. At the highest level, executives need to commit to becoming data-driven. But too often, executive expectations are misaligned with the reality of the business because they don't understand that what the data organization is delivering is difficult and complex.

“The biggest challenge for most organizations is the gap between expectations and the lip service that’s paid to using data within the business and the on-the-ground, in-the-trenches reality of what it takes to achieve a level of data competency.”

Randy Bean, Founder and CEO, NewVantage Partners, a Wavestone Company

To overcome this gap, executives must do more than just talk about becoming data-driven. They must:

- Truly commit to becoming data-driven, while also realizing it's a long-term effort that will not deliver results overnight
- Ensure that they invest in data initiatives so that their business can keep pace with their data-driven peers and achieve their data goals
- Empower their CDOs not just with funding, but with authority and people



Prediction #2

The focus of CDOs will shift from solving data technology problems to solving business problems.

In many organizations, CDOs emerged as a way to fill a gap in the organization. As CIOs struggled to deliver on the remit of using data as a strategic weapon against the competition, the need for CDOs evolved, enabling them to take responsibility for - and realize the value of - data. But many times, these newly-minted CDOs focused solely on the data ecosystem and the technology that supported it.

Today, times are changing, and many CDOs realize that they must broaden their scope to embrace the context in which their enterprise consumes data and to treat the front line business owners as true partners. They are evolving their role from data technologist to strategic business partner, enabling them to forge stronger bonds with leaders across the organization to solve business problems.

“The business of the CDO is business, not technology.”

Dr. Richard Wang, Director, Chief Data Officer & Info Quality Program at MIT and Professor at University of Arkansas, Little Rock



Prediction #3

Many firms will increase investment as they continue to play “catch-up” relative to spend on data and data-related initiatives.

The pandemic accelerated digital transformation for many organizations. And when digital transformations accelerate, investments in data initiatives often follow suit.

“Digital acceleration has created the need for hyper-decisive organizations. And hyper-decisive organizations build sophisticated data management and advanced analytics capabilities in order to raise their competency to make decisions.”

Jamie Popkin, Digital Business Strategies, Dresner Advisory Services, LLC

According to a Tamr survey of data leaders, over 90% of organizations have data initiatives planned for 2023. To realize the benefits of these initiatives, organizations must ensure their investments deliver real business value all along the way and avoid technology-oriented projects that attempt to “boil the ocean”.

“We’re seeing a shift. Executives are willing to dedicate a greater share of wallet to building a best-of-breed tech stack when it helps them to deliver the clean, accurate data needed to drive better decision-making, business value, and ROI.”

Melissa Campbell, Chief Revenue Officer, Tamr Inc.



Prediction #4

The roles of CDO and CAO will converge and their scope will expand.

For many years, the CDO and Chief Analytics Officer (CAO) have been evolving as relatively separate roles, each emerging with its own responsibilities and remit. But as data continues to evolve and change, these roles must converge, empowering CDOs to embrace a more holistic view of data consumption across their organization.

Similarly, CAOs are realizing that without clean, curated, continuously-updated data, they cannot deliver on their remit of democratized analytics. In fact, most CAOs now believe that clean, curated, high-quality data is the bottleneck to their analytic initiatives.

As a result of this convergence, today, there are more Chief Data and Analytic Officers (CDAO) than in the past. And many more CDAOs will emerge in the years ahead.

“Some people say ‘this is data analytics’ and others say ‘no, this is data management.’ These roles need to converge - and be empowered with budget, authority, and resources. And my hope is that, ultimately, every organization has a chief data officer and a data department, just like you have a marketing department today.”

Dr. Richard Wang, Director, Chief Data Officer & Info Quality Program at MIT and Professor at University of Arkansas, Little Rock



Prediction #5

As data citizens gain business context through more accessible tooling, the role of the citizen data scientist will continue to rise.

Historically, data scientists have struggled to succeed in many organizations. And many times, it was because they lacked business context for the problems they needed to solve.

But as organizations implement self-service tooling, data citizens gain greater access to data. And because they understand the business context, it gives way to more sophisticated analysis, which is causing the role of the citizen data scientist to rise, especially in larger, more data-driven organizations.

“I’m seeing that the rise of the citizen data scientist is much more at the top end of the curve. Tier 1 organizations almost ensure now that you have some level of Python and/or SQL, regardless of your role in the organization. But at the tier 2 level, there isn’t as much of a data analysis mindset required in order to join the organization. That said, I see things moving down the path where you’ll need to understand data models, data architecture, or the no code/low code solutions that let you respond to business needs faster leveraging data toolkits.”

Chris Napoli, Head of Wealth & Asset Management, Snowflake



Prediction #6

Data engineers will become an even hotter commodity.

As the number of data scientists inside organizations exploded over the past decade, many organizations realized that great data science without next-gen data engineering is putting the proverbial data science cart before the data engineering horse.

This is a natural evolution. A decade ago, Tom Davenport **called data science a “sexy” job**. Data engineering is much less sexy, but it is now becoming the bottleneck for organizations who are trying to realize all the data science and AI-related promises.

Data engineers are already in high demand. But as organizations realize they need more data engineers, these skilled individuals will become highly-desirable and expensive, making them even more difficult to find, recruit, and hire. That’s why organizations looking to increase investment in data engineering will realize that finding qualified data engineers will continue to be a barrier on their road to data-driven success.

“Typically, what I see as the bottleneck in data processing is data engineering. It’s a critical role, though, because without it, there is no data platform. The data won’t flow. And it won’t be scalable or timely - all things you really need. And as the demand for data and data products increases, you really need data engineering.”

André Vargas, Chief Data Officer Creative Artists Agency (CAA)

To overcome this challenge, some organizations are encouraging software engineers to morph into data engineers since many of the principles and skill sets are similar. Other organizations are trying to retrain their database administrators. Sometimes this works, but in our experience, data engineers are most successful when they are retooled software engineers or computer scientists with a database systems focus.

As data engineers continue to be a sought-after resource, we’ll see more organizations explore creative ways to fill these gaps. But inevitably, many will need to pay up to get the best talent.



Prediction #7

Business people will proactively embrace the need to build their analytic muscle and the importance of data quality.

Business analysts are frustrated by the lack of data organization and quality, as well as the fundamental inability to find the data they need. To address this frustration, these analysts, often called analytics engineers, will assume responsibility for the collection of data across the organization to ensure that it is consolidated and organized as appropriate for their needs. This is a natural evolution of business people becoming data- and analytically-driven.

“This is the biggest shift we’re seeing overall. Many more of the people attached to the business and doing these one-off data analyst workflows are now trying to figure out how they can centralize and curate the larger body of knowledge.”

Tristan Handy, CEO & Founder, DBT Labs



Prediction #8

The CDO will view data products as the primary artifact they deliver to their organization.

As companies shift their thinking toward data products in 2023, a new role will emerge for the CDO: the data product owner. Similar to a software product owner, the data product owner's role is to manage the data products organized around key, logical entity types in the organization such as customers, suppliers, products, and employees. Most organizations will have hundreds of data products that they deliver to their organization, and they will require a great deal of new infrastructure, people, and processes to deliver these data products reliably.

“A data product owner is a newer concept, but it's a recognition that data products are going to be the preferred way of consuming data for the vast majority of potential consumers.”

Justin Borgman, Chairman & CEO, Starburst



Prediction #9

Data marketplaces exploded onto the scene but will continue to stall because of the gap between the data in the marketplace and internal data in modern data platforms.

The major cloud data platforms have invested significantly in creating data marketplaces. These marketplaces contain data from many third-party data providers, and the process of mapping and matching the third-party data into a company's tables on their cloud data platform is a very manual and time-intensive task. In order for data marketplaces to realize their potential, we need to close the last mile between the files that third-party providers deliver to these marketplaces and the context of the actual tables that companies have in their cloud data platforms.

Data products will help to close this gap by providing curated data sets that organizations can then map and match with third-party data sets that have the potential to offer significant enrichment of the company's key data.

"It's always been such a pain point to find and leverage external data and integrate it with your internal data. We definitely think that [data marketplaces] are a huge opportunity going forward."

Wayne Eckerson, President, Eckerson Group



Prediction #10

Companies will realize that the most accurate version of their data lives outside their organization.

As the volume of data within an organization grows, many are realizing that the quality of their internal data has decayed. And in order to improve the quality, they need to enrich their internal data with data from external, third-party sources. Think about it. Where is the most accurate version of a customer's address and phone number likely to be - in your internal systems or on their company website?

“One thing that is going to change is that there will be increasingly less distinction between internal and external data. People have realized that much of their internal data was fat-fingered by individuals and that to make it complete, they need to enrich it with data from the outside world.”

Andy Palmer, CEO, Tamr Inc.

In 2023, more organizations will embrace data enrichment to significantly improve the quality, accuracy, and relevancy of their data.



Prediction #11

Organizations will realize that data is one of their most valuable and underutilized assets.

We believe that a company's data is a valuable asset. And in order to compete in the digital economy, companies must treat their data as a product.

“At its core, every business is a data business. Which is another way of saying every business should have data products and think about managing their product - which they might think is software or retail or healthcare. But it's not. It is, in fact, the data. And they should manage that asset like a product.”

Anthony Deighton, Chief Product Officer, Tamr Inc.

In 2023, forward-thinking companies will realize that data is an underutilized secret sauce, and they will begin to develop strategies to manage their data as a key, strategic product. They'll invest in data product strategies and uncover ways to monetize their data, enabling them to leapfrog their competitors by integrating data into their fundamental value proposition and optimizing their capital efficiency by spending less and reducing risk in their business.



Data Integration

All organizations have data silos. And when an organization attempts to integrate data from multiple source systems, continuously cleaning and organizing the data for consumption is a significant challenge.

Operational systems inside a company generate data. And that data contains the idiosyncrasies of the organization and the people that built and own the system. These idiosyncrasies are a fixed reality, and when multiplied by the thousands of systems in a large organization, the differences in data across systems become the primary bottleneck in producing clean data that is useful for potential data consumers.

“If the focus is on getting all the technology around the data, you’re missing the point - because the data is still dirty.”

Arturo Morales, PhD, Vice President, Data & Technology Solutions, WCG

Improving holistic data quality and integrity is the **#1** thing organizations can do to become more data-driven.

Nearly **95%** of data leaders anticipate investing more in AI/ML in 2023.

This is why data mastering is critical. Modern data mastering is both machine-driven and human-guided, enabling organizations to consolidate, clean, organize, and continuously-update internal and external data to power the insights they need.

AI and machine learning provide many benefits, but as more and more organizations employ AI/ML, they are also realizing that they need to ensure the data the machine is using is clean and accurate. That’s why, as we head into 2023, more organizations will realize that they must keep humans involved too.

“I’m not going to predict that AI will be everywhere, but I think that we will have more helpful AI. And that’s something I would love to see. Not AI replacing jobs, but AI helping people to do their jobs and making their jobs less boring.”

Christophe Blefari, Data Engineer

The Power of Machine-Driven, Human-Guided Data Mastering

If you’re wondering how to strike the right balance, here is a formula that illustrates how the best modern data mastering solutions work:

**80% machine
+ 10% humans
+ 10% rules**

**100% Modern
Data Mastering**

This is Tamr. Our cloud-native, machine learning-driven approach to data mastering provides the clean, curated, continuously-updated data your organization needs to power analytic insight and accelerate business outcomes. Organizations benefit from the power that machine learning provides and the valuable feedback only humans can contribute. Tamr’s **patented** technology solves many of the challenges traditional MDM solutions cannot overcome, allowing businesses to accelerate critical analytical insights by reconciling internal and external data at scale. The result is cleaner, more accurate data and higher levels of trust.

Prediction #12

It's not about big data - it's about clean, curated, continuously-updated, accurate data.

Big data has been a big topic for many years. Yet today, many believe big data is overhyped.

“Big data is a term that's overhyped. Little data is just as good as big data if you're analyzing it for a purpose to come up with a result or to make money.”

Greg Wessel, Senior Technical Expert in Cyber, Networking, and Analytics

That's why, in 2023, we think the conversation will shift away from big data and instead, focus on the need for clean, curated, continuously-updated, accurate data. After all, if your data, big or small, is dirty, it's impossible to realize value from it.



Prediction #13

Dirty data isn't going away, but the tooling organizations have to fix dirty data will improve significantly.

Business leaders make bad decisions when the underlying data from which they draw insights is dirty, disorganized, inconsistent, or outdated. That's a fact. And unfortunately, we can't predict that this issue will go away in 2023. However...

We do believe that organizations will get significantly better at fixing their dirty data. They'll have access to better tooling, like data mastering, which will make cleaning their data easier and more automated.

"Data mastering is like spell check but for data. It employs machine learning to fix errors in the data, freeing up your team to focus on making better decisions."

Anthony Deighton, Chief Product Officer, Tamr Inc.

As organizations invest in tooling like data mastering and enrichment, they'll reduce the amount of time they spend fixing dirty data, enabling them to focus more using data to drive better decisions across the organization.



Prediction #14

AI/ML will continue to grow, pushing us towards a tipping point between automation and manual processes.

Historically, many of the processes companies used to clean and manage their data throughout its lifecycle were manual and time-consuming. But as more businesses invest in data, demand for more efficiency and greater scalability will grow, as will the need for clean, accurate data. That's where AI and machine learning come in.

As we head into 2023, organizations will reach a tipping point. It's no longer possible to rely on traditional, rules-based master data management (MDM) processes that are primarily human-driven. They simply don't scale. Instead, businesses need to embrace ML as a way to automate the mastering of data across the organization so they can break down data silos and deliver clean, accurate data for analytics and decision-making.

"This is the biggest prediction for 2023: DataOps and AIOps have traditionally been a manual set of processes - the whole lifecycle. I think we've reached a tipping point where the business demand globally is such that we have to start making these intelligently automated, otherwise we'll never scale to meet demand."

Mark Clare, Former Chief Data Officer, JPMorgan Chase and HSBC

Embracing AI and machine learning will enable organizations to resolve data and data entities at scale and with speed. Using the machine to organize and clean data enables the inclusion of many more data sets as input because the machine can integrate new datasets at a marginal cost that is at least the same, if not less.

While we believe that machine learning-driven data mastering will outpace traditional MDM, we also believe that AI and machine learning alone are not enough. Which brings us to our next prediction.



Prediction #15

Hybrid AI will rise.

As AI and machine learning continue to rise in popularity, many organizations are now realizing that machines alone are not enough. They also need to keep humans in the loop to provide critical feedback and context to ensure accuracy of the model results.

“[In our industry], you want to have mathematical significance but also clinical relevance. Those two are really important. You can’t do clinical relevance because you have too much volume - you can’t do it by hand. And you can have outliers that are great - and outliers that are bad. Context is important.”

Arturo Morales, PhD, Vice President, Data & Technology Solutions, WCG

Humans have the domain expertise needed to provide feedback on the model and ensure that the results are correct. The more feedback they provide, the more accurate the model becomes, which drives greater trust across the business. And when organizations trust their data, they are more likely to use it for analytics and decision-making.



Prediction #16

Many organizations will get caught up in the hype of “no code AI” tools, but will fail to solve an actual business problem.

Many data companies today tout their no-code AI tools. But buyer beware - not all no-code AI tools are created equal!

“There is not one shoe that fits all. I think it’s overhyped if an AI vendor says ‘just use my AI and your problems are solved’ without actually understanding my business problem.”

André Vargas, Chief Data Officer, Creative Artists Agency (CAA)

We believe the right no-code AI tools have their place in the data ecosystem. But organizations looking for a quick fix should **proceed with caution**.



Data Governance

Data governance has been in the spotlight the past five years. In fact, many predict that the interest in data governance will continue to grow, but that it is shifting away from source-based governance which mostly focuses on data cataloging and toward consumption-based governance.

“Data governance - the management of data itself - has been done really badly, if it's been done at all, by and large. And organizations are waking up to the fact that they have large amounts of data in one place that they've ingested, and now they have to manage it in order to meet regulatory requirements.”

Jonathan Holman-Rohwer, Chief Examiner & Visiting Scholar, The London Institute of Banking & Finance; Head of Digital/Transformation, Multiple Banks

Data privacy has taken center stage in recent years, too. Between new regulations such as GDPR and CCPA, data privacy and security has evolved, swinging the pendulum from the wild west of years past to an environment where data is significantly more locked down. And while the jury is out on where the pendulum will land, many believe our current, conservative environment is here to stay, but that a focus on data consumption rather than data sources is critical to appropriately protect the privacy of data.

“When it comes to data privacy and security, the pendulum has swung way far to the right. It may come back a little, but it will never come back to the middle.”

Greg Wessel, Senior Technical Expert in Cyber, Networking, and Analytics

Prediction #17

Companies will (finally!) get governance right as they move away from source-based governance and towards consumption-based governance.

For years, organizations have focused their efforts on source-based governance. But they experienced diminishing returns. And while many have successfully implemented comprehensive data catalogs, other expectations from data governance initiatives have fallen short.

That's why organizations will begin to shift their focus: from source-based governance to use- or consumption-based governance. They will continue to leverage data catalogs because they contain important data on all the sources that exist within an organization. But organizations will also provide the tools consumers need to ensure they are using data appropriately and safely based on the information access policies in their organization and the consents of the data owners.

“The future of data governance will focus on understanding how people are consuming data and enabling them to consume data in a way that doesn't get themselves in trouble.”

Andy Palmer, CEO, Tamr Inc.



Prediction #18

As we enter a new era for data privacy and security, organizations will ramp up their focus and investments to ensure they remain compliant.

Across the industry, there is strong consensus that data privacy and security is more important than ever before. No company can afford the devastating reputational harm that results from a data breach or the steep fines associated with non-compliance to privacy regulations. But even though organizations acknowledge the importance of data privacy and security, today, many organizations are still not appropriately investing in it.

“One area that people realize is important, but I don’t think they realize the magnitude of the problem, is around security, information security, and PII.”

Rich Miner, Android Co-founder, and Board Member, Tamr Inc.

As we head into 2023, we’ll see organizations begin to realize that they need greater focus - and more budget - to ensure they remain in compliance and can realize value from the promise that their data holds.

“In the next 5-10 years, data privacy and security will be in the forefront. We continue to struggle as an industry to make sure that our data is well-protected and used in a way that is appropriate and respects the people that are contributing information to it.”

Paul Balas, Data Strategy Consultant, 303Computing



Prediction #19

Machines will play a more significant role in identifying patterns and predicting data breaches.

As data volumes grow, organizations will find it increasingly difficult to detect suspicious patterns or potential threats. But when they employ machine learning, they'll be able to proactively identify and address privacy and security issues before they cause widespread harm.

“From a data privacy perspective, using ML to identify patterns and breaches is going to be a very large industry. Small, mid sized, and large companies are going to have to invest more in data privacy and security.”

Paul Balas, Data Strategy Consultant, 303Computing

Companies are already investing more in AI/ML than ever before. And as these investments grow in 2023 and beyond, we'll begin to see more organizations allocate some of this investment to improving the security and protection of their data.



Data Architecture

Companies have more data - and more data sources - than ever before. But as data continues to grow, so, too, does data chaos. Data is messy, data silos proliferate, and data is constantly decaying, making the task of integrating data across an organization extremely difficult. Impossible, in fact, without the help of the machine.

To tackle these challenges, organizations must consider the best way to structure their data architecture. But as new approaches and strategies emerge, they must also assess which trending topics are worth the hype - and which are overhyped.

Take, for example, data lakes. For years, data lakes vowed to deliver on the unfulfilled promise of data warehouses. Many organizations dumped their ever-growing body of data into Hadoop Distributed File Systems (HDFSs)/data lakes with the hope that they would finally have their data in a place that met their data consumer's expectations. But data lakes are overhyped. That's why many organizations who implemented data lakes are realizing that what they actually created is a data swamp.

“A data lake without any governance is a data swamp.”

Chris Napoli, Head of Wealth & Asset Management, Snowflake

Data mesh is another trending topic, particularly for large firms. Data mesh is an approach that enables organizations to deliver a decentralized architecture that groups and curates data by a specific business domain with the goal of providing a more consistent view of enterprise data resources. It helps to solve the challenge that organizations face when it comes to standing up a single point of access that can query data wherever it lives.

“Data is everywhere, and [data mesh] is a solution to that challenge. It provides self-service access that enables faster time to insight and better decision-making, across all of the data that we have. It's a solution to decentralization. And not only a solution to it, but it also turns a weakness into a strength.”

Justin Borgman, Chairman & CEO, Starburst

And while data mesh is receiving a lot of buzz in the industry, debate continues about whether it's hype - or overhyped.

Heading into 2023, new data architecture strategies will gain momentum, with many claiming that they provide the silver bullet that will solve all of an organization's data challenges. But savvy organizations will realize that anything that sounds like a panacea is probably wrong. Instead, these organizations will focus on the most important goal: ensuring their data is clean and curated.

Prediction #20

Companies will realize that data lakes are dead.

Data lakes held great promise for many organizations, but turned into data swamps for the most part.

"Data lakes are dead and dying, and organizations are aggressively moving away from them. HDFS was a fine file system, but it was built for a specific purpose and irrationally applied more generally, to the disappointment of many."

Andy Palmer, CEO, Tamr Inc.

Having clean, curated data is valuable. That's a fact. But dirty data swamps are not. That's why, in order to have accurate, integrated data, organizations must prioritize three things:

1. Appointing a Chief Integration Officer charged with creating best practices for integrating and mastering data
2. Developing a strategy for eliminating data silos
3. Making the cleaning of data everyone's responsibility

"The most important thing enterprises can do is to appoint a Chief Integration Officer who is responsible for creating an enterprise-wide strategy for stamping out data silos and developing best practices for data integration and mastering. Because if you integrate your data and it isn't clean, you've done nothing."

Michael Stonebraker, Adjunct Professor at MIT



Prediction #21

Data storage costs will drop while performance and availability will accelerate.

For many years, data storage was a hot topic in data architecture. But as more companies move to the cloud, data storage has become a commodity. And that's not a bad thing. While data storage remains important, it is no longer the focus of the conversation. Companies today assume that data storage is inexpensive, highly-available, and performant. And that availability and performance will continue to improve at a rapid pace while costs continue to decrease.



Prediction #22

As the debate surrounding centralizing vs. decentralizing rages on, many organizations will find themselves landing somewhere in the middle.

Companies have debated centralized vs. decentralized data for decades. But the reality is that both centralized and decentralized data architecture design patterns are necessary for an organization to manage their data as an asset or product, and neither option is sufficient on its own. For that matter, betting solely on the rationalization or standardization of systems and data is misguided. Rationalization and standardization are necessary, but not sufficient, to realize the value of the holistic enterprise data opportunity.

That's why we believe organizations will land somewhere in the middle, where they can experience the best of centralization and decentralization methods such as mesh and fabric, enabling them to rationalize and standardize as much as possible.

“Companies flip-flop between centralized and decentralized models. Each of them have their downsides, and when they hit the downsides, then they flip-flop to the other side. But what they really want - or need - is a hybrid model or federated model. The goal is to get the best of both worlds - the standardization, the efficiency, and the economies of scale of centralization with the agility, speed, and flexibility of decentralization.”

Wayne Eckerson, President, Eckerson Group



Prediction #23

When it comes to data mesh, many organizations will ignore the hype, focusing instead on the functionality and value it provides.

We believe that data mesh is certainly experiencing a lot of hype. But as we enter 2023, we also believe that organizations will ignore the hype and focus on the value a data mesh approach can provide, especially when paired with **data mastering**.

For data mesh to work, organizations need to cleanse and standardize their enterprise data. That's why data mastering is so critical. It can serve as both a complement and an augmentation to their distributed data initiatives. Data mastering provides standardized keys for data that can be understood across systems and domains and creates useful mappings between data identifiers across the organization, both of which are often critical bottlenecks in a data mesh strategy.



The Year Ahead: Are You Ready?

Clearly change is afoot as we enter 2023. From culture and integration to governance and architecture, organizations must embrace change in each of these areas if they want to reach the ultimate goal of having clean, accurate, integrated data.

Are you ready to make 2023 the year you finally become data-driven?

Get started at www.tamr.com.

A special thank you to all our contributors!



Paul Balas
Data Strategy Consultant, 303Computing



Randy Bean
Founder and CEO, NewVantage Partners



Christophe Blefari
Data Engineer



Justin Borgman
Chairman & CEO, Starburst



Melissa Campbell
Chief Revenue Officer, Tamr Inc.



Mark Clare
Former Chief Data Officer, JPMorgan Chase and HSBC



Anthony Deighton
Chief Product Officer, Tamr Inc.



Wayne Eckerson
President, Eckerson Group



Tristan Handy
CEO & Founder, DBT Labs



Jonathan Holman-Rohwer
Chief Examiner & Visiting Scholar, London Institute of Banking & Finance. Head of Digital/Transformation, Multiple Banks



Rich Miner
Android Co-founder, and Board Member, Tamr Inc



Arturo Morales
PhD, Vice President, Data & Technology Solutions, WCG



Chris Napoli
Head of Wealth & Asset Management, Snowflake



Jamie Popkin
Digital Business Strategies, Dresner Advisory Services, LLC



André Vargas
Chief Data Officer, Creative Artists Agency



Richard Wang
Director, Chief Data Officer & Info Quality Program at MIT
Professor at University of Arkansas, Little Rock



Greg Wessel
Senior Technical Expert in Cyber, Networking,
and Analytics



Tamr is a highly scalable next-generation data mastering platform that allows customers to use data product templates to consolidate messy source data into clean, curated, analytics-ready datasets. Organizations benefit from Tamr Mastering, the industry's first suite of data product templates that combine human curation, patented machine learning, mastering rules and enrichment with first- and third-party data to accelerate business outcomes and deliver business-changing insights. Tamr's cloud-native and SaaS solutions enable industry leaders such as Toyota, Western Union and GSK get ahead and stay ahead in a rapidly changing competitor environment. Tamr's work in the public sector includes the U.S. Air Force, U.S. Army, Department of Homeland Security and various other federal and local government agencies.

Visit www.tamr.com and follow [@Tamr_Inc](https://twitter.com/Tamr_Inc) on Twitter and LinkedIn for more information about Tamr, its partners, and investors.