

Forbesinsights

---

**Intelligent  
Data Catalogs:**  
At The Forefront  
Of Digital  
Transformation

IN ASSOCIATION WITH:



Informatica



**T**rust in information is the glue that moves today's data-driven business forward. By integrating and leveraging information from any and all points across the enterprise, executives have opportunities to build digital businesses with little or no latency in decision making, whether by humans or by cognitive systems. Extreme confidence in real-time or near real-time decisions depends on the robustness of information, and how wide of a view it offers of the enterprise, its customers, its markets and the rest of its environment.

It's essential to identify and integrate a variety of information types, which is key to a successful digital transformation strategy. There has never been more pressure on organizations to improve products and services, move closer to the customer, stave off competitors and move fast to be disrupters themselves. A successful digital transformation strategy—built on new thinking and new technology—paves the way to success in this hyper-competitive economy. Every digital strategy requires analytical power, and such power requires that data be inventoried, its value assessed, and that it be made available across enterprises when and where it is needed.

The challenge today is a disconnect between the business and technology sides of enterprises, which interferes with and hamstring attempts at data-driven digital transformation. This disconnect takes several forms:

- Business users not being able to integrate and leverage data, as they don't know where the data is, what it is, or even who to ask for guidance.
- Business and IT teams not seeing eye to eye because they speak different languages.
- Business users often disagreeing over conflicting data points, often from different sources.
- The "democratization" of data—making data available and accessible to all who need it—is still in its early stages.

This disconnect occurs within many of today's organizations due to a lack of cohesive data governance; a lack of ways to find IT-certified, trusted data; and a general lack of visibility of enterprises' data landscape and related applications. As a result, seeking out the right information is more of a treasure hunt than informed analysis. The time has come for both the business and IT to enable their data assets to reach their full potential.

Intelligent data catalogs manage and provide access to metadata that underpins data-driven and digital enterprises. They point to critical data that fuels the analytical engines that ultimately power insightful decision making. This thought leadership paper explores the emerging role of intelligent data catalogs, and the growing role artificial intelligence and machine learning will play in enabling enterprises to organize and discover their data assets.

"Most organizations use only a small percentage of the data they have access to—in my experience less than 5%—even though they continue to collect and store terabytes of data," says Shervin Khodabandeh, partner and managing director with Boston Consulting Group. "For example, many utility companies have been installing smart meters with the promise that the intelligence hidden in this data will transform their business. However, most have not yet utilized any of this information in any meaningful way. Similarly, many retailers have been collecting data on customer transactions and behaviors online but are only marginally using this information to create more personalized customer experiences."

This is a common problem across just about every industry, he adds: "Too much data, too many variables, what to do with it and how best to utilize it. As a result, most organizations focus only on a sliver of data—that part that they intuitively understand and have experience in—and treat the rest as metadata. In reality, there is no such thing as metadata. It is all data. Today's data was yesterday's metadata."

## INTRODUCTION: THE ROLE OF INTELLIGENT DATA CATALOGS

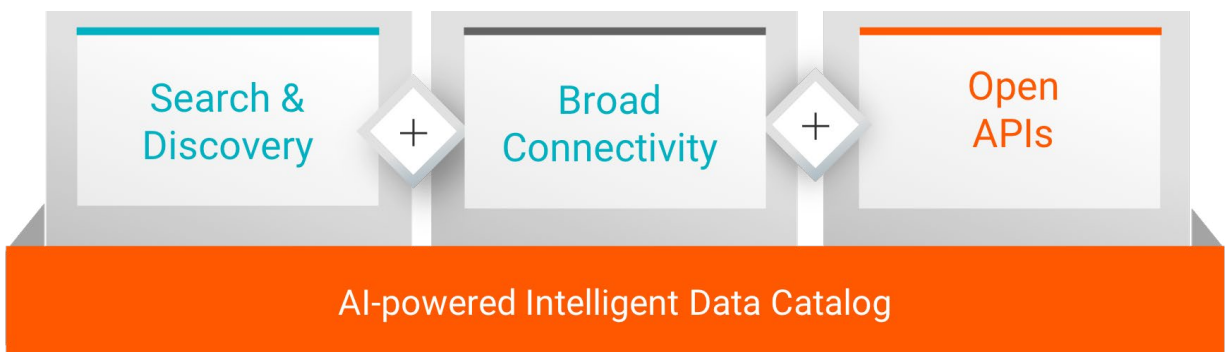
For years, data integration has been an elusive goal for organizations, with various forms of data hidden away in silos, departmental applications and external sources. With the rise of the digital enterprise, which depends on data from every corner of the enterprise and beyond, things are only getting more complex. Data sources and assets are exploding in size, creating a need for more automated ways to track and discover data. Intelligent data catalogs meet these requirements.

### INTELLIGENT DATA CATALOG, DEFINED

Intelligent data catalogs pick up where traditional data catalogs, metadata repositories and data lakes leave off. This new generation of catalogs not only provides directories of data assets but also incorporates automation that helps enterprises get the most out of ever-increasing stores of information.

Intelligent data catalogs represent the next evolution of metadata tools and repositories. These new catalogs take “metadata management from its backwater silos to a centralized cross-platform facility that is feature-rich and comprehensive,” according to Philip Russom of TDWI. “Imagine metadata extracted from all sources—whether on premises or in the cloud, whether within the enterprise, on the Internet, or at partnering firms. Metadata thus amassed is then improved and shared across an enterprise and beyond for unprecedented consistency, productivity, trust, and governance.”<sup>1</sup>

Intelligent data catalogs help business and IT users realize the full potential of their enterprise data assets by providing a unified metadata view powered by AI that includes technical metadata, business context, user annotations, relationships, data quality and usage. These catalogs are also employed to assist in the curation, governance and management of data.



<sup>1</sup> *The Data Catalog's Role in the Digital Enterprise*, Philip Russom, TDWI, 2017; <https://tdwi.org/research/2017/11/ta-all-informatica-the-data-catalogs-role-in-the-digital-enterprise.aspx>

---

## WHY ORGANIZATIONS NEED INTELLIGENT DATA CATALOGS

---

Data has exploded, to the point where it's difficult for humans to effectively manage the large volumes of data stores, structures and sources. In addition, data warehouses and even data lakes are inadequate for ever-expanding bases of data consumers. Put simply, today's knowledge workers and decision makers don't have the time to be tracking down or attempting to verify the data sources. How do you make them trust these assets and make informed decisions? How do you provide business and technical context about the data they are pulling from? Whom do they ask if they have questions? How do you democratize data?

These are all issues being tackled by Brian Conneen, CTO and CSO for Marlette Funding, an online consumer finance service. "We're never for lack of data—we have data coming from all sorts of sources," he says. "We have data coming from what the customer gives us when they apply; we have data analytics and tools that are measuring the web interaction; we have data coming from credit bureaus as part of the application and verification process; and we have interesting data from applications themselves, such as performance data." To help decision makers and analysts understand what and where these resources are, Conneen's team maintains a data catalog with access to 20 to 30 different data sources.

The previous generation of data management and integration solutions—especially data warehousing—may be too slow and cumbersome for today's fast-moving digital businesses. "Try to meet the data where it lives," Conneen says. "Don't try to over-engineer all the data with integration solutions. You may have a data source that you really love, and you spend three months getting it into your data warehouse. Then, three months later, the data isn't good anymore. So you just expended three months' worth of time and effort."

The challenges of democratizing data and making it available to decision makers include:

- **Data, especially unstructured data, is often unseen, siloed and unknown to decision makers.** While an estimate from IDC puts the amount of unstructured data at 90% of all digital data, a study from its parent organization, IDG, also shows that structured data is still a priority. Eighty-three percent of IT professionals said structured data initiatives were a high priority at their organizations, compared with 43% who assigned top priority to unstructured data projects. This suggests "that many organizations are missing out on an opportunity to derive significant business value from properly harnessing unstructured data."<sup>2</sup>
- **Numerous data sources.** In today's data-driven organization, data comes from many different sources, such as IoT devices, enterprise systems and social media. "Technical metadata documents data's structures, components, and data types. This is a foundation for data extraction and load, other computerized processes, and highly technical interfaces," according to TDWI.
- **Data on multiple platforms.** As organizations evolve to be more digital, they'll rely on both on-premises and cloud systems to store and process their data. As a result, data management takes on new dimensions, with data being stored in multiple places. Along with the need to identify and lock in new sources, enterprises need to ensure that data available from multiple types of systems is viable, accessible and secure.
- **Inability to fully leverage data analytics.** The companies that are entering markets with new innovations and ways of doing things are likely to be deploying data analytics to make it happen. Most of today's organizations have large amounts of data from various sources. However, many have not been able to leverage it to achieve competitive advantage.

<sup>2</sup> "Solving the Unstructured Data Challenge," Jaikumar Vijayan, CIO, June 25, 2015.  
<https://www.cio.com/article/2941015/big-data/solving-the-unstructured-data-challenge.html>

- **Systems are complex, and interactions between data sources and systems need to be better understood.** In an era when enterprises are turning to hybrid computing to leverage assets both within the cloud as well as in on-premises systems, data will be pulled and analyzed from a range of sources.
- **Lack of understanding around where data comes from.** One of the greatest challenges of ensuring trust in data is the inability of users to understand where their data comes from, how it was created and the legal rights to use of the data.
- **Real-time capabilities require well-vetted data sources available on demand.** The growing requirements for real-time interactivity, especially through artificial intelligence and machine learning interfaces, require that data be even more trustworthy and transparent.
- **Compliance mandates require searching and validating data.** A great deal of enterprise staff time and resources are being consumed in finding the right information and validating it for ongoing reporting purposes.
- **Users need—and expect—Google-like semantic searches for enterprise data.** Everyone — from consumers to employees to customers—is now thoroughly accustomed to the instantaneous results delivered by social media sites like Google or Twitter to queries. This sets a new standard for delivery of internal enterprise information.

## ARTIFICIAL INTELLIGENCE ENHANCES INTELLIGENT DATA CATALOGS

The emerging generation of intelligent data catalogs don't just manage, tag and store metadata, they also employ artificial intelligence (AI) and machine learning, which lead to a deeper and broader understanding of data.

When used with intelligent data catalogs, AI:

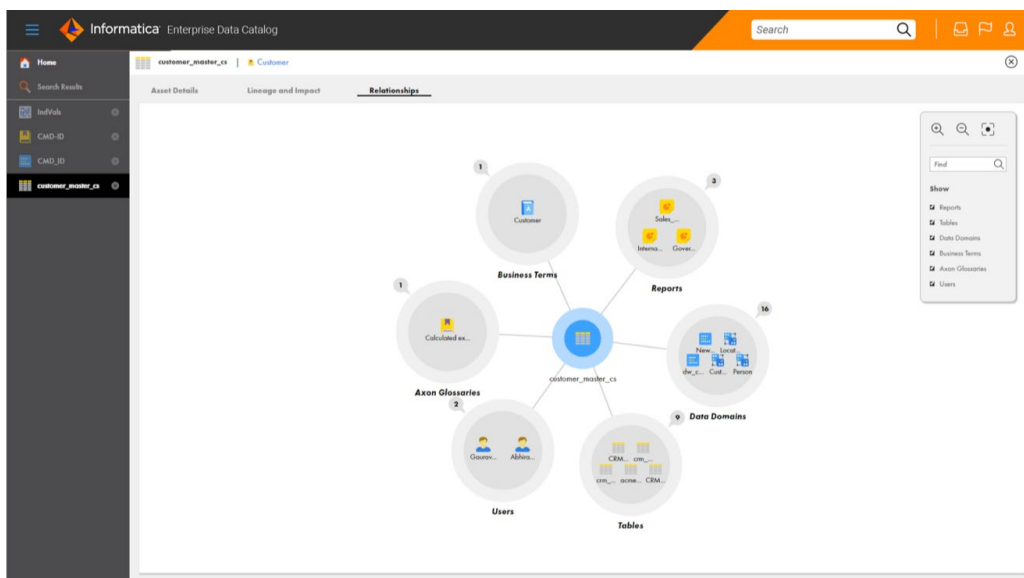
- Enables users to discover entities within unstructured data
- Uncovers highly complex relationships among data
- Provides an understanding of data lineage through end-to-end analysis of data from sources to consuming applications

Machine learning also enables semantic inference and recommendations. Some examples include discovering the structure of complex data (e.g., machine data); identifying similar and duplicate data sets; discovering composite entities across data sets; and recommending alternative or similar data sets based on usage, quality or provenance, among other factors.

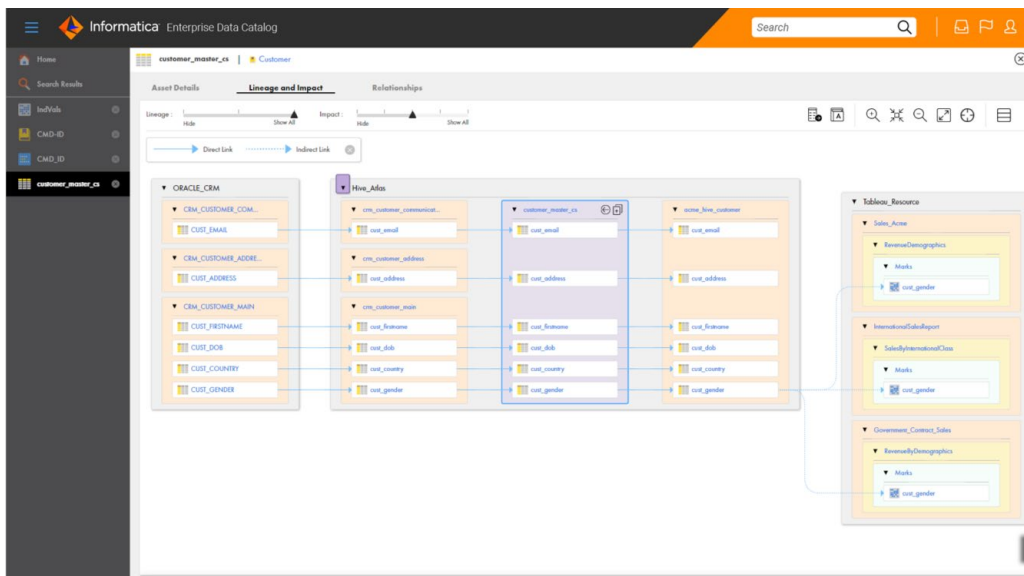
# THE IMPACT OF INTELLIGENT DATA CATALOGS

Intelligent data catalogs are part of many enterprises today, providing the foundation of digital transformation efforts. These catalogs not only point to the right data sources but also help manage the workflows associated with retrieving the data.

## HOLISTIC VIEW OF DATA ASSETS AND RELATIONSHIPS



## UNDERSTAND DATA LINEAGE AND IMPACT ANALYSIS



Flavio A. Zanetti, chief of staff and strategy and planning, digital transformation senior leader at Cisco, says it's key that the company's teams of 800 engineers, based across 42 countries, not only have immediate access to the information they need to serve customers but also understand how engagements are being handled. "We have different data instances, different databases, different data lakes across the company," he says. "We have to pull all this into the same place to find the best location for the content and the best source that our engineers are looking for."

Zanetti says his team has already seen a range of benefits from the use of data catalogs, mainly in the form of reducing time spent looking for documents and data and thus allowing engineers more time to focus on higher-level activities, like development, demos, proposals, customer engagements and discussions. "It varies, based on location, based on volume, but at the very least, it has given us 15% to 20% more productivity gains that we've seen so far," he says.

It is no longer just a handful of analysts or managers pulling data; almost everyone in an organization uses data for decisions. A data-driven organization with truly "democratized" data capabilities not only makes the right data available across the enterprise in a secure way but also has it managed by teams representing key disciplines. For example:

- Data scientists and analysts can take advantage of the self-service analytics associated with intelligent data catalogs to find relevant data assets, visualize lineage and get a 360-degree view of information.
- Data stewards and data owners using intelligent data catalogs can employ data governance that helps connect business glossaries to technical objects, verify business to technical lineage and track key data elements compliance.
- Data architects and developers can use intelligent data catalogs to employ data asset management, analyzing column-level lineage and change impact, as well as viewing transformation logic.
- Governance and security administrators.
- Data consumers.

There are seven key use cases in which data catalogs play a crucial role:



**1. Data Discovery and Understanding.** Unstructured data may be seen as the key that unlocks many new competitive insights, but within many enterprises, this information is hard to find, if not outright hidden. One report puts the percentage of organizations with limited or no understanding of their unstructured data assets at 78%.<sup>3</sup> An intelligent data catalog—powered by AI—enables users and applications to quickly find and discover trusted data from sources across the enterprise and beyond. It enables a unified view across all data stores, platforms and services. Curation and consumption is supported by automated inference of domains and entities, as well as search recommendations.

Intelligent data catalogs also provide the long-sought 360-degree view of all pertinent data—including data domains, users and related data sets—along with detailed end-to-end lineage, automatic scanning and tagging of data. This also opens up the advantages of crowdsourcing metadata based on usage, inference of data structures and avoiding data duplication.



**2. Data Governance.** The successful data-driven enterprise also requires effective data governance. As data catalogs scale and grow with the business, it's important that data be secure, and be aligned with business policies and priorities. Compliance mandates may also introduce requirements for the transparency and management of data assets. Data catalogs provide such alignment, automatically applying rules and policies to data assets as they are established within enterprises. While metadata has been closely tied to associated silos, applications or functions, an intelligent data catalog brings information about data sources into a centralized enterprise environment.

<sup>3</sup> "Five Steps to Unstructured Data Security," Michael Fimin, *ITProPortal*, January 25, 2017. <https://www.itproportal.com/features/five-steps-to-unstructured-data-security/>



Intelligent data catalogs also enable more seamless governance, regardless of how large or unwieldy data assets may grow, by automating such processes, rules and policies for compliance and data vetting. In addition, intelligent data catalogs can “automatically record data’s lineage as data is handled via metadata and other functions of the cataloging platform. Lineage should be detailed down to the entity and attribute level, with links among business terms within the glossary. With these details, data lineage enables fast and deep insights into data provenance and impact analysis,” according to Russom.



**3. Collaboration.** The automated governance capabilities delivered through intelligent data catalogs enable greater collaboration as data processes are opened to teams across the enterprise and users are automatically granted access to files based on their permissions levels. For example, users can tag, comment on, like or recommend data sets to other project team members. In addition, usage stats and metadata can help surface popular and relevant data sets, which users can share in a secure fashion. Data masking can be employed to obfuscate sensitive data, ensuring that it can be shared among authorized users without risk of exposing personally identifiable information or other sensitive data types.

Data catalogs not only foster greater communication but also help avoid duplication of effort across enterprises. “If somebody has a question or a need in Australia, we might have the same engineer answering the same issues in the U.K.,” says Zanetti. “If they’re all spending time researching the same question, we are wasting our resources.” Cisco’s sales team now provides access to proposals, specifications and all other related data in a single place, enabling users to see who’s doing what and share information accordingly.



**4. Real-Time Enterprise.** In today’s economy—which depends on IoT and the ability to address customer requirements as they arise—the ability to process and deliver insights in real time or near real time is critical. The traditional approach would require that users or analysts sift through all possible data sources. “That would take an analyst a couple of days to figure out the data,” says Conneen. Having an intelligent data catalog “cut down our time to 15 minutes to arrive at the same answer. We’re almost able to see live data, almost in real time.” At the same time, traditional data environments and data warehouses would also require considerable staff resources as well as time to effectively integrate data. “If we tried to integrate this data into one giant data warehouse, it’s all we’d ever do, spending time to integrate it.” The key is to have data cataloged “where it is, to make it available more quickly,” he says.



**5. Self-Service Analytics.** Today’s successful data-driven and digital enterprises have strong self-service ethics, meaning they actively seek to empower end-users to access the data they need, when they need it. Intelligent data catalogs make it easy for business users to identify and locate the data they need. This is critical, as business users—from the executive suite down—often don’t know what data assets their companies have, or where these assets may be located.

Previously, knowledge of data assets has been shared among users and teams by word of mouth or other offline means. Therefore, much of their time is spent searching or tracking down appropriate data. With data sources and volumes constantly growing or evolving, and end-user bases expanding, it’s difficult for IT departments to keep users updated. The majority of time has been spent searching for data, with less devoted to analysis.

Ideally, a self-service environment, supported by AI-driven data catalogs, will enable users and teams to devote more time to analysis, versus searching. “We’re taking disparate data sources and using technology to layer SQL-type syntax on top of them, so we can expose them to users’ dashboards—providing them self-service,” says Conneen. “We tell them, ‘this is a data set for application data,’ ‘this is a data set for web performance data,’ or ‘this is a data set for customer interaction data.’ They don’t really have to know all the different sources the data is coming from. They just need to know they can drill in and find reporting and dashboards.”



**6. Journey to Cloud.** Data stores are not only growing exponentially in size, but they are also drawing on sources beyond the enterprise walls. Business queries may increasingly be running against constellations of cloud-based sources, requiring an ability to discover and manage such data as seamlessly as that within enterprises. Both data sources and data targets may be in the cloud, or part of a hybrid system. Intelligent data catalogs open up all external data sources for discovery as part of the essential 360-degree view needed of customers, markets, operations and other important environments.



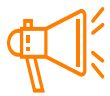
**7. Managing Organizational Transitions.** As organizations evolve through growth or mergers and acquisitions, there may be uncertainty about data assets within newer parts of the business or duplication of data between business units. Data catalogs will play a role in keeping track of data assets, regardless of their origin.

## HOW TO GET THE MOST OUT OF INTELLIGENT DATA CATALOGS

To compete in the fast-growing digital economy, enterprises need to have reliable and accessible data. An intelligent data catalog is a critical tool for ensuring that data is available and ready for the opportunities and challenges ahead. Here are three recommendations for getting started and making the most of intelligent data catalogs within your company:



**1. Work closely with the business to identify critical data sources.** Business users have critical data sets they regularly need to access as part of their usual work. In addition, there are likely many areas in which business users get delayed or work is bottlenecked due to the amount of time it takes to seek out data. Data impacted by compliance mandates is another area that needs to be managed in a more automated way. Rather than attempt to put everything into the intelligent data catalog all at once, start by identifying those areas that are most critical to users. In addition, it's important to understand the pain points of business users who use and consume data through self-service tools. For example, they may not understand the meaning behind technical terms associated with information they pull from data assets.



**2. Promote enterprise-wide engagement.** Just as data sources will come from across and beyond the enterprise, there needs to be assurance that data is not only available, but of the highest quality and as accurate as possible. This requires working closely with data providers as well as data consumers. Collaboration is key here—every employee has the opportunity to be a contributor as well as a user.



**3. Wrap the catalog within strong analytics and data governance.** The business needs to determine how the metadata will “allow better targeting, prospecting, issues resolutions, customer servicing, and so on,” says Khodabandeh. Other factors to consider when developing use cases are determining which variables are going to be predictive, he adds. This all hinges on strong analytics, as “often patterns are not obvious through traditional regression-based techniques but become evident through machine learning.” In addition, the catalog needs to be built within the guidelines of a data governance process “that allows systematic adding and removal of this metadata as customers change their minds, as well as when regulators come up with new regulations.” The data needs to be “removable, without models collapsing,” he adds.

To learn more, visit:

[www.informatica.com/products/big-data/enterprise-data-catalog.html](http://www.informatica.com/products/big-data/enterprise-data-catalog.html).



# Forbes insights

## ABOUT FORBES INSIGHTS

Forbes Insights is the strategic research and thought leadership practice of Forbes Media, a global media, branding and technology company whose combined platforms reach nearly 94 million business decision makers worldwide on a monthly basis. By leveraging proprietary databases of senior-level executives in the *Forbes* community, Forbes Insights conducts research on a wide range of topics to position brands as thought leaders and drive stakeholder engagement. Research findings are delivered through a variety of digital, print and live executions, and amplified across *Forbes'* social and media platforms.

### FORBES INSIGHTS

#### **Casey Zonfrilli**

DIRECTOR, ACCOUNT MANAGEMENT

#### **Tori Kreher**

PROJECT MANAGER

#### **Todd Della Rocca**

PROJECT MANAGER

### EDITORIAL

#### **Erika Maguire**

EXECUTIVE DIRECTOR

**Kasia Wandycz Moreno** DIRECTOR

**Hugo S. Moreno** DIRECTOR

**Joe McKendrick** REPORT AUTHOR

**Zehava Pasternak** DESIGNER

### RESEARCH

**Ross Gagnon** DIRECTOR

**Scott McGrath** RESEARCH ANALYST

### SALES

North America

**Brian McLeod** EXECUTIVE DIRECTOR

*bmcleod@forbes.com*

**Matthew Muszala** DIRECTOR

*mmuszala@forbes.com*

**William Thompson** MANAGER

*wthompson@forbes.com*

**Kimberly Kurata** SALES EXECUTIVE

*kkurata@forbes.com*

EMEA

**Tibor Fuchsel** MANAGER

*tfuchsel@forbes.com*

APAC

**Serene Lee** EXECUTIVE DIRECTOR

*slee@forbesasia.com.sg*