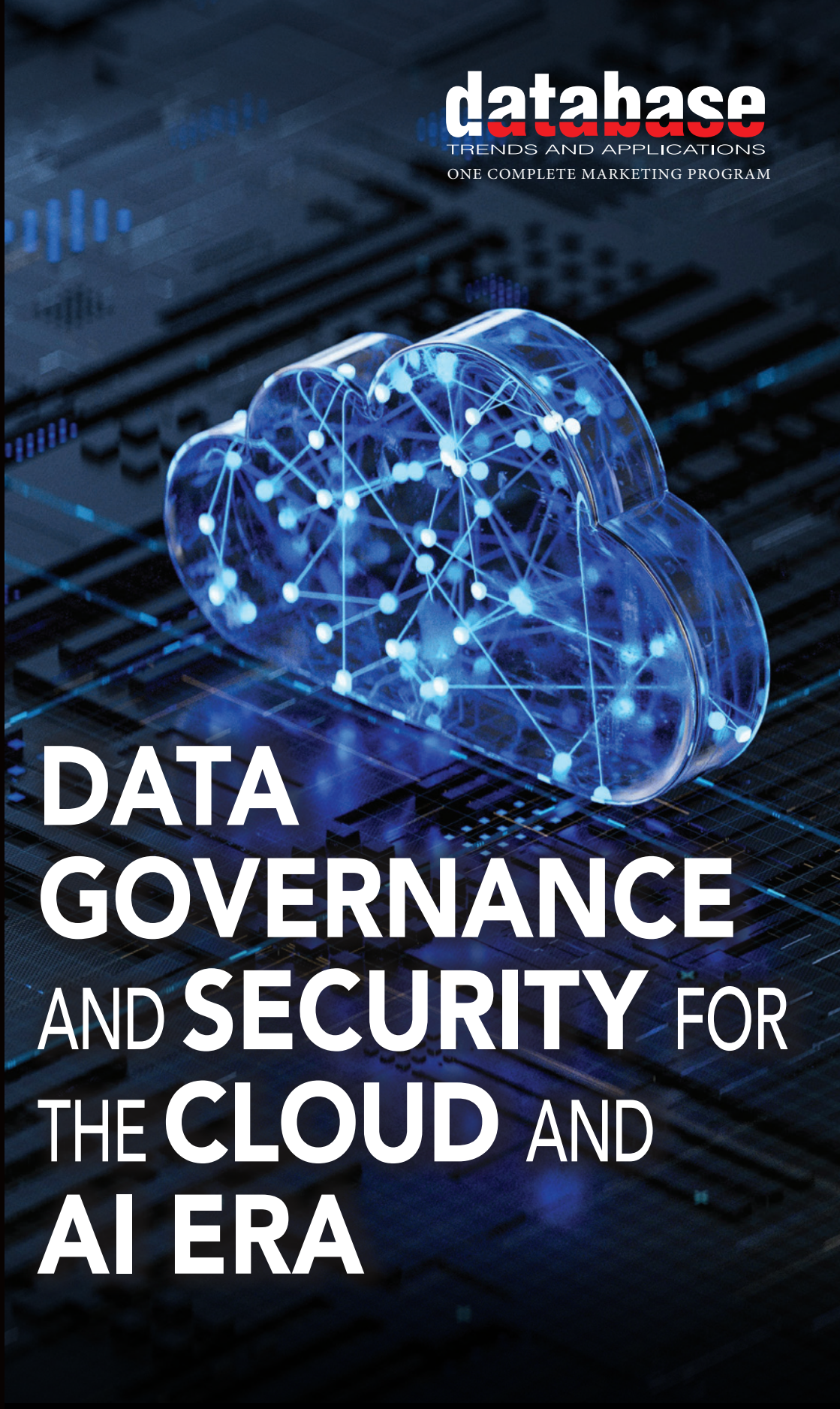
database
TRENDS AND APPLICATIONS
ONE COMPLETE MARKETING PROGRAM



DATA GOVERNANCE AND SECURITY FOR THE CLOUD AND AI ERA

database
TRENDS AND APPLICATIONS

Best Practices Series

DATA GOVERNANCE AND SECURITY FOR THE CLOUD AND AI ERA

Best Practices Series

As AI and machine learning (ML) take hold across enterprises, there's a push to ensure that models and algorithms associated with these emerging technologies are viable for the business and that the data employed is secure. This calls for new standards for both secure and responsible AI-based data. The data needs to be protected against cyberthreats and deliver fair and unbiased results to decision makers and business leaders.

As organizations develop new data analytics and AI projects to compete in today's digital economy, the protection and management of data flowing into AI systems have become challenging aspects for IT and data leaders. This is key, as many companies are betting their businesses on AI-powered applications such as predictive analytics, customer relations, and product development.

That's why the data flowing into AI systems needs to be of the highest possible integrity and quality. As businesses aggressively move into both generative and operational AI, they require massive amounts of accurate and timely data. However, only 23% of data managers surveyed by Unisphere Research, the research arm of *Database Trends and Applications*, expressed full confidence in their data. Close to one-third said that data quality is a constant, ongoing issue. What's more, new data analytics and AI projects are now surfacing many of these issues. The survey, part of a series launched in 2021, reports no letup in the growing confidence gap in the data needed to support next-generation initiatives ("Unfinished Business: Taking on the Data Quality Challenge in the Age of AI," Unisphere Research, a division of Information Today, Inc., November 2023).

The stakes are much higher now. The rise of large language models (LLMs)—both publicly available as well as contained within enterprises—to support business decision making and customer communications means data is being pressed into service in new and highly demanding ways as training data and real-time streaming feeds. The models need to support growing demand for intelligent, customer-facing applications such as chatbots and conversational interfaces, as well as intelligent assistants for internal enterprise operations.

AI data governance, of course, isn't limited to protecting a company's assets and brand—it also helps protect society at large, especially in terms of compliance with laws and regulations issued by governments.

Add to this the ongoing incidents of cyberattacks, hacks, and security breaches that have become almost daily events. If ever there was a solid business case for AI data governance, this is it.

AI data governance, of course, isn't limited to protecting a company's assets and brand—it also helps protect society at large, especially in terms of compliance with laws and regulations issued by governments. For much of the world beyond enterprise walls, AI has become a hot-button issue.

The following are key considerations for building an AI data governance framework:

- **Begin the AI data governance process as a business initiative.** AI data governance is an initiative that needs to extend well beyond the IT department—ownership and responsibility for AI should reside with anyone who works with technology. Technical teams can implement solutions and ensure that systems, models, and data are delivered. However, it's up to the business to establish the mission and purpose of AI initiatives along with the guardrails that ensure data isn't compromised. For many organizations, this may mean a cultural change, including a push to incorporate responsible AI into day-to-day operations, as well as longer-term strategies. Education and training are essential to developing an AI-ready culture.
- **Assess the current AI footprint.** Inventory your current AI systems and assets. This may require working across enterprise lines to better understand what AI tools, platforms, or applications are being used or considered—including the more informal use of outside AI resources, such as ChatGPT.
- **Understand that trust is at the core of governance.** While enterprises and their employees are enthusiastically embracing AI—particularly generative AI (GenAI)—to assist in their work, there is still a lingering mistrust of AI-generated decisions, especially among executives and managers, who continue to rely on gut-level decisions. Effective governance will help assure that the insights produced by AI systems will be based on well-vetted and relevant data. AI data governance should be geared to instill trust in the data that supports AI insights and recommendations.
- **Integrate data governance with AI governance.** These two areas are closely connected. For instance, governing data applies to AI, which is now part of the realm. Data governance, which seeks to oversee the security, quality, and business value of data, lays the groundwork for AI implementations. There are already standards and procedures, developed across several decades, governing data use. The challenge now is to expand upon these governance tenets to extend to AI data. Efforts to bring these two realms together, however, can prove to be challenging, as AI and data teams tend to work separately and need to be brought together.
- **Employ AI to also power governance solutions.** Ensuring the viability of data going into AI initiatives is one side of the coin. The other is the emerging role of AI in supporting governance, through automatic monitoring and providing course corrections for governance and security, based on continuous learning cycles. Solutions are emerging that not only employ GenAI to build SQL queries but also ensure security and compliance.
- **Understand that AI data governance is about security—and much more.** Successful AI data governance addresses potential AI risk factors, as well as assessing business value, tracking data use, providing transparency, assuring privacy, flagging ethical risks, and assuring compliance. In addition, AI presents potential risks with violating intellectual property—even if it is restricted to internal use.
- **Remember that AI data governance can't be "outsourced" to cloud or third-party providers.** The onus is on enterprises to put processes in place to ensure their data is secure and relevant. Cloud providers can deliver broad security from the outside, but it's up to organizations to oversee internal access and adherence to policies and procedures.
- **Keep humans in the loop.** No systems, no matter how current the algorithms and data feeds, are perfect. It's important that the people charged with deploying and using AI are trained in the ethics and risks of these systems. Importantly, people need to be encouraged and ready to hit the "stop" switch and challenge the results delivered by AI systems.

Comprehensive AI data governance will help deliver the new capabilities AI is poised to offer to organizations. Ensuring AI data is both secure and responsible stems from regular and sustained engagement by decision makers and technology teams. This will build trust in AI and the results it delivers, which promise to accelerate business growth in today's economy. ■

—Joe McKendrick

Powered by Data: E.ON's Journey

E.ON powers data readiness, metadata management, data governance, and data modeling with erwin® by Quest® to ensure energy continuity for 51 million customers.

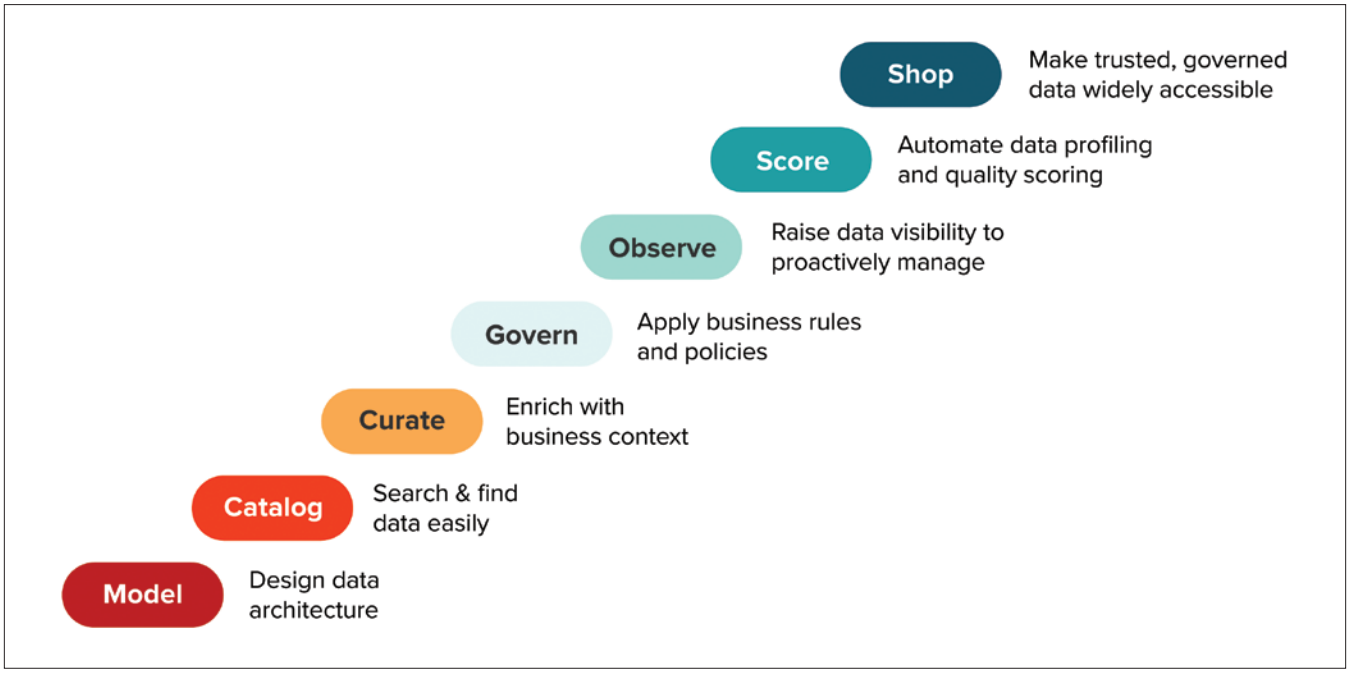
E.ON began its data readiness (DaRe) journey in 2018 after realizing that taking better advantage of data and data intelligence would help improve customer services, increase efficiency in energy production, and make it easier to meet the regulatory requirements of the countries where E.ON operates. With E.ON's data landscape being enormously heterogeneous—including several thousand applications where data is stored (SAP, Microsoft Azure, Oracle, old legacy Windows servers, and others)—the company needed a data intelligence software that could map as many data sources as possible. The E.ON DaRe team spent months evaluating metadata management and data governance capabilities to ensure they would select the right data intelligence software provider that could help them not only kickstart their data journey but be integral in helping them achieve their long-term goals.

"Our purpose is to ensure enterprise-wide deployment of our data as a strategic asset."

—Romina Pyplacz, Head of Data Management and Governance, E.ON

Guided by its vision, E.ON selected erwin by Quest as its software provider. erwin Data Intelligence by Quest was initially deployed for automated metadata management, data lineage documentation, and impact analysis. With these capabilities, E.ON has improved data quality, detected data redundancies, and achieved IT efficiencies that were not originally planned for when first starting their data readiness journey. "We use erwin for the auto-configuration of data pipelines," said Romina Pyplacz, Global Head of Data Management and Data Governance at E.ON. "I did not imagine in 2018 that I would use a data catalog to automatically configure data pipelines, but I can. I can do things that I didn't even think about when I got started."

The E.ON DaRe team has used the data governance and data literacy capabilities of erwin Data Intelligence to deliver a standardized, business-focused data catalog to improve IT and business collaboration, as well as alignment. By defining and standardizing common business metrics such as KPIs and making associated data asset relationships clearly visible within erwin Data



Intelligence mind maps, E.ON has been able to provide clarity for business stakeholders.

Emboldened by the success of these initiatives, E.ON expanded its erwin partnership with a focus on data modeling harmonization by adopting erwin Data Modeler by Quest, and began to explore the addition of data quality, data provisioning, and data-as-a service as key E.ON data readiness offerings.

E.ON's DaRe team began as a team of only two E.ON employees in 2018 but has expanded to include more than 30 experts operating within 24 business units. erwin by Quest is used by more than 600 E.ON users today for data modeling, intelligence, and governance.

BENEFITS

- E.ON identified more than 2 million Euros in business impact after 24 months.
- E.ON estimates a savings of 30 percent on external data management costs and a 50 percent reduction in time spent on data discovery because better data availability and quality have a direct impact on productivity for each data-driven activity across the enterprise.
- By delivering "data governance as a service" for one-stop data access and lineage, E.ON has standardized its vocabulary, met regulatory compliance requirements, and has been able to understand and then optimize how data is used across global operations.

Read the full case study at: https://www.erwin.com/casestudy/powered-by-data-eons-journey-to-sustainable-energy/

7 WAYS TO MAXIMIZE THE VALUE OF YOUR DATA WITH ERWIN BY QUEST

With proven, award-winning, and integrated **erwin Data Modeler** and **erwin Data Intelligence** software, you can rapidly mature your organization's data intelligence and governance efforts ensuring the right people have the right data, for the right use, at the right time.

Visit www.erwin.com to learn how data modeling, data catalog, data quality, data literacy, and data marketplace capabilities can help you deliver high-value, trusted data widely throughout your organization. ■

LEARN MORE ABOUT THE 7 STEPS TO DATA MATURITY

https://www.erwin.com/docs/7-steps-to-maximizing-the-value-of-your-data-ebook-ebook-30983.pdf

Data Governance & Security for the Cloud and AI Era

The era of the data-defined business is upon us. The most successful organizations are those that effectively and efficiently capture, analyze, and leverage data as the central operating model for every decision, innovation, and investment. In the data-defined business, everything starts with data. And while this may seem intuitive and straightforward, in reality it is rife with complexity.

The biggest challenge isn't the lack of advanced technology or a dedicated team. It's the inability to quickly and securely enable data sharing and collaboration. Siloed data use across lines of business, manual processes that are slow and burdensome, and concerns about inadvertently exposing data, or failing to comply with regulations, all hold teams back from being able to access and share data effectively.

"In the age of cloud and AI, data security and governance complexities are mounting. It's simply not possible to use legacy approaches to manage data security across hundreds of data products."
—Sanjeev Mohan, Principal at SanjMo

Data sharing has always been beneficial, but it may now be the tipping point for organizational success. The ability to share data fuels BI, ML, and AI applications and workloads, giving companies a tangible competitive advantage. And while some of these technologies have existed for years, GenAI in particular is a new frontier that has quickly become a critical tool powering organizations forward.

But a longstanding challenge looms as an obstacle that could hinder the rapid development and implementation of AI applications: balancing fast data access with security and governance.

When it comes to AI, the choice between speed and security takes on greater urgency than ever before. AI technologies are evolving mind-bendingly fast. Open source vendors offer catalogs of models that accelerate development. Alongside foundational and large language models (LLMs), smaller, domain-specific models are quickly gaining traction as data-defined businesses seek sharper insights and more precise outcomes. And the speed at which data can be provisioned directly correlates to how much value and competitive advantage AI offers.

THE AI SECURITY CHALLENGE

But AI and systems of intelligence are unfortunately black boxes that offer little transparency about how models are built, trained, and deployed. What data is being fed into AI applications? Is it authorized to be used for that purpose? Is there any sensitive information that could be compromised? What is being done to preserve privacy while avoiding bias?

Shadow AI is also a problem. With users implementing their own programs outside the remit and control of centralized data teams, effective data governance becomes even more elusive. In a survey of 700 data professionals, 88% said their organizations' employees are using AI, but just 21% had established acceptable use policies. It's no wonder that half of respondents reported that their company's data security strategy is failing to keep up with the pace of AI evolution—a pace that is only accelerating.

Another study found that while 73% of people viewed generative AI tools as a business threat, 34% were willing to implement them quickly despite the potential consequences. It's a sign that the promise of AI's results is considered worth risking noncompliance, inadvertent exposure, or other harmful repercussions.

Since ChatGPT exploded into the public domain in late 2022, the flurry of fast-moving AI developments has gained the attention of lawmakers. Legislation like the EU AI Act, the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, and New York City Local Law 144 aim to wrangle how AI applications are developed and deployed, but even they are unable to keep up effectively. In the absence of a wide-reaching regulation like the GDPR, it's incumbent upon organizations to use AI responsibly. The question is not why, but how?

ADAPTING TECH AT THE SPEED OF AI

As this report makes clear, navigating the complexity and uncertainty of the changing AI landscape requires a proactive approach to data security and governance. This is particularly true as the volume of data sources and users continue to grow

exponentially, and the pace of data analytics accelerates. Data is a currency that is now ubiquitous and necessary to compete; organizations that manage and leverage it with as little risk as possible will be the most successful.

But in this AI era, traditional approaches to security and governance are quickly becoming obsolete. Unstructured data is the primary input for training and fine-tuning models, supplemented by structured data. While most existing solutions protect structured data in compute layers, today's teams need the ability to enforce controls on unstructured data in storage as well. Governing unstructured data is an entirely different—and more difficult—task that requires different approaches to access control and monitoring.

Further complicating matters, most organizations have a diverse tech stack, comprising several cloud platforms. To ensure these platforms are truly interoperable, they must have standard, consistent controls on all data, no matter where it lives. Trying to manage this manually on a platform-by-platform basis is not feasible at scale, particularly when AI workloads are involved. Mapping roles to policies, and ensuring those policies are consistently enforced and maintained across all platforms, is a full-time, error-prone job.

A comprehensive data security platform like Immuta, which automates data discovery, access control, and monitoring at scale, protects data for any use case—including AI. By enforcing controls and advanced masking techniques at the data layer, Immuta provides the speed and flexibility needed to ensure that data is only used by the right people for the right reasons—no matter where it lives.

Immuta is platform-agnostic, so a single, granular policy is able to cover multiple platforms with no intervention required. And since it sits between the data and the models, policy enforcement is context-based—in other words, access is neither too broad nor too narrow to satisfy data needs. To ensure policies are working as intended, continuous monitoring and auditing capabilities proactively identify anomalies for remediation. If data is inadvertently exposed via an LLM or other AI application, you will know before it's too late.

For Swedbank, the largest banking group in Sweden and third largest in the Nordics, moving from on-premises to cloud storage and compute was a business imperative in order to support deep learning and AI. Building out their AI capabilities would allow Swedbank's data team to detect financial fraud and suspicious activity, saving time, money, and resources, while maintaining compliance with strict industry standards.

"Organizations have to continue to embrace new technology to remain competitive and relevant in today's economic landscape. Still, the introduction and integration of AI-based solutions create complexities for security teams, who will have more to manage and oversee than ever before."

—Mike Scott, CISO at Immuta

DE-RISKING AI TO DRIVE INNOVATION

By integrating Immuta into its tech stack, Swedbank doubled its data use cases, improved process efficiencies by 3x, and accelerated onboarding of new users. The ability for more people to confidently access more data proved to be a business driver, with no manual overhead required.

Beyond these outcomes, more efficient and scalable data security allows companies like Swedbank to leverage the latest and greatest advances in AI. From generating chatbots and predictive maintenance needs, to improving customer service and hiring processes, there is virtually no end to the ways AI and ML will deliver new value to modern enterprises.

What will separate the most successful organizations from the rest will be their ability to de-risk data so that they can leverage it with confidence for novel, innovative use cases. Data introduces new possibilities; de-risked data delivers them safely. ■

IMMUTA
https://www.immuta.com

Prioritize the Accessibility and Searchability of Data with Indexing, Metadata, and Artificial Intelligence

Information today is growing exponentially, and it's impossible to manually look for files at the same rate at which content is generated. A whopping 78% of organizations are already overwhelmed by the vast volume, velocity, and variety of information created from technology usage. And the amount of information entering organizations is only expected to continue growing, with over 60% of that data being unstructured.

Greg Grospitch, Senior Consultant at Access, put this challenge into perspective by explaining that, "every year we're not growing [in information] by some clean factor of X. It's exponential growth every single year. And at some point, the infrastructure and the planning are going to be too much to get your arms truly wrapped around."

This challenge is made even more troublesome for information professionals with our society's heavy reliance on swift and precise access to vast amounts of data. Nobody wants to wait days for a piece of information to be delivered to them, nor does anyone want to spend that much time searching for it.

LOCATE INFORMATION FASTER WITH METADATA AND INDEXING

To meet this demand of retrieving information rapidly, organizations must have a thorough records management strategy that includes metadata indexing.

The metadata schema defines the attributes and characteristics of the data, while the index provides a streamlined way to access that data quickly. A robust data system often utilizes both a well-structured metadata schema for detailed information and an index for swift and efficient retrieval—eliminating the need for extensive digging.

Any filing system you implement, whether numeric, alpha-numeric, or terminal digit, must achieve these key goals to be successful:

- Systematic, predictable metadata tagging
- Systematic, predictable file storage
- Physical or logical grouping based on topic similarity

Like well-managed paper files that are stored using helpful categories for future retrieval—such as client name, year, employee ID number, etc.—electronic files must adopt a similar approach with structured, hierarchical metadata and indexing.

Imagine navigating through an unfamiliar city without a map or any street signs while searching for a specific restaurant – that's what searching for a document without metadata is like. Progress is slow, and reaching your destination becomes uncertain.

A well-constructed metadata indexing schema will be effective in any environment because it contains the critical information necessary for successful searching and fast document retrieval.

HOW TO START APPLYING METADATA TO YOUR DOCUMENTS

Now that you understand why metadata is crucial, you are likely considering the lack of resources you have in place to start applying it to your documents. Digitizing files and indexing are time-consuming, so partnering with a vendor may be the solution you need to help reduce workflow inefficiencies.

By leveraging the latest advancements in artificial intelligence (AI) to identify and classify record types and metadata, Access Unify™ | Active File Service ensures your documents are stored and digitized in a way that prioritizes the accessibility and searchability of data. So, instead of searching high and low for a document, with Access Unify | Active File Service, you simply request it via the Unify portal or through your line of business system, and Access will deliver the file digitally in under four hours.

The sooner you begin to implement metadata indexing, the quicker you'll support your employees in day-to-day operations and witness a swift transformation toward a digital-first model.

Don't wait to get your document management strategy in place. The journey from an information-challenged to an information-empowered organization is about moving from repetitive, unproductive cycles to a streamlined, efficient, and strategically focused workflow. Access Unify | Active File Service stands out as a tool that utilizes AI and machine learning to address the immediate challenges of information management and paves the way for digital transformation and long-term business success—a win-win for everyone involved. ■

FOR MORE INFORMATION on Access Unify | Active File Service, speak with an Access consultant today at https://www.accesscorp.com/contact-us

Governing Data for Cloud & AI Environments

IRI Voracity
An Insatiable Appetite for Data

As stewards for the newer sources of data in the cloud and feeding AI models, today's GRC teams and developers must grapple with increasingly complex requirements for data security, compliance, and quality.

THE NEWER LANDSCAPE

Recent studies by Database Trends and Applications (DBTA) confirm that IT departments are dedicating more time and effort to data governance and security than ever before. Let's delve into the factors driving this trend:

1. **Data Proliferation:** Organizations are facing an explosion of data sources and formats. From structured databases to unstructured logs, the sheer volume of information demands effective governance.
2. **Distribution Challenges:** Data transcends traditional boundaries. It flows rapidly between cloud, on-premise, edge and remote platforms. Ensuring consistent governance across this diverse landscape can be hard.
3. **Insatiable Demand:** Business users crave fast, easy access to data insights. Whether for reporting, analytics, or AI model training, the pressure to deliver timely information is relentless.
4. **Security Imperatives:** Evolving cyber threats pose risks to data integrity and confidentiality. Protecting sensitive information while enabling data access requires a delicate balance, in both production and test environments.
5. **Compliance and Quality:** Stricter regulations demand adherence to data privacy rules. Simultaneously, maintaining data quality standards is essential for accurate decision-making.

MODERNIZING AND IMPROVING DATA GOVERNANCE

IRI Voracity is a comprehensive data management platform designed to address these challenges head-on.

Voracity supports a range of data governance functions—especially for cleaning bad data and masking sensitive data—for a myriad of data sources on-premise and in the cloud. It uniquely offers:

1. **Unified Capabilities:** Voracity integrates data discovery, integration, migration, governance, and analytics within a single, intuitive interface. Built on the Eclipse™ framework, it streamlines the entire data lifecycle.
2. **Integrated Data Quality:** Standalone or during ETL, masking, wrangling, reporting, or migration jobs, Voracity users can validate, filter, scrub, de-duplicate, enrich, and standardize data in structured tables, files, and streams.
3. **Robust Data Masking:** Voracity excels in data masking and test data synthesis—critical data security activities. By classifying, discovering, and consistently obfuscating or synthesizing data, you can comply with data privacy regulations while enabling data utilization and preserving referential integrity.

REDEFINING TRADITION

Modern data governance and security practices must be scalable, agile, and adaptable to accommodate cloud and AI use cases. Voracity embodies this shift from traditional approaches, enabling organizations to harness the full value of their data while simultaneously governing it.

More specifically, Voracity and its included "IRI Data Protector" suite component products like DarkShield in particular, can search, report on, and mask a broad range of data classes pertaining to multiple data privacy laws across an even broader range of structured, semi-structured, or unstructured data sources.

Voracity also supports the ongoing cloud and AI movements through:

1. **Design Efficiency:** Voracity streamlines data processes, optimizing workflows from discovery to analytics. Its unified data management capabilities and consolidated data manipulation reduce project and design complexity, enabling teams to work faster and smarter.
2. **Runtime Speed:** Powered by CoSort and its built-in resource management features, Voracity saves time through smarter algorithms and job configurations.
3. **Team Collaboration:** A familiar GUI and common metadata foster a shared job design environment, bridging typical gaps between IT and business intelligence teams. As collaboration becomes seamless, governance and insight improve. ■

LEARN MORE ABOUT Voracity and the data-centric governance and security capabilities at https://www.iri.com/products/voracity/technical-details#capabilities.