00.00.00

Velotix

SCALABLE AND SECURE DATA ACCESS:

HOW TO BUILD THE BEST-OF-BOTH-WORLDS DATA SECURITY PLATFORM

WWW.VELOTIX.AI

EXECUTIVE SUMMARY

It's one of the biggest challenges in data governance today: how to make data accessible and secure at the same time. The size of this challenge is matched only by its complexity, with data residing across warehouses, lakes, and even static spreadsheets or within databases. Multiple decentralized locations also means multiple forms of legislation across industries and regions, compounding the challenges for those tasked with managing governance, control, and access.

Within the business there are also different drivers to balance. IT may be tasked with provisioning access while also maintaining security. End users, such as data science teams, seek access to relevant, enriched, and timely information. Data governance teams have to maintain visibility for compliance and audits. Public-facing employees have to reassure customers that their data is being processed correctly.

It's possible to align these disparate elements. We know because we've witnessed and supported the transformation for many customers – often operating within tightly regulated industries. Having trodden this path many times, we've now decided to distill the learnings and share them with you.

In this paper you'll find six key requirements for scaling your data access for greater insights – without compromising on security:

- **Leveraging Policy-Based Access Control (PBAC)**
- **Enforcing access policies dynamically**
- **Establishing your single source of truth**
- **Distributing responsibilities and unifying control**
- **Centralizing audits for greater visibility**
- **Architecting a flexible future**

WHAT A DATA ACCESS CONTROL FRAMEWORK SHOULD LOOK LIKE

Let's start with the destination: a secure data platform that provides data access dynamically. Rather than one based on a rigid approvals process and delays that make data out of date, irrelevant, or less useful.

Accurate metadata should underpin the data, indexing and surfacing the content in a way that bridges silos. There should also be integration with any existing platform, to maximize existing investments and minimize the need for extensive change management.

The result is the right information, made available to the right people, at the right time. Every part of the data lifecycle is mapped, to bring business value across:

- **Creation**
  Business data is captured, tagged, and categorized for later use
- **Storage**
  Costs and vulnerabilities are minimized by using dynamic data from a centralized epository, rather than relying on duplicated databases
- **Usage**
  Insights are based on users knowing, understanding, and trusting the data available, without any explainability or transparency risks from "black-box" AI
- **Archiving**
  Data is protected, labeled, and archived accordingly, with data set insight coming from assigned data stewards, making it easier to scale and localize data governance
- **Deletion**
  Actions are based on a visible and transparent lifecycle, ensuring compliance and avoiding unintentional data loss

LET’S NOW EXPLORE WHAT’S
NEEDED TO REALIZE THIS REALITY.

1. LEVERAGE POLICY-BASED ACCESS CONTROL (PBAC)

Go back to the start of most organizations' data control strategies and you're likely to findRole-Based Access Control (RBAC). Mirroring the nature of internal corporate structures, and granting access and privileges to individuals according to their role. This also reflected the top-down way in which information would be passed through the company.

Over time, this role-based approach started to show its limitations as data – and business –evolved to be more fluid, dynamic, and unstructured. Information started arriving from multiple sources, requiring input from more systems, which reduced visibility. That's when businesses turned to Attribute-Based Access Control. Of course, this provides more granularity in terms of users, resource classifications, and environments (location, IP, time zone, etc.).

Time to go beyond Attribute-Based Access Control

However, limitations remain with ABAC. It's still systems-based, with evaluation of characteristics such as the subject making the request and their environment. Added challenges come externally, through the increasingly fast pace of regulation. China, Brazil, and the US are all expected to introduce new data protection and privacy legislation during 2023. New laws mean new roles, audits, and authorizations. At scale, managing a patchwork approach becomes impossible for organizations seeking to grow, especially across various jurisdictions requiring different forms of compliance.

These trends are why it's time to redefine the concept of access control once again. This time, by going beyond roles and attributes. The solution is Policy-Based Access Control (PBAC), where access privileges are determined dynamically, based on contextual and finely grained policies. PBAC also means that access is available to all applicable users, based on the validity of the request in real time. Factoring in further conditions, such as location, helps add a more precise and automatic layer of control.

There's more finely grained fluidity too, with the permissions and rights tailored to the policy, rather than the person's attributes. Changes can be made based on new regulations or internal recommendations, without having to update role criteria and attribute privileges. Reports and logs are automatically adapted and updated, ensuring high visibility and transparency. PBAC's dynamic nature also makes it ideal for large organizations with complex authorization, control and access requirements.

Also, by adopting this method of control, data owners and controllers can code policies using natural language – an advantage over ABAC platforms which may be written using complex XACML principles. Naturally, this gives non-technical users further opportunity to manage data access and benefit from platforms that offer self-service. The end result is greater agility for the data access system across different areas of the business.

2. ENFORCE ACCESS POLICIES DYNAMICALLY

Data has evolved to be more dynamic and cloud-centric. But many data storage system remain static, relying on replicated data sets and simple access/deny requests. These involve manual configuration, such as applying Rights Management Services protection to encrypt Microsoft files containing sensitive information.

This means that policy decisions have to be made on static copies of data, with attributes defined and matched to the relevant policy. Resources are required to replicate the data and make it available to users wherever they are. Any masking has to be applied via slower batch processes, creating access request bottlenecks. What's more, because the replication is a snapshot of a specific point in time, data can soon become redundant after a change.

For the end user, this can limit time to action for data access requests. Alongside costs of data storage and integration, companies also face increased risk from unrestricted copies of sensitive data being held across different apps.

Granularity without complexity

Instead, access control has to be as dynamic as data. This calls for automated policy management and implementation, powered by AI. With methods including anonymization, tokenization, and masking, all of which ensure granularity and security without complexity.

This combination of intelligence and automation allows for access based on variables such as data classification and user status. Complex cases requiring human input can include user behaviors, exceptions logged from edge cases, and large-scale regulatory changes.

By focusing on the data, rather than the user, access controls are decoupled from the manual revoking and reassigning of permission. Instead, a change in one criterion doesn't have to impact all permissions. This means access controls are more resilient to changes in roles and regulations. It also means you go beyond the limitations of coupling data to a data warehouse, which by its nature means data isn't available for real-time processing, monitoring, or access. Decoupling data from applications means policy modifications can be made from a central access-based data network, rather than through hundreds of code changes.

3. ESTABLISH YOUR SINGLE SOURCE OF TRUTH

Data granularity and visibility have a symbiotic relationship within data catalogs. You need both working in harmony, so that you can build a unified and enriched metadata layer. Categorization can be used to implement rules around naming, matching, and parsing. This provides the framework for how your business will use and access information, as well as how you control the quality of information.

That's the theory – in practice there are often delays and inconsistencies. For example, imagine a change made in the data warehouse with new data being added. A traditional approach may involve the data owner logging the changes in a spreadsheet, for governance to validate. The spreadsheet can then be imported into the data catalog, with metadata, tags, and descriptions also being added.

It's not possible, or desirable, to replicate this lengthy process at scale. Instead, it requires a central repository for accessing information about users, accompanied by a single dashboard to grant and revoke data access permissions. Classifications should be established centrally, with rules to avoid inconsistent labeling.

Unifying and cleansing the data brings structure to relationships between entities, and lays the foundations for automated metadata management. The ultimate objective is to move away from batch updates and metadata that's only as accurate as the last manual update, crawl, or ingestion. Reducing the time between updates, ideally down to real time, is crucial with a view to avoiding the risk of metadata or schema drift. It also makes it easier to detect anomalies and unmoderated or unauthorized changes, even at enterprise level and at large scales.

Bridging silos to build collaboration

Data tagging is an essential part of the process, standardizing taxonomy across aggregated data catalogs. The resulting enriched metadata queries can then go down to rows and columns. This granularity is the level necessary for sensitive data, PII, and other information subject to regulation. There's more discoverability, a unified single source of truth, and unlocked value by surfacing relevant and related items your organization might not even realize are available.

By improving data discoverability with tagging and cataloging, the data fabric layer is improved. This also builds trust for users and data subjects by aggregating and optimizing data and its related rules and models. Exceptions and outliers can be more easily identified and processed as edge cases with employees adding insight in the role of human-in-the-loop. Meanwhile, AI can be deployed to automate and standardize the more routine data access requests functions.

4. DISTRIBUTE RESPONSIBILITIES AND UNIFY CONTROL

With data residing in one centralized platform, it becomes possible to quickly propagate global policy updates. Governance standards can be maintained with consistency, even at enterprise levels of scale and complexity. Access management can be localized and delegated.

As a result, data access is democratized among business users, minimizing the risk of bottlenecks or overload from stretched IT teams. It also supports remote working and the associated reduced visibility of systems and devices.

Adapting access to your organization

To bring about this unification, start by evaluating data access infrastructure. If this is the responsibility of IT, what happens when the data arrives from unstructured channels, or even shadow IT sources? How is this gathered, when is the governance team involved, and how are the inter-departmental silos managed?

Achieving the necessary alignment and integration starts by identifying data stewards. They can be appointed based on their understanding and experience of uploading data sets and administering flows, controls and standards. Responsibilities are mainly for the third of the three P's of data management: procedures. Data stewards are required to be aware of rules for aggregation and identifying data entities, documenting the relevant evidence. They should also be tasked with identifying any redundancy and optimizing data quality management.

By operating across departments, they bring the flexibility and accessibility required for onboarding and permissioning. Responsibility can be delegated to the stewards and administrators, who can easily and consistently replicate the necessary governance standards.

5. CENTRALIZE AUDITS FOR GREATER VISIBILITY

Audit logs are as varied as the systems used for managing and reporting on data. This usually means limited understanding and scalability when it comes to policy enforcement and access. With no single audit standard across the different systems, organizations often struggle to establish transparency from both ends of the data lifecycle.

However, achieving this visibility is crucial for managing dark data, particularly among larger enterprises within highly regulated industries. Whether that's unstructured data from users, inaccessible data stored in legacy systems, or simply data coming from new channels and sources. Without an aligned view, the data remains siloed and the schema is inconsistent. Costs can also creep up through cloud-based storage with pay-as-you-go models. Inconsistent classification and expiry labels also pose risks around data hygiene.

Solving the "dark data" challenge

Given that 80%–90% of data collected is unstructured, solving dark data is business-critical. Not just for uncovering untapped business intelligence, but also for identifying possible violations, vulnerabilities or malicious activities. Visibility and control of data access starts by auditing:

1. **The location of the data**
2. **Who in the organization is currently accessing the data**
3. **Who in the organization should be able to access the data**

Establish these three elements, and you gain a central foundation for understanding transformations and links between data. Data lineage can be mapped to track changes and permission statuses. Standardized metadata can be added, boosting future searchability and allowing partial and flexible searches across data catalogs. Dark data becomes discovered data – findable, auditable, and enforced with the relevant data retention policies. For example, by automatically classifying incoming data as PII and restricting access.

Centralizing also helps mitigate the challenges when governance teams set policies from the top down, without knowing if policies are being enforced or if data is protected. It also becomes possible to trace who has had access and for what purpose – across the entire data access lifecycle. Over time, as your organization and systems gain understanding of usage, it becomes easier to design and manage more effective policies.

6. ARCHITECT A FLEXIBLE FUTURE

Integration underpins every aspect of building the data platform. From a technological perspective and integrating AI and automation, to a behavioral perspective and integrating departments and adopting self-service.

With so many of these moving parts, start by looking at what already exists within your organization. For example, the existing attributes from catalogs or identity management systems.

Creating order from distributed disorder

Achieving this reality requires the flexibility that comes from API-driven architecture. After all, APIs are already being used by many platforms to share data, or via applications built on top of them. It's a natural progression to use APIs and related open source tooling for data access.

APIs offer key architectural components for developments such as open banking and the EU's PSD2 directive, where consumers can mix and match products to create a personalized form of banking. The consistency of APIs makes them integral to securing the associated verifications, payments, and data flows between institutions and third parties.

With a standard interface, APIs also offer a way to unify control of complex back-end systems, helping to increase speed, agility, and observability. API components also offer logging for programmatic tracking of the data lifecycle.

ANSWERING THE “WHAT COMES NEXT?” QUESTION

This paper is designed to start discussions and explore your organization’s level of maturity around secure data access control.

Use the questions below to help your key stakeholders get a sense of what’s needed to evolve further:

- **What** types of access control are we using – RBAC or ABAC – and are these enabling or preventing timely access requests?
- **Is there** an impact on data granularity when we try to scale our data governance processes?
- **What’s** the current level of data catalog usage, and how up to date is our metadata?
- **Do** we have a form of data stewardship, and does it meet the needs of non-technical users?
- **Do** dark data and shadow IT systems pose a risk to current data governance strategies?
- **How** much are APIs supporting our data fabric flexibility?

PARTNERING FOR PROGRESS

In the past, data governance may have often been primarily regarded as a way to ensure compliance and meet regulatory requirements. Introducing analytics redefines the concept of data governance from reactive to proactive. Where insights are generated for users to access and act on.

The stages highlighted above will help shift the focus of data access control to the point when the data is being requested, rather than when the data is created or stored. This will require consideration of the trade-off between increased resources or increased storage.

After all, it's more systems-intensive to provide centralized data that's been filtered and made compliant.

Other factors are around risk, and highly regulated industries such as finance, healthcare, or education. Given that some of the requirements above also involve AI and automation, organizations have to choose a partner able to demonstrate scalability, explainability, and transparency throughout the data lifecycle.

Whatever your current level of maturity, your data platform journey will involve many steps. Want to explore how to make your next step the right step? Talk to us about how to design a platform for scalable and secure data access.

OUR MISSION

TO ENABLE THE SECURED, EFFICIENT, AND COMPLIANT USE OF DATA.

Velotix

WWW.VELOTIX.AI

PARTNERING FOR PROGRESS

In the past, data governance may have often been primarily regarded as a way to ensure compliance and meet regulatory requirements. Introducing analytics redefines the concept of data governance from reactive to proactive. Where insights are generated for users to access and act on.

The stages highlighted above will help shift the focus of data access control to the point when the data is being requested, rather than when the data is created or stored. This will require consideration of the trade-off between increased resources or increased storage.

After all, it's more systems-intensive to provide centralized data that's been filtered and made compliant.

Other factors are around risk, and highly regulated industries such as finance, healthcare, or education. Given that some of the requirements above also involve AI and automation, organizations have to choose a partner able to demonstrate scalability, explainability, and transparency throughout the data lifecycle.

Whatever your current level of maturity, your data platform journey will involve many steps. Want to explore how to make your next step the right step? Talk to us about how to design a platform for scalable and secure data access.

OUR MISSION

TO ENABLE THE SECURED, EFFICIENT, AND COMPLIANT USE OF DATA.

Velotix

WWW.VELOTIX.AI