IAM Can’t Solve Today’s Data Security Problems

Mark Settle and Manav Mital

Conventional IAM tools lack the scope, granularity, visibility, and responsiveness to safeguard the practical usage of data resources found in every modern corporation.

IAM is the Cornerstone of Corporate Security

Identity and Access Management (IAM) has its roots in the use of passwords to gain access to a company’s corporate network. When all of a company’s IT resources were housed in its proprietary data center and all job-related activities were performed on company-owned devices, one-time network authentication via passwords was considered to be an effective means of protecting a company’s IT assets. Zero trust principles employed at that time were primarily focused on establishing trust between networks or user directories, not between end users and individual IT resources.

This basic IAM paradigm was adapted over time to manage access to cloud-based services and systems. Single Sign-On (SSO) procedures pioneered by **Microsoft** and **Okta** extended conventional IAM practices to SaaS applications. CIEM (Cloud Infrastructure Entitlement Management) vendors offered highly granular solutions for accessing cloud computing and storage resources, also employing classical IAM principles. These developments fostered the belief that ‘identity is the new perimeter’, and IAM became a cornerstone tool in every company’s security architecture.

Conventional IAM practices blur the distinction between identity authentication (AuthN) and resource authorization (AuthZ) from an end user's perspective. Although security professionals appreciate this distinction, end users typically experience authentication and authorization as a single event. Successful authentication automatically grants them a specific set of resource usage privileges, obscuring the underlying differences between these two very important and very complementary safeguard mechanisms.

Conventional IAM practices do not provide adequate protection for the diverse data resources found in a modern corporation. End user authentication is the first line of defense in any data security strategy. Many conventional IAM tools possess sophisticated procedures for verifying end user identity and are capable of enforcing stringent AuthN safeguards. However, these tools fail to provide adequate AuthZ controls on data usage for a variety of reasons that are unique to data environments.

> IAM is seen as the key to security, and in practice, it often blurs the line between user authentication (AuthN) and resource authorization (AuthZ). Conventional IAM practices lack the AuthZ controls that modern corporations need for data security.

Data Access Poses Unique Challenges

Access Controls Struggle to Cope With The Proliferation And Continuous Metamorphosis of Data Resources

Conventional IAM tools were built to control the usage of IT resources that are relatively finite, static, and long-lived, such as a collection of SaaS applications or network connections. Data is different. Data resources expand exponentially, are frequently ephemeral, and rarely static. Data can be endlessly transformed into derivative products and easily shared across teams, departments and companies, making it difficult – if not impossible – to establish and enforce consistent AuthZ controls on data usage.

Security organizations have attempted to cope with this complexity by employing one or more of the following strategies.

- *Cataloging tools* such **Alation**, **Collibra** and **Informatica** can be used to track the lineage of newly created or derivative data assets. Lineage relationships with accompanying information regarding the nature and sensitivity of the data contained within an individual asset can, in turn, be used to define access controls that are appropriate for that asset.
- *Data Security Posture Management (DSPM) tools* such as **BigID**, **Cyera** and **Symmetry Systems** conduct continuous surveillance on the usage of data assets and identify gaps in the enforcement of existing security policies.
- *Data Detection and Response (DDR) tools* such as **Cyberhaven**, **Dig Security** and **Laminar Security** complement the capabilities of DSPM products by providing a means of flagging potential instances of data misuse or misappropriation and automating the response to such events.
- *Walled gardens* are an infrastructure-based approach to safeguarding the use of data resources. Access to assets in a walled garden is controlled through stringent network AuthZ controls. Such assets may even be segregated into a dedicated virtual or physical hosting environment. Users who gain access to walled gardens are at liberty to use its resources in whatever way they please. However, they're restricted from exporting or sharing such resources outside the confines of the garden perimeter. **Databricks**, for instance, aims to provide such a walled garden through its Unity catalog.

Access Controls Are Not Sufficient to Restrict Many Forms of Data Misuse

'Access' has become a catch-all term to describe the ways in which the use of IT resources should be controlled. Unfortunately, it's a very coarse-grained concept that has limited applicability to data resources.

Effective controls on data usage are built upon a combination of access permissions, action privileges and granular entitlements. *Access permissions* grant entry or access to a particular resource such as a data store, database or file. *Action privileges* describe the actions that a user can perform after gaining access. CRUD controls (Create/Read/Update/Delete) are classic action privileges that can be exercised in whole or in part by users who have gained access to a specific data resource. And finally, *granular entitlements* may further restrict the records or fields that an individual user may view or manipulate. Grouping permissions, privileges and entitlements under the general term ‘access’ can be confusing, misleading and potentially dangerous. Vendors would be well advised to underscore these distinctions in describing the capabilities of their products. Buyers of such products would be well advised to seek clarity about what a specific product can and cannot do.

Access Controls

Access Permissions

Access permissions determine who is allowed entry or access to a specific resource, like a database or file.

Action Privileges

Action privileges define the actions a user can perform once they gain access, such as creating, reading, updating, or deleting data.

Granular Entitlements

Granular entitlements further restrict what specific records or fields an individual user can view or manipulate.

> Conventional “access control” is not enough for data security, and needs to be broadened to include access permissions, action privileges (e.g., CRUD), and granular entitlements. Vendors and buyers should differentiate these terms when describing and evaluating products.

Data misuse extends far beyond the consequences of a data breach or the inadvertent exposure of sensitive information. Many forms of customer, supplier, and partner information are acquired under specific consent or non-disclosure agreements that need to be enforced during all subsequent phases of data use. In many cases, these agreements not only place restrictions on the sharing of information with third parties but also on the *business purpose* or *context* in which data can be used within the acquiring organization.

Contextual controls are particularly difficult to define because data is frequently used for purposes its creators or collectors never intended or imagined. This is another important distinction between data and other types of IT resources. Business applications are purposefully designed to support specific business processes. Cloud computing instances are designed to support specific types of coding, testing and deployment activities. Data is different. Data is frequently used in ways that were never anticipated or envisioned by its creators, vastly complicating the definition and application of appropriate AuthZ controls.

It's interesting to note that 'least privilege' principles commonly employed in establishing AuthZ controls for infrastructure and application resources are much less frequently discussed with regard to data security. This is an implicit recognition of the complexity of layered permission/privilege/entitlement controls and the difficulty of applying least privilege principles to such controls across all data usage scenarios.

In summary, the use of conventional IAM tools to manage data AuthZ controls is somewhat similar to using a blunt knife to perform open heart surgery. The user group structures, service accounts and privileged account mechanisms used by conventional IAM tools to assign permissions, privileges and entitlements lack the specificity and business awareness required to be effective in a data environment.

The Semi-Infinite Uses of Data!

Corporate data is routinely used in ways its creators or collectors never intended or envisioned. Customer data may provide the most graphic illustration of this phenomenon. Every company maintains some type of master customer database or data platform. Although the sales and marketing uses of such information are obvious, customer data can also be used for a variety of other purposes such as the following:

- **Product development:** Analysis of the sales or use of commercial products and services across a company's existing customer base can guide investments in future products.
- **Inventory management:** Seasonal sales patterns and the buying behaviors of key segments of the existing customer database can be used to minimize the working capital tied up in merchandise sitting on the shelves of company warehouses and distribution centers.
- **Retail store planning:** The demographics of an existing customer base can be combined with publicly available demographic information summarized at a zip code level and historical eCommerce sales information to target the location of new retail outlets.
- **Product placement:** A/B testing can be performed on product placement in retail stores with similar customer demographics to determine placement patterns that optimize store revenue or expand the sale of higher margin products.
- **Product delivery commitments:** Dynamic information concerning product location and customer location can be used to forecast delivery dates or to ensure that delivery SLA commitments to premium customers are being consistently achieved.
- **Bias reporting:** The U.S. government may require certain businesses, primarily within the financial services industry, to demonstrate that they are not engaging in discriminatory practices in sales of their products or services.

- **Acquisition decisions:** Companies seeking to expand through acquisition routinely evaluate the overlap between their existing customer base and the customer base of a prospective target company to estimate potential gains in revenue and profitability that might be achieved through acquisition.
- **And the list goes on!**

Many Data Access Channels Don't Support Identity Persistence

Conventional IAM tools serve as gatekeepers between end users and IT resources. Users maintain identity accounts within IAM tools and personal, named accounts within specific resources, such as a SaaS application or videoconferencing service. Once they are successfully authenticated by an IAM tool, they are passed to their personal account in the targeted resource. Their identity information persists within the targeted resource and can be used to control the ways in which they exercise or extend their permissions/ privileges/entitlements.

That's not always the case in the use of data resources. The identities of individual end users accessing data through business intelligence applications, data analytic notebooks or ML/AI modeling platforms, or via system-to-system data contracts, APIs or microservices may not persist in the targeted data resource. It's obviously impossible to establish identity-based controls on data usage under these circumstances.

A unique challenge with data security is that access is often through applications (visualization tools, notebooks, APIs, etc.) that don't persist the user identity when connecting to the data store.

It’s Not Just A Technology Problem – Data Usage Policies Struggle To Keep Up With Data Sprawl, User Sprawl and Scenario Sprawl

It’s unfair to attribute the limitations of AuthZ data controls solely to the limitations of conventional IAM tools. Quite frankly, it’s difficult for data usage policies to keep pace with the myriad of data resources, end users and business scenarios that are continuously evolving within a modern corporation.

Controls may be established and administered at multiple levels of an enterprise by individual work groups, functional departments and at an enterprise-wide level. Control accountability is difficult to enforce because the *authorship of a data resource is not necessarily equivalent to ownership*. Individuals who had no role in creating or constructing a specific data resource may be assigned responsibility for defining and enforcing its usage policies solely on the basis of their organizational roles or responsibilities.

Timely policy creation and administration is also a practical concern. Authorization restrictions need to be imposed at ‘the speed of the business’. Action privileges may need to be modified on the fly as a data analyst researches the root cause of a critical business issue. Granular entitlements may need to be exercised at query time, dynamically obscuring specific fields or records in response to individual queries or table joins. Most, if not all, authorization systems struggle to enforce effective policies without introducing undue and highly contentious friction in everyday business operations. IAM tools are simply one of many security safeguards that are continually trying to balance tradeoffs between protection and productivity.

AuthZ data control limitations cannot be solely blamed on IAM tools. Adapting data usage policies to the evolving landscape of data resources, users, and business scenarios within a modern corporation is challenging. Controls cross enterprise levels and often lack clear accountability. Timely policy creation is essential, and adjustments need to be made to keep up with business requirements. Enforcing effective policies without disrupting daily operations is a common challenge across authorization systems, including IAM tools, which aim to balance security and productivity.

The Proliferation Of Data Platforms And Analytical Systems Further Complicates The Consistent Enforcement Of Data Usage Controls

Many large data platforms such as **Snowflake**, **Databricks**, and **Google BigQuery** have intrinsic security architectures that employ uniquely defined terminology, user role restrictions, usage protocols, and security services to manage data usage. Any attempt to manage permissions/privileges/entitlements on a consistent basis across an enterprise must necessarily be translated into the security 'languages' of these individual platforms and other data-intensive systems. In most cases, it's difficult – if not impossible – to map IAM user group structures and service/privileged account mechanisms into the security architectures of data-intensive platforms and systems. More often than not, the coarse-grained AuthZ controls employed by IAM tools hinder a company's ability to leverage the intrinsic controls of sophisticated data management and analysis platforms.

NextGen Solutions

The business threats posed by data misuse can never be completely eliminated but they can be minimized by a variety of new technologies, some of which have been referenced already. The following discussion highlights emerging capabilities and concepts that are particularly promising.

AI-Assisted Data Destruction

The threat of misuse increases in direct proportion to the size of a corporation's collective data holdings. One of the simplest means of minimizing misuse is to simply discard or sequester data that is likely to have marginal value in the future. Data purists may consider this to be heresy and claim that it's impossible to anticipate the future value of data. Therefore *all* data needs to be retained indefinitely. However, as a practical matter, data destruction is occurring today, largely as a cost control measure. In many instances, destruction occurs randomly based upon the personal intuition of individual data analysts or engineers and not on the basis of any formal business principles or evaluation procedures.

The amount of 'dark data' in major corporations is staggering. A 2019 survey conducted by **Splunk** revealed that 55% of the data retained by business enterprises is unused or, in many cases, unknown. A more recent report published by **Zippia Research** similarly indicates that only 40% of corporate data is actually used**.** Vast amounts of data go unused precisely because corporate data holdings are expanding so rapidly. Data professionals recently surveyed by **G2** indicated that the collective data holdings of their companies were growing by more than 60% on a monthly basis.

This represents a significant opportunity for AI technology. Continuous monitoring of asset usage throughout a company can provide valuable insights into the likely value of assets in the future. AI algorithms can be used to discard data on a probabilistic basis based on these insights, allowing individual companies to specify their risk tolerance for discarding data that might prove useful in the future. Alternatively, some assets might be sequestered in low-cost storage environments for an AI-determined period of time prior to destruction, as a hedge against future usage scenarios that are difficult to anticipate at the present time.

Active Metadata

The metadata used to describe data assets has traditionally consisted of semi-static information regarding asset age, ownership, location, schemas, field definitions, security tags, etc. More recently, greater emphasis has been placed on defining the lineage of data assets as well.

Traditional forms of metadata are rapidly being supplemented with operational information regarding how assets have been used in the past and how they should be used in the future. More specifically, dynamic information is being captured concerning authorized hosting environments, authorized users, recent usage patterns, field and record level encryption requirements, value range fluctuations, run time security protocols, etc. The expansion of traditional metadata frameworks to include operational information and quality metrics is popularly referred to as 'active' metadata.

Data assets must ultimately become more *self-describing* in terms of their characteristics, quality and usage. Past usage patterns provide a powerful means of establishing future usage controls but they must be readily available and not buried in system logs to be of practical use. Widespread adoption of active metadata concepts and products can significantly improve the sophistication of data AuthZ control mechanisms in the future.

Prukalpa Shankar, the Co-Founder of Atlan, has been a particularly powerful evangelist of active metadata benefits. Zhamak Deghani, the founder of the data mesh movement, espouses a similar concept in the form of policy 'sidecars' that are an integral component of individual data products.

AI-Assisted Dynamic Policy Modification

Conventional AuthZ controls are rule-based. *If* the use of a specific asset is being requested by a particular individual under a specific set of circumstances, *then* approve/deny the request. Day-to-day business activities don't always operate in this neat if/then fashion. Business purpose or context may have an overriding influence on allocating permissions/privileges/entitlements for individual tasks or activities. For example, an HR analyst updating compensation tables during the annual merit review program does not need to view employee SSNs. If the same analyst is constructing a compensation table that ADP can use to administer bi-weekly payroll payments, then he/she needs to be entitled to view and export SSN data fields for tax withholding reporting purposes.

AI can potentially play a critical role in adapting AuthZ policies on a probabilistic basis to adjudicate usage requests under a wide variety of usage scenarios. Control policies for a specific asset may be inherited in whole or in part from predecessor products that were employed in its creation. These policies may be subsequently modified based on the roles and responsibilities of individual users, the virtual environments in which they are performing their tasks, the nature of those tasks, etc. Dynamic policy modification is in an embryonic stage of development precisely because it is so difficult to perform correctly and consistently. In principle, AI algorithms can conquer this complexity and make risk-based recommendations regarding data usage on the fly, even during a single work session. AI-based recommendations can be accepted or overruled based on the risk tolerance of specific teams, departments, or enterprises.

As noted earlier, AuthZ controls are based upon conscious tradeoffs between employee productivity and data protection. In theory, AI algorithms can quantify these tradeoffs on a much more consistent basis than human administrators. AI routines are particularly adept at risk optimization, provided that policy administrators can establish clear guidelines for risk tolerance under different circumstances.

Externalized Authorization Management (EAM)

EAM is the only sane way of coping with the proliferation of data assets, storage platforms, analytic applications, and business usage scenarios. EAM tools abstract policy definition to an external control plane that can be integrated with the idiosyncratic security frameworks of individual data platforms and systems. Data security teams 'write policy once' on this control plane and then enforce it everywhere via these integrations.

Cyral and **Immuta** have pioneering offerings in this space. Their ultimate success or failure will ultimately be determined by the breadth of integrations they are able to achieve and their ability to incorporate dynamic policy modification capabilities in their offerings. EAM tools hold great promise but they need to overcome these challenges to become truly effective and achieve widespread adoption.

Authors

Mark Settle is a seven-time CIO, three-time CIO 100 award winner and two-time book author. His most recent book is *Truth from the Valley, A Practical Primer on IT Management for the Next Decade*.

Manav Mital is the CEO and Founder of **Cyral**.

Cyral

Cyral provides controls for privacy, compliance, governance and protection thereby reducing risk, complexity, and cost for managing structured data. The Cyral platform discovers data, unifies access controls for users and applications, enables fine-grained authorization policies and provides complete monitoring and reporting. This comprehensive coverage enables risk-based governance, limits the blast radius of data-related incidents and reduces overhead and costs. Cyral's technology allows customers to implement data security controls using their existing, centralized entitlements, thereby simplifying administration and automating remediation. Customers use Cyral to accomplish least privilege, data minimization, spillage prevention and Zero Trust.

For more information, visit **www.cyral.com** or follow **@CyralInc** on X.