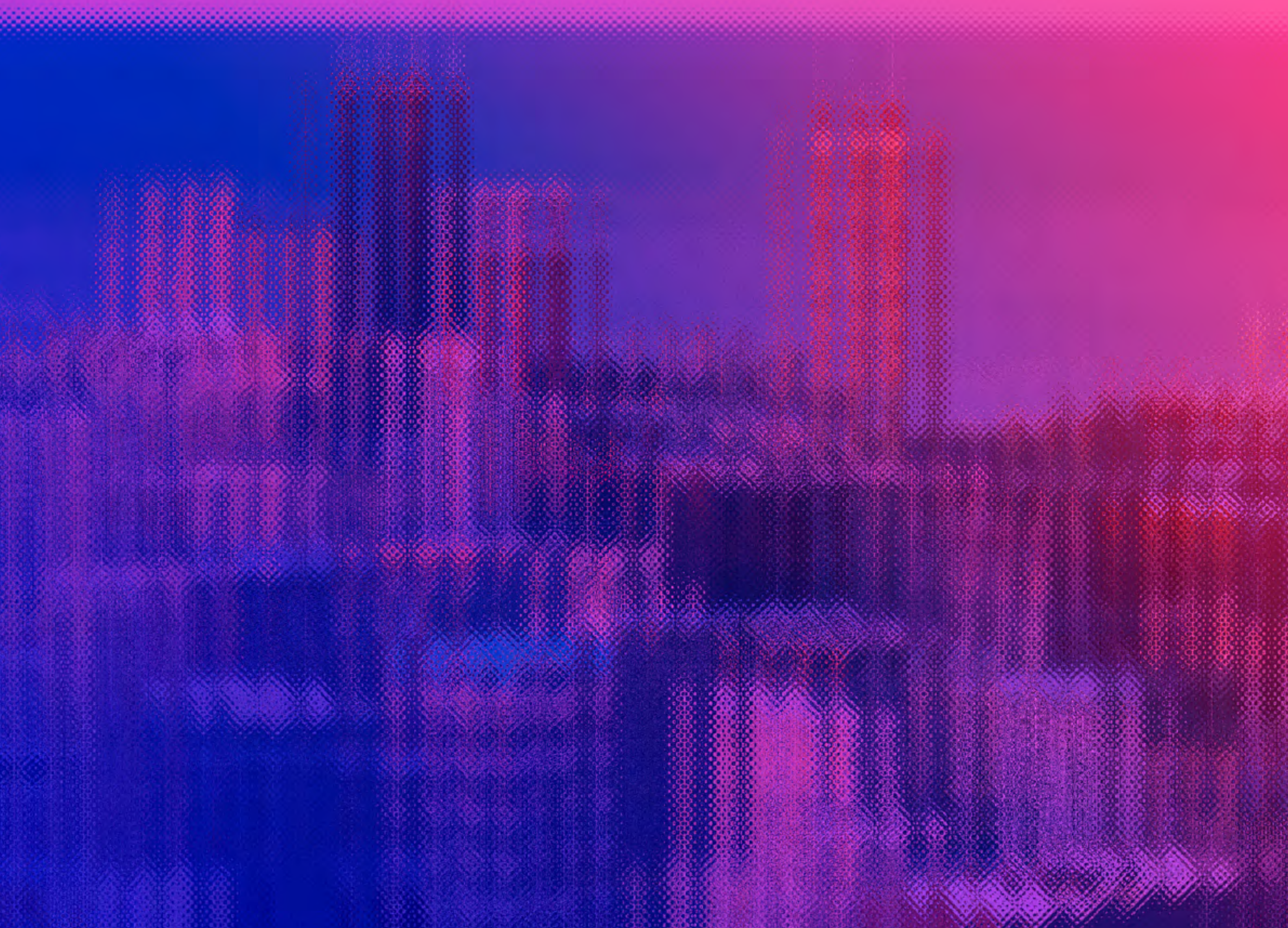
PRIVITAR



The Data Protection Framework

Create a Data-driven Bank with Privacy and Accountability Built-in

As financial firms seek to be data-driven and leverage all forms of data for competitive advantage, there comes a realization that the crux of that effort is to analyse personal and other types of sensitive data. At the same time, there is growing regulatory and reputational risk that calls for the protection of that data. The challenge then becomes how to rapidly protect sensitive data without losing its analytical value.

The Practical Nature of the Data Protection Framework

Data protection is not a choice, but how your organization affects a solution is. For banking and finance organizations, a mature, flexible, and scalable data protection framework is needed and it can only be derived through an enterprise-wide cross-functional perspective.

Although that sounds like a big ask, this paper walks you through a story of how such a framework naturally fits within the efforts of becoming a data-driven bank.

Enjoy!

Contents

This paper describes how to address that challenge by painting a picture of business and engineering teams working together to become a data-driven bank. In doing so, we walk through a number of areas.

- A base model for a data-driven bank.
- The foundation of a data protection framework based on subject matter expertise and accountability.
- The risks and benefits of leveraging sensitive data
- Practical steps to bringing that framework to fruition starts.

Many of the deepest insights are derived from sensitive data, but responsible enterprises must ensure that such data is safe for wide usage.

The Data-Driven Business Conundrum

Calls for digital transformation to become a data-driven business resonate loud and clear throughout the financial services industry. These terms can be nebulous, but for those whose job it is to develop novel ways to modernize and automate processes and decision making, the steps to get there are clear, at least initially.

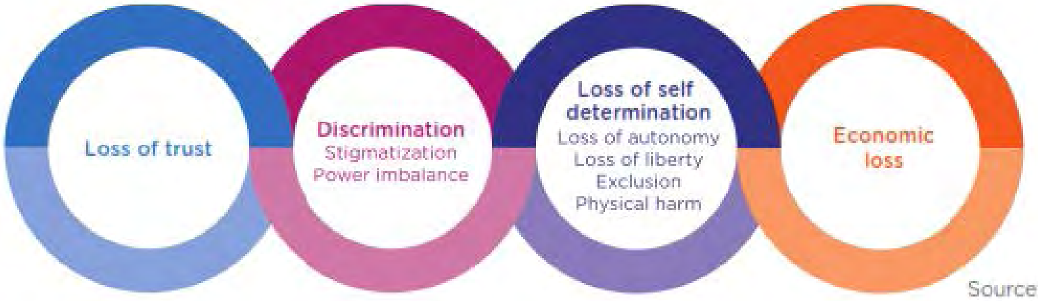
As they peel back the onion, however, the challenges of using data to drive better business outcomes become more complex as technological, operational, organizational, social, and regulatory layers get exposed to reveal the following: the transition to a data-driven business is a sophisticated balancing act.

Negative Consequences and Harm

On the one side of the scale, the deliberate or unintended misuse of data has negative consequences ranging from the dramatic to the mundane: class action litigation, law enforcement actions, executive firings, regulatory penalties, and/or higher cyber insurance premiums.

While the effects on large enterprises are sometimes newsworthy, it shouldn't be lost on us that data missteps can cause real privacy harms and social disadvantage. For a deeper understanding of what could be, consider the diagram below that lists four types of harm: loss of trust, discrimination, loss of self determination, and economic loss.

As diagrammed, these categories overlap, and any specific event could lead to a snowball effect.[1]



Source: Adapted from NIST[13]

Figure 1: Categories of privacy harm.

- Loss of trust is often an outcome of one or more of the harms listed here. While loss of trust is commonly seen as a reputational risk for an organization or an industry at large, it is a personal harm. As a result of lax data protections or an organization's outright inappropriate behavior, the violation of personal space, even virtually, can affect emotions of anger and helplessness.

- Discrimination most often comes to mind as an event in which an individual is offered sub-optimal information, services, pricing, or shut out of an opportunity altogether based on their personal profile. Targeting, is another type of discrimination. It is a form of manipulation, where the data about an individual is used to influence their choices that could induce addictive on-line behavior.[2]

- Loss of self-determination is a broad category and covers a range of possible harms, such as a loss of control if data is collected without an individual's consent. It could also come in the form of emotional distress as a result of information shared over social media.

- Economic loss is commonly caused by fraud resulting from a data breach. Identity theft, for example, is when criminals use breached data to impersonate an individual and make purchases or commit some other fraudulent activity. Spear phishing is another example in which criminals use a subset of personal information to persuade them to release more information.

Positive Outcomes and Data Utility

On the positive side of the scale, the innovative but proper use of sensitive data is hugely valuable. A number of advanced data analytic use cases express the utility of data as the ability to generate revenue while mitigating risk. For example, while machine learning on customer data can reveal behavioral patterns that lead to successful upsells of products and services, that same deep customer knowledge can help to disrupt fraud attempts soon after (or even before) it is identified. The data reflects this. Some estimates show that personalization provides a 10-30% lift in revenue[3] while machine learning captures 50% more fraud attempts.[4]

Financial institutions capture a lot of their customers' data simply through the course of doing business. For example HSBC, one of the largest banks in the world, reported that 87% of their retail banking transactions are digital.[5] By itself, that data is quite revealing, with cash transactions patterns, in-store purchases, and reward point redemptions providing a clear picture of a person's buying preferences and proclivities.

Hyper-personalization is a banking industry imperative that takes the above to the next level, not only to understand a customer's transaction history but also how they think. Hyper-personalization leverages the nuanced details of a customer's real-time omnichannel experiences, including the content of call center chats, meetings with financial advisors, online digital behaviors, biometric readings, and the time and

$392million:
How much it costs to recover from a mega-data breach[11]

location of each interaction.

Ironically, in a world that is very sensitive to data privacy concerns[6], hyper-personalization is a consumer driven demand. Over half of consumers expect banks to anticipate their needs and make relevant product suggestions[7] but at the same time, they also expect to be protected.[8]

Machine learning, AI, and other big data technologies significantly magnify the negative and positive impacts on data, often simultaneously. Addressing that paradox is a challenge that can't be solved through a binary response of locking up data for protection or releasing data for utility. Banking and financial institutions require both and that balance is best brought about through a collective effort.

The Enterprise-Wide Response

A data-driven business is heavy on data analytics, a discipline not limited to a centralized group of data scientists and technologists. In large enterprises, data analytic functions are performed across a patchwork of organizational domains, each producing, consuming, analyzing, and acting on data in ways that reflect their priorities and expertise. However, operating through a patchwork of domains is challenging because societal, regulatory, and market trends do not align with internal organizational structures. As such, facing the world at large with an efficient and unified front is difficult at best.

In the omni-channel multimodal world of hyper-personalized banking, every touch point gives customers, partners, regulators, and even bad actors a much broader plane of interactions than can be acted upon by any one domain because each is the subject matter expert on only one aspect of the larger picture. A cross-functional, enterprise-wide digital response better completes the picture.

A Model of a Data-Driven Bank

The vision for a data-driven bank is both technical and organizational. It is also a federated one meant to let each domain effectively and proactively respond to the new opportunities and risks of a multimodal world, regardless of how unstructured and unplanned it may be.

From a technological point of view, many organizations evolved centralized big data platform architectures like data warehouses and data lakes. At the time, they were adopting technology innovations that solved the functional and economical challenges of storing massive volumes and varieties of data.

The Fine Line Between Helpful and Creepy
Data is data and big data is just more data, but how you use it matters. With data you can be helpful or you can overstep your bounds into someone's virtual personal space. Listen to Uber's Chief Privacy Officer, Ruby Zeffo describes the fine line between being helpful and being creepy.

Digital Transformation: The Driver of Post-Pandemic Economic Recovery (@5:25)

From an organizational point of view, this centralized paradigm helped to break down data silos and enable cross-functional collaboration between lines of business to streamline processes and create new revenue opportunities.

The success of the latter revealed the bottlenecks of the former. While the lines of business worked to respond to organic, event driven market demands, the linear and top-down approach of centrally curated data architectures exhibited a dragging effect on new business. In a world that rewards anticipation, recommendation, and on-demand delivery, the excessive time needed to centrally collect, transform, and load data into generic models makes the new product offering or fraud prevention opportunity moot.

The vision of a data-driven bank is transformative, not so much by adopting novel data technologies and patterns but by giving each domain more autonomy while demanding more accountability. This model better addresses dynamic use cases that balance the need for data protection and data utility. Below describes how your enterprise data engineering colleagues would enable that.

Data Engineering Enables Business

The modern role of enterprise data engineering teams is to enable business. They are responsible to deliver the data capabilities, tools, and platforms that enable teams across the organization to effectively adopt the latest trusted technologies and to do so in an agile, governed, protected, and modern way.

Having dealt with the challenges and concerns brought by multiple lines of business, they have developed a broad perspective of the organization over time and appreciate the challenges inherent to enterprise-wide organizational dynamics. They also understand that the requirements for data protection and data utility are not diametrically opposed.

To solve this puzzle, many data platform engineering teams are working towards a data mesh[9] architecture, which figuratively aligns with the patchwork of organizational domains

A Winning Offense and Defense

In sports, a great offense and a strong defense usually makes for a winning team. The skillsets of each are quite different and how a team adapts their respective talents while crossing half court or midfield in either direction is critical to the team's success. An offensive business model is proactive, it ekes out profitability in commoditized markets and generates additional revenue streams through new products and services that are timely and relevant to individual customers. A defensive business model, on the other hand, prioritizes protection against misappropriation of those products and services and the security of the full range of data assets. CTOs, CDOs, and CIOs and their respective data engineering teams work to support both types of business models.

CTA: Hear how ABN Amro's Head of Data Engineering describes this described above. The concept behind a data mesh is that data ownership is federated across each domain and each provides data to the enterprise as distinct products.[10] In terms of this data pattern's value to the bank, the autonomy provided to each domain, if implemented well, enables faster time to value due to agility of analysis and accountability for high quality data.

Subject Matter Experts Enforce Accountability

The vision of a data-driven bank is to give each domain and their technology teams opportunities to quickly respond to market events and customer behaviors. Effective action is best taken by the team with particular business context expertise. By decoupling data owner and consumer functions, the respective subject matter experts can be accountable for the content, quality, and use of the data.

Furthermore, decoupling helps to ensure agility because each team retains the autonomy needed to experiment and figure out how to deliver more and faster value with data.

But a third participant is needed in this scenario from a data protection perspective. One who can advise and keep score across all parties. Below we summarize the three.

Data Production Experts

The data owners within each domain are inherently knowledgeable about the data they produce because they work with it day in and day out. They understand the nuances of the data, discerning if an anomaly is a sign of something significant or simply a technical hiccup. That experience translates into developing the most efficient way to extract, transform, load, and deliver that data. Taking all of that into account, the data owners are therefore in the best position to be accountable for its content and quality.

Data Utility Experts

Those that consume data do it for a reason, be that building new products, optimizing services, or diminishing risks that reflect their deep understanding of their customer, the competition, and regulatory regimes. They are not limited to one data owner and could join a number of data sets with an eye on what combination is appropriate to the market, how to leverage it, and what is on the horizon. Therefore, data consumers are the subject matter experts in the utility of the data that they use.

Data Protection Experts

There are a number of roles that fall under the data protection category. Data Protection Officer, Chief Information Security Officer, Data Governance Officer, and in part the Chief Risk Officer, Chief Regulatory Office, Chief Compliance Officer, and this list goes on. They put together and rely on teams of individuals to apply their expertise as it relates to current and future best practices, regulatory patterns, and the threat horizon.

The vision of a data-driven bank necessitates that each advise the other, and it is that cross-functional enterprise-wide collaboration that makes a data protection framework.

The Data Protection Framework

The diagram below illustrates how a data protection framework would align with the

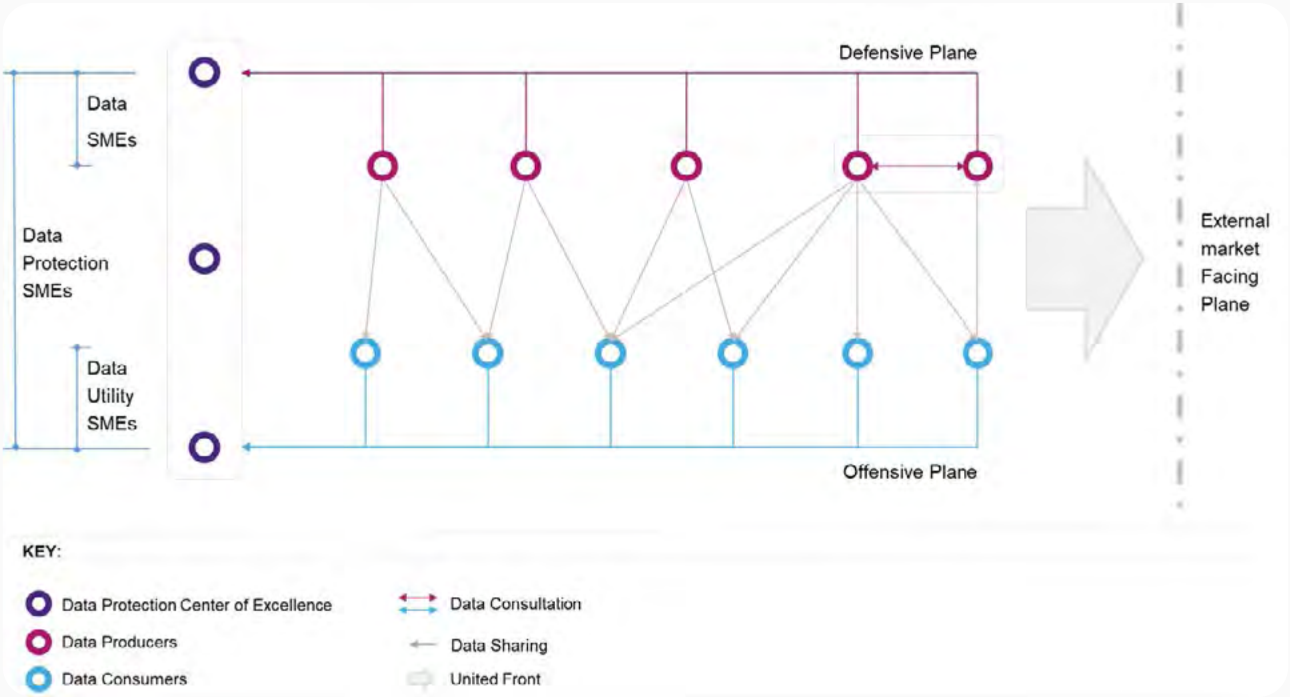


Figure 2: Concept diagram of a data protection framework in action

guiding principles of a data-driven bank. From left to right, the framework hinges on the collaboration between the subject matter experts described above: data protection, data production, and data utility. With that in mind, data consumers and data owners, in consultation with the data protection experts, are free to aggregate and share data as best fits their strategies.

In effect, an offensive plane of data consumers and a defensive plane of data owners, each in consultation with data protection experts, is created that provides a unified front to the external market. The result is that data privacy, governance, and accountability are directly built into data and product development lifecycles. The consultative element is critical to this framework and below we describe how that would work from an organizational point of view.

Data Protection Center of Excellence (DP-COE)

The data protection subject matter experts work to establish enterprise-wide data protection alignment and, as a group, are often referred to as a Competency Center, a Capabilities Center, or as a Data Protection Center of Excellence (DP-COE).

The DP-COE is an enterprise-wide service that provides leadership, best practices, research, support and training with regard to protecting the data of the institution as a whole. It should be seen as a coordinating function to effectuate enterprise-wide benefits as listed below:

- Collaborate with enterprise data engineering teams to reduce data provisioning time while ensuring the points below.
- Maximize the utility of datasets while ensuring they're properly safeguarded.
- Ensure that privacy is standardized across the organization's data operations.

- Galvanize data privacy as an enabler rather than a blocker.

In doing so, there are three main functions performed by the DP-COE.

1. Co-design the data protection framework that protects data throughout its lifecycle. This includes data security, privacy by design strategies, and the adoption of privacy-enhancing technologies and engineering methodologies.

2. Monitor and remain current on the threat horizon and global regulatory trends to ensure that the data protection framework meets all data protection requirements across regional jurisdictions.

3. Audit, report, and oversee operations to ensure that, in practice, the data protection framework is fully employed and appropriate risk mitigation techniques are in place.

The work of the DP-COE is critical to developing a mature data protection framework that aligns the internal data supply chain with customer demands and varied regulatory requirements.

Privacy-Enhancing Technologies and Engineering Methodologies

A safe provisioning framework is only as good as the underlying mechanics that let it function. Above, we described the organizational parameters, while below, we summarize some of the functional and technical factors that are the critical pieces of the puzzle.

There are a number of ways to protect data. You can lock it away, delete personal data, or obfuscate it. Many scenarios would consider all three but there are additional layers that need to be considered [For a quick primer, see Data Privacy 101: Guide to De-Identification]. For example, can what was initially thought of as anonymized data be used to target individuals when combined with other data sets? If an individual has been de-identified, do you then need to re-identify them in order to provide a service? The answer, in a hyper personalized world and in both scenarios is, yes.

Under the leadership of the DP-COE, there are a number of privacy-enhancing techniques and methods that can be applied. But since every situation is different as it applies to any combination of data owner(s)/data consumer(s) at hand, standardized workflows are needed for efficiency and automation but open enough to

Manage a Data Protection Framework

Protecting customers' privacy can seem like a monumental task, particularly as data collections have grown exponentially during 2020's pandemic-driven lockdowns. With a collaborative, risk-based approach to data privacy, organizations can appropriately leverage even sensitive data.

Watch on Demand: In:Confidence fireside chat with Michelle Dennedy, Privacy Leader and Co-author of the Privacy Engineer's Manifesto to better understand how to maximize your innovation with data privacy and safe analytics.

address the immediate revenue generation or risk mitigation opportunity.

For example, different domains require different levels of data utility. The right solution will be able to apply the appropriate privacy technique based on requirements and allow you to optimize data down to the column level. The data received from direct customer interactions may be acceptable for product recommendations, but would not be appropriate for credit risk analysis.

Understand Your Level of Maturity

More often than not, knowing where you are is the first step in moving forward.

It's important to understand that the models and frameworks described above would not be the result of a big bang one-and-done event. It would be an evolutionary approach that, probably, your organization is already in the midst of. In that case, it makes sense to first understand a few things before moving forward.

1. Where does your domain sit within the larger organization? Is it a line of business or an enterprise service provider?

Consumption Patterns and Supporting Architectures
For a framework to work, it needs to apply to a variety of data patterns and architecture form factors. Data privacy policies can be deployed to any state of data in your system. Here are a few examples:

- **Batch**: For scenarios that require data on a pre-planned basis like end of month financial statements or overnight analytic routines. Tailored data protection techniques can be automated to provide safe data sets that match privacy requirements of each data consumer use case.
- **Event Streaming**: To capture transaction events at the source, like a credit card swipe or trade execution, the data powner publishes data events as topics to which data consumers chose to subscribe. Data protection policies are applied on a topic by topic basis.
- **API**: This is a request/response mechanism built off the event streaming pattern. In this case, instead of getting individual events in real-time, the consumer can request the most recent 10-20 transactions on an account for example. Application developers create customized workflows that adhere to data protection policies.
- **Private, Public, Hybrid, and Multi Cloud**: Deploying data protection policies across a variety of data environments is possible. For example, a Global Exchange Administrator can publish protected datasets to multiple data exchanges, incl. Internal ones (e.g. UAT vs prod), and external ones (e.g. Snowflake, AWS).

2. When it comes to data, is your domain a data owner, consumer, or both? Are there collaboration opportunities with others?

3. With regard to data protection, what is your organization's level of maturity?

Points one and two can probably be answered on your own. Point three requires a bit more context. For that, below are the basic levels of a data protection maturity model.

Level 1 – Ad-hoc: There is no framework in place to adequately control privacy protection. Privacy and data security are considered in isolation.

Level 2 – Siloed: data protection is controlled at a domain level only and is not meant to be shared.

Level 3 – Coordinated: Data is protected based on usage, regardless of domain.

Level 4 – Centralized: Data protection, governance, and accountability are directly built into data and product development lifecycles.

Level 5 – Innovative: Data protection enables new business initiatives.

Most organizations find themselves at Level 1 or Level 2 but with aspirations to level up and extract the maximum utility from their data. The diagram above represents a pathway by which your organization could ascend Levels 3, 4, and 5 to simultaneously operate at the highest levels of business value and privacy capabilities.

For context, the data protection framework and the data-driven bank model described in this paper depicts a Level 5 maturity scenario.

The purpose of the maturity model is to help level set where you are in the journey. It is also meant to establish a common vocabulary and shared vision amongst your colleagues by which

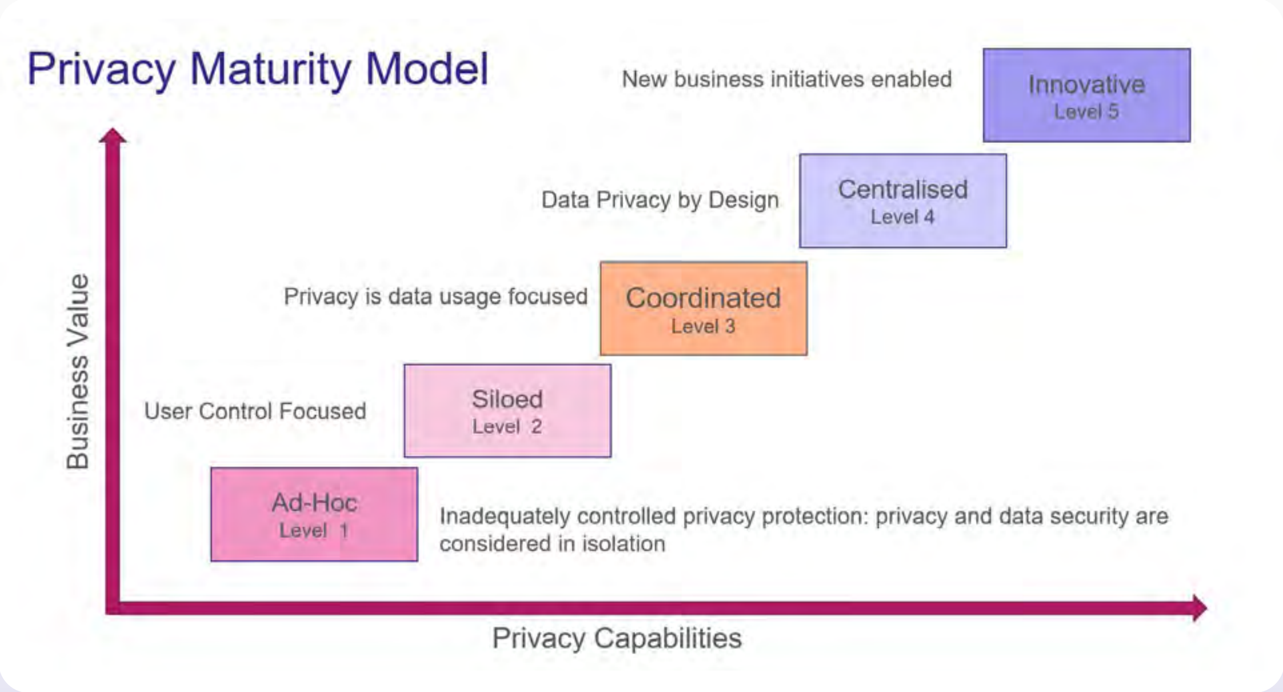


Figure 2: Concept diagram of a data protection framework in action

to elevate data provisioning in your organization and maximize adoption.

Find Success from Balance

Many of the deepest insights are derived from sensitive data, but responsible enterprises must ensure that such data is safe for wide usage. From that basis, the entire enterprise can find, understand, and collaborate across datasets to unlock next-level data innovations on the way to maturity Level 5.

The data protection framework described in these pages naturally fits with the efforts to become a data-driven bank. It is meant to simplify the processes by which to generate more utility from sensitive data and to safely democratize data science.

Business acceleration is a constant balance between risk mitigation and revenue generation. In a data-driven world, either end of the scale is weighed by how data is treated.

The institutions that succeed in balancing data availability, data protection, and data utility are those that evolve an operating model that is agile to business innovation, indifferent to location, adaptable to evolving regulations, and establishes accountability across all involved. Accelerating business with privacy and accountability built-in is only possible through a collective cross-functional and enterprise-wide effort.

About Privitar

Privitar provides technology and service solutions that help organizations ascend the data maturity curve in the pursuit of extracting maximum value from their data.

Privitar’s flexible Data Privacy Platform applies privacy protection to data at massive scale. By combining The Privitar Data Privacy PlatformTM and data privacy best practices into their analytics and data science operations, customers accelerate insights while building trust and mitigating regulatory risk.

Founded in 2014, Privitar is headquartered in London, with regional headquarters in Boston and locations throughout the US and Europe.

Contact Us

info@privitar.com
+44 203 282 7136
www.privitar.com

Endnotes

1 Privitar. “Understanding privacy harms for risk management” p 18.

2 For a deeper dive into the manipulative aspects of targeting, read the Scientific American article, Cambridge Analytica and Online Manipulation by By Marcello Ienca and Effy Vayena. (March 30, 2018)

3 The Financial Brand. Personalization Pays Off Big Time for Financial Marketers. May 2019

4 Paymts.com. Deep Dive: How AI and ML Improve Fraud Detection Rates And Reduce False Positives. September 2020

5 HSBC. Banking of the Future: Finance in the Digital Age. November 2019

6 Cite Privatar research

7 Deloitte. The Future of Retail Banking November 2020

8 TBD - cite this from somewhere.

9 Dehghani, Zhamak. How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh. May 2019

10 Moses, Barr. What is a Data Mesh — and How Not to Mesh it Up. July 2020

11 IBM Security. Cost of a Data Breach Report 2020. July 2020

PRIVITAR

PRIVITAR.COM