LightBeam.ai

Zero Trust Data Protection

A framework for organizations to unlock true potential of their business data while strengthening data trust protection with enhanced data security, privacy and governance posture.

WHITEPAPER

LightBeam.ai

Table of Contents

Executive Summary

Nearly every organization carries personal and/or sensitive data in one form or another. Customers, consumers, contractors, employees, and other data subjects share their data with an expectation that the receiving organization protects their data and treats it with the same respect as the organization treats them. However, locking up data within the organization is generally a non-starter, as businesses need data to function efficiently and effectively. So the question becomes how can the two seemingly competing requirements of data security and privacy vs. data utility be aligned into a single unified data governance framework that achieves the right balance of data protection and successful business outcomes?

This whitepaper explores the concept of a **zero trust data protection** approach to practical data governance in greater detail. It breaks down zero trust data protection as a function of data security controls and data privacy readiness. The author's objective is to highlight the various facets of data security and data privacy often going beyond the regulatory requirements, which are deemed necessary and important, but not always sufficient.

A world where sensitive data is secure and protected, and still retains its utility is a world where decisions can be made about data without violating consumers' data privacy and security rights.

Audience

This document is intended for enterprise security teams, data privacy/protection officers (DPOs) who are helping their organizations stay compliant with global regulations like the GDPR, Québec Law 25, CCPA and numerous other privacy regulations, or for compliance teams in charge of sectoral regulatory requirements such as PCI, GLBA, HIPAA, or IT and data teams helping their organizations to

leverage artificial intelligence (AI) to drive the next growth arc of their business. Finally, business leaders can use this whitepaper as a framework to bake-in data governance policies right from project inception through completion, or to help privacy and security impact assessments be more efficient, which can minimize last minute surprises and avoid expensive project costs and time overruns.

Introduction

Zero Trust Data Protection is a concept that assumes no entity, whether inside or outside an organization's network, should be trusted by default when it comes to accessing, using and sharing sensitive data. This approach challenges the traditional security model that assumes that everything inside an organization's network is safe, and where access is controlled broadly at a group or role level. With Zero Trust, the granularity of data protection moves to an entity, where the entity could be a system user, data subject or individual or device or application.

A well-architected Zero Trust Data Protection model encompasses the following concepts:

- **No Trust by Default**: Trust is never assumed. Whether it's based on the internal or external location of the data or the user's identity. Every user, device, or system is treated as potentially untrusted.
- **Continuous Privacy and Security Impact Assessments**: Instead of being periodic and manual, impact assessments are on autopilot, associated with every document update, planned/unplanned data/log collection, code check-ins, and support tickets amongst other actions that change stored data.
- **Identity Centric Data Inventory:** Instead of the legacy data element centric approach of discovery and cataloging, Zero trust promotes an “entity” based approach to data protection in which the entity or data subject and all of

their related personal information has been identified and recorded and can be monitored for change.

- **Spillage Safeguards**: Inappropriate data is automatically redacted and/or anonymized by default policies that protect gold source data and enable schema observability for rapid development and decision making.
- **Highlight External/3rd Party Data Sharing**: Any external access to company resources must be assessed, approved, and apparent. Records for 3rd Parties using personal data are well tabulated for quick reviews, with automated controls minimizing risk of sensitive data getting exposed in the event of a 3rd party breach.
- **Continuous Verification**: Continuous authentication and authorization are required, even for users and devices that are already inside the network, access is granted on a least-privileged basis.

Implementing Zero Trust Data Protection helps organizations enhance their security and privacy posture in a world where traditional network perimeters are becoming less relevant. It's particularly crucial as organizations embrace cloud services, remote work, mobile technologies, and AI everywhere. In these modern times, the traditional strong perimeter castle-and-moat approach to security becomes a less effective model, while Security and Privacy by Design models work to build in required protections.

Continuous monitoring and data leakage prevention add up to a Zero Trust environment

Why Zero Trust Data Protection Now

In the movie, “The Truman Show”, Truman Burbank goes about living his life in an almost perfect, if boring, setting. Arguably his life is secure. Living life as part of a reality TV show, watched by millions of people, is nothing if not secure in general. But privacy is another matter altogether. In 1998, the movie was quite ahead of its time in laying out the trade-offs between privacy and other comforts of life. Having come to the realization that his privacy has been traded off his entire life, Mr. Truman reacted negatively and realized that he didn't like the intrusion one bit. The irony is that the movie came at the start of an era where nearly the entire world traded their privacy for free services such as free email, free search, free video games et al. In the first two decades of the twenty-first century we all have been living in “The Truman Show”, perhaps unbeknownst to most of us. It is only recently that consumer privacy, individual rights and AI governance has become a cause celebre around the world accelerated by regulations like GDPR, and followed by CCPA, and ADMT (automated decision making technology), with numerous other emerging regulations that look to build more privacy into our lives than Truman had.

On one hand, and thanks to the emerging regulatory environment, consumer awareness, and internal realization, some organizations are moving to implement basic controls to achieve a “checkbox compliance” approach to data protection and governance. Improperly done Opt-out/Unsubscribe options, Bookshelf Privacy Policies, Unfulfilled retention schedules and lack of governance are all examples of a checkbox approach. Conversely collecting user data has become more important to business and bad actors have become more sophisticated in targeting the most profitable segments of our ecosystem like retail, financial services, healthcare,

public education etc to get access to personal data which can be sold to the highest bidder. The rise of AI is another reason to introduce Zero Trust Architectures at this time. AI has become a key consumer of data given the fact that AI models are trained on an organization's transactional data sets. AI models are becoming commoditized and nearly all the value is coming from the data sets AI systems have access to and are trained on.

Unless we all are comfortable becoming Mr. Truman, and we know we are not, organizations have clear incentives to leverage a zero trust data protection approach that can provide greater respect for the personal information they process and win the trust of their employees,consumers, and customers. Recently Apple has done a good job of this in portraying data privacy as one of the biggest selling points for Apple devices.

Privacy and Security Interplay: Scenarios

It is easy to say that privacy and security are the two sides of the same coin. But let's dig into this a bit further and consider two simple scenarios:

Scenario 1

An organization does its utmost to adhere to privacy regulations. They care about their consumers' right to access their own data, and the right to be forgotten. They provide their consumers control over who their data may be shared with including all 3rd parties. They manage and track consumer consent and opt-outs properly so that their consumers are not getting bombarded with unwanted campaigns when they have opted out. However, this organization routinely suffers from data breaches. Consumer data is stored in a variety of systems across the organization with little visibility, control or security protections applied.

As a consumer, would you feel comfortable doing business with such organizations, and sharing your data with this company?

Scenario 2

An organization does its utmost to secure all sensitive data, including the consumer data they have. They know exactly where all their consumer data is located whether across structured and unstructured data repositories, file shares, or in 3rd party SaaS services. They know who has access to sensitive data within their organizations, and with whom that data is shared external to their organization. Data is encrypted at rest, and even in-transit to the extent possible. Controls are analyzed and balanced to ensure no loss of utility that the data provides. With these controls in place, this organization rarely suffers from data breaches. However, this organization is yet to implement all the necessary regulatory controls and primarily operates on checkbox privacy control capabilities, leaving this organization left wanting.

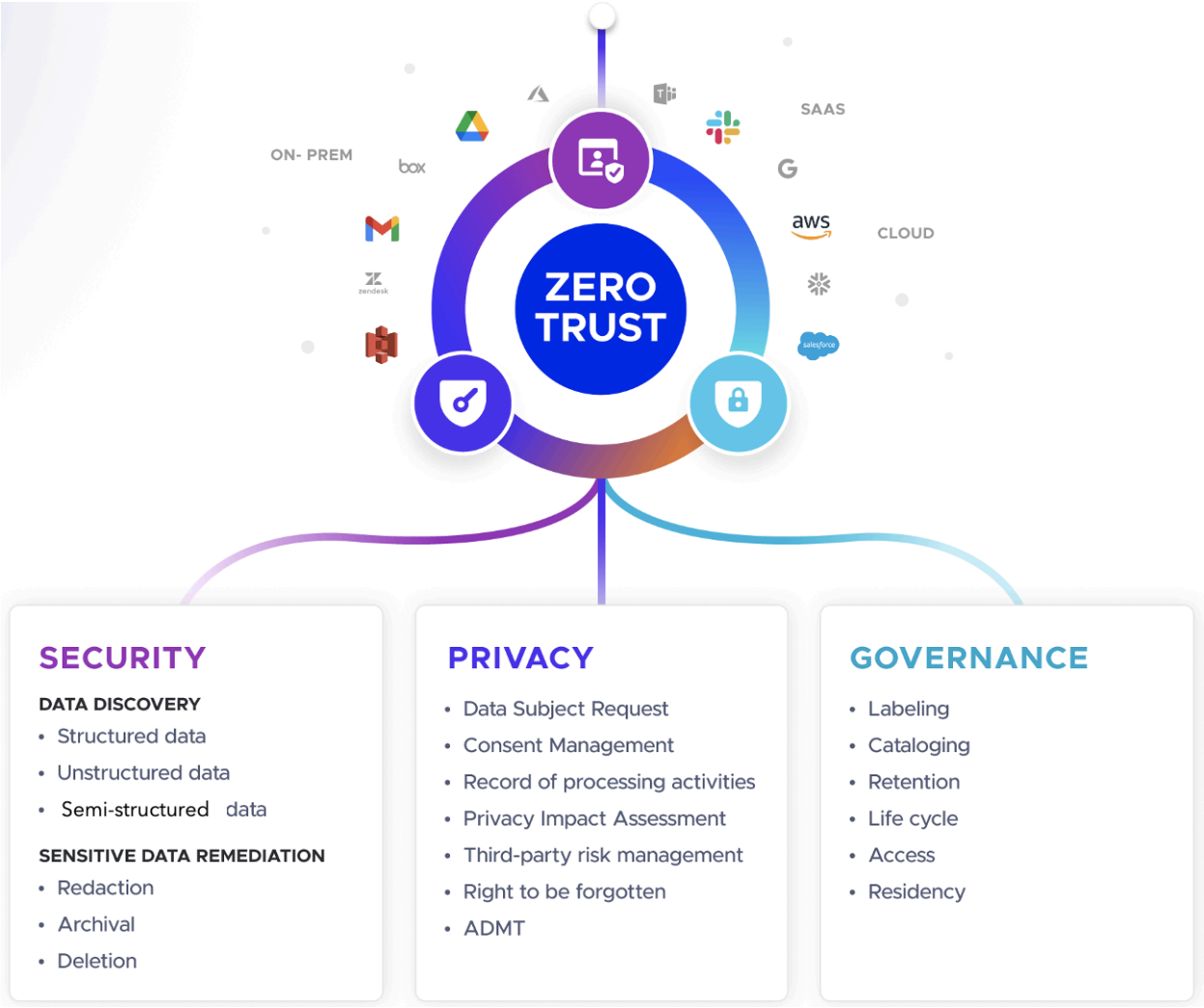
The question is simple - which of these two organizations would you feel comfortable doing business with? Obviously, neither of these organizations are in an ideal spot, but the question is do you care more about your privacy or your data security controls? In this light it may be hard to know what is best. An organization that can assure the security of my data probably deserves the business more than one that on the surface adheres to all the checkboxes but fails on the most important control of protecting personal information from inappropriate access and use.

Zero Trust Data Protection: Design Components

As explained above, it is clear that data privacy readiness can be seen as a force multiplier in developing the trust of your customers. However, if your data security readiness is missing, no matter how prepared you are for data privacy, customer trust in your business will suffer. A lack of data security and data governance readiness leading to security breaches and sensitive data exposure can wean away the hard earned trust of your customers, consumers, and employees.

The Zero Trust Data Protection framework is organized in three actionable components of Data Security, Data Privacy, and Data Governance. While the specific controls needed will depend on the business and operational context (The type of business, the geographic location of operations, and the specific data elements and who they belong to) implementing manageable controls from these three areas are the key in establishing a Zero Trust environment where personal information is both well protected and used to support business goals.

You can have security without privacy, but you cannot have privacy without security.



Data Security Controls

Data Security controls are a key part of any data protection program. By focusing on the data being processed, organizations can elevate the level of confidence and capability over the protection of data. Intelligent data protection controls can not only secure data, but also can guide its use through redaction and deletion policies to reduce the overall risk of data exposure. By automating key technical controls, data security and data governance can all work together to ensure the Confidentiality, Integrity, and Access of the data being processed.

No.	Data Security Readiness Criterion	Score your organization low if:	Score your organization high if:
S1.	**Attribute 360 (the WHAT)** A complete view of all sensitive data your organization carries within its premises.	You have a general understanding of the types of sensitive data your organization collects and keeps.	You have an up to date record of all the sensitive data elements your organization collects and stores, including the sensitive data about your customers, employees, patients, students, partners etc. You are alerted when new sensitive data is identified as entering your organization. Reporting is near real time, as scanning is continuous.
S2.	**Structured Data Map (the WHERE part I)** A complete view of all sensitive data stored in structured data repositories.	You know the structured data systems such as databases, data warehouses, data lakes where sensitive data resides.	For every structured data system within your organization, you know and can verify exactly the type of sensitive data present in them. Any sprawl of sensitive data within any structured data system (e.g. addition of a new database, or a table or a column with sensitive information) is almost immediately detected.
S3.	**Unstructured Data Map (the WHERE part II)** A complete view of all sensitive data found in unstructured data repositories.	You know the unstructured data systems such as emails, messaging platforms, file repositories, ticketing systems where sensitive data resides.	For every unstructured data system within your organization, you know and can verify exactly the type of sensitive data present in them. Any sprawl of sensitive data within any unstructured data system is almost immediately detected.
S4.	**Entity 360 (the WHO)** Whose data it is that you are carrying within your company.	You have a general understanding of the data of entities who have provided their data for processing. Data of patients, students, employees, customers etc. is processed but specific	You know exactly whose data it is that you carry within your systems. You can pinpoint all data belonging to a particular entity such as John Doe, Jane Doe etc. You can group the data you carry amongst various types of entities (customers, employees etc.).

		details are not reviewed.	
S5.	**Partner 360** How is your sensitive data getting shared (or getting leaked) outside of your organization	You understand the partners with whom you may be sharing data. This is inferred largely on the basis of partnership agreements you have in place.	You can pinpoint exactly the type of data, and the specific data elements that have been shared with each of your partners individually. You can also, if necessary, instruct your partners individually to delete the data they have received from you by providing them a detailed list of data that has been shared with them over a specific time period.
S6.	**Data Automation** How are risks contained within your organization once detected.	You can use a ticketing system to raise requests to quickly delete or mask sensitive information that might have been detected at unapproved locations. The actual correction and risk containment might take days to weeks to months.	Any sensitive information that might have been detected at unapproved locations can be automatically redacted/masked or tokenized. You accomplish this via a set of policies you define as part of your data security process.
S7.	**Access Automation for Structured Data** Policy based authorization granting data access to the right individuals	You can provide individuals/employees within your organization access to structured data systems (including parts of data warehouses, data lakes or databases) based on your roles/profiles.	Entity-centric access control: Your access control not only designates whom access is being provisioned to (e.g. your employees, members of a group etc.) but also ensures whose data it is that can be accessed. E.g. Some members of a group can access a column containing highly sensitive information such as pay records while others can't even though they have access to the database / data lake.
S8.	**Access Automation for Unstructured Data** Continuous monitoring of unstructured content being accessed by individuals	Access to unstructured content has been put in place leveraging data systems native capabilities. E.g. Access to certain files/folders within Google Drive managed via Gdrive permissions, private	Attribute and entity-centric global access control policy has been defined for each data source independent of native data systems capabilities. E.g. A policy with automated enforcement is in place ensuring no file or message containing SSN/payslip is shared with groups of employees.

		channels in Slack. No controls for access creep.	Any access creep is immediately acted upon with access revocation to a file containing sensitive info, or deletion of messages containing sensitive data etc.

Table 1: Data Security Controls

Data Privacy Controls

Let's dive deeper into data privacy controls and how automating privacy can lift the program above mere compliance and actually enable business operations and data use. Many organizations have implemented privacy controls, but historically privacy controls are very manual in nature, This can severely restrict their utility and value. Privacy teams are typically small organizations and if manual processes are not embedded into day to day operations they can become ineffective. Privacy controls should not be treated merely as legal policies although they are an important part of the privacy puzzle. To attain continuous privacy governance, an automated approach that scales and helps the business expand faster can be the difference between confident growth and stagnated compliance for companies that process personal information.

No.	Data Privacy Readiness Criterion	Score your organization low if:	Score your organization high if:
P1.	**Cookie Consent** On your website, visitors can opt out of accepting anything but the necessary cookies.	Your website provides a privacy notice and a link to your privacy policy but customers have no choices that can be made. Choices are not tracked.	You make it easy for your visitors/customers to opt-out of non essential tracking cookies, while logging and managing their responses when appropriate.
P2.	**Consent Mgmt.** Customers' consent	You have a way of logging responses at least from	You log responses from every channel of consumer interaction across all departments. You maintain

	expressed through any channel is logged, managed and acted upon centrally.	one inbound channel.	a single unified view of consumer consent and actively use that to calibrate your consumer interactions.
P3.	**Data Subject Access Requests** Your customers can make a request to you to share any and all data you are carrying about them.	You provide methods for your consumers to submit individual Rights Requests. Information that you provide to users is based on what you carry about them in your main "source of truth" database.	You have a foolproof yet easy way to validate people making these requests. The information you provide includes data from all data repositories you have - structured and unstructured. A verified user can see the data you carry about them at a moment's notice.
P4.	**Right to be Forgotten (RTBF)** Your customers can easily make a request to have you delete any data about them, subject to legal/regulatory reasons for data retention.	You provide a link to your consumers to make RTBF requests but do not have an efficient process to fulfill them.	You help your consumers understand the type of data you carry about them, the purpose and the time duration such data would be carried. You can adhere to users' RTBF requests except where data needs to be retained for legal reasons.
P5.	**Consumer Control over Data Sharing** Your customers control what data you share and with whom.	Your privacy notice explains to consumers what type of data you might be sharing with different organizations including specific names of organizations, and the purpose of that sharing, the retention period for each of those organizations	You provide an accurate view of all data that has been shared with your partners over the past 90 (or xx) days for a given consumer, along with its purpose and retention period. Furthermore, you can disable data sharing on consumers' requests, and how your services might be impacted, if any.
P6.	**Records of Processing Activity (RoPA)**	You can generate a manual RoPA report for your company and its key	RoPA reports can be generated at any point in time at any level - company, process groups, individual

	Your ability to conduct a regular sensitive data audit and generate a RoPA report.	processes, but it is a intensive and manual process.	processes. RoPA report generation has been automated with near-real time visibility into all data present in your data repositories, their purpose, active data maps and so on.

Table 2: Data Privacy Controls

Data Governance Controls

As we look at how data is governed we want to consider everything that is done to ensure data is secure, private, accurate, available, and usable. It includes the actions people must take, the processes they must follow, and the technology that supports them throughout the data life cycle. Understanding and documenting the data being processed is a large part of meeting today's demanding data protection regulations.

No.	Data Governance Readiness Criterion	Score your organization low if:	Score your organization high if:
G1	**Labeling** The types of sensitive information being processed are identified across applications.	Data resides natively in various applications without additional identification efforts.	The company has applied labels to connect to applications such as log repositories, ticketing systems, project management tools, databases, emails, messaging platforms, and file repositories to discover and label sensitive information.

G2	**Cataloging** Data Cataloging is the practice of organizing and categorizing data elements according to predefined criteria.	Data is undefined as to type, sensitivity, or structure.	The company has a method to define and classify various data types. Automatic tag assignments for structured, unstructured, financial, healthcare data are made automatically.
G3	**Retention** Data is stored for a specific period of time. Policies are created to monitor data types and retention periods.	Data is kept forever and not disposed of. Schedules for data retirement have not been created or are not enforced.	Records are identified by catalog type with specific disposal time period limits. Action is taken to remove data once its intended use is complete. Policies exist to monitor and alert data owners when retention periods have been met.
G4	**Lifecycle** Data is understood and monitored from collection through storage, processing and retirement.	Data is collected and used as determined by the business with little overall visibility.	Upon entering the organization's environment new data is identified,and classified and its use is approved through a PIA process. Retention schedules are created and data is monitored for continued use, and disposed of properly.
G5	**Access** Access to data is on a need to know basis and is controlled with administrative and technical controls	Basic role assignments drive access to data, but an accurate picture of who has access at all times is limited.	Access to sensitive data by third parties is actively monitored and action taken where inappropriate access is discovered.
G6	**Residency** Data storage is documented and not copied off premise for analytical or non transactional purposes.	SaaS based application contracts do not govern data use controls.	SaaS based analytical tools that utilize on premise sensitive data sources for analysis purposes do not replicate offsite copies of production data.

To attain continuous privacy governance, an automated approach that scales and helps the business expand faster can be the difference between confident growth and stagnated compliance for companies that process personal information.

Conclusion and the Way Forward

The data protection journey to winning your customers' trust doesn't and shouldn't stop with check the box solutions over the sensitive data being processed, or checkbox based manual data mapping exercises or even posting a comprehensive privacy policy for visitors to your website. Integrated privacy and security programs need to understand and control the data being processed by whom and for what reasons. Traditional privacy control areas that focus on data use and approval are also needed but they are not enough. Truly caring about customers' sensitive data will take you to places where data observability into every nook and cranny of your organization, to where data might be stored including engineering, marketing, finance, and operations systems amongst others is an automated and routine operation. It will lead you down a path of figuring out WHAT data you carry, WHOSE data you have, WHERE is that data stored, WHY do you have that data, WHO has access to that data, WHO are you sharing it with, and WHEN can you get rid of it.

But creating data observability is just the important first step. Upon being able to observe your data, you will move to the next level of building policy based automation such that any data risk gets contained before it can do any harm. Unwarranted exposures will get acted upon automatically before a malicious actor can get access to that data. In the unfortunate event of a partner getting compromised, you will know exactly what and whose data was shared with them

enabling you to take precise actions. Being prepared to react to a breach post-facto is a start, but adding preemptive data measures will elevate your data protection significantly. With a tabulated view of all sensitive data your organization might have shared with each of your partners, you can automatically send notices to each of your partners asking them to delete precisely the data you have shared with them.

The manual processes that we all put in place to manage and adhere to data privacy regulations over the last decade were necessary for they were the best we could do. Moving forward though, the call to action is to focus on data security with the same rigor we apply for network security. If network security controls are first line of defense (firewall), and data governance is the second line then a Zero Trust Architecture is an overall third line that can help you keep your sensitive data secure on an ongoing basis even when the first line of defense falters.

Data Protection is too important to be left to a checkbox approach.

Appendix

Revision History

Document Update	Date	Author(s)
Generally available.	Jan 24, 2024	Priyadarshi Prasad pd@lightbeam.ai Bill Schaumann bill@lightbeam.ai

About LightBeam

LightBeam streamlines and converges data security, privacy and governance, so businesses can accelerate their growth in new markets with speed and confidence.

Leveraging generative AI, LightBeam has gained industry leadership by pioneering a unique identity-centric and automation-first approach to data security. Unlike siloed solutions, LightBeam ties together sensitive data cataloging, control, and compliance across structured and unstructured data applications providing 360-visibility, sensitive data risk remediation, compliance with PCI-DSS, GLBA, GDPR, CPRA, Québec Law 25, HIPAA among other regulations. The continuous monitoring with full data residency ensures ultimate protection against ransomware and accidental exposures.

LightBeam is on a mission to create a secure privacy-first world helping customers

For any questions or suggestions, please contact us at: sales@lightbeam.ai.