CHAPTER 2

DATA GOVERNANCE & MANAGEMENT

Data is the **lifeblood** of organisations. As data grows from diverse sources — **customer interactions, IoT devices, social media, transactional systems** — organisations face a **dual challenge**: harness its potential while ensuring its **integrity, security** and **compliance**. Data governance provides the **frameworks, policies and practices** to turn raw data into a **trusted, actionable asset**.

CHAPTER ROADMAP

1. What is data governance?
2. Data classification
3. Data storage
4. Data integrity
5. Data security
6. Data stewardship
7. Metadata management
8. Compliance with data regulations
9. Emerging technologies in data governance

1 WHAT IS DATA GOVERNANCE?

DEFINITION

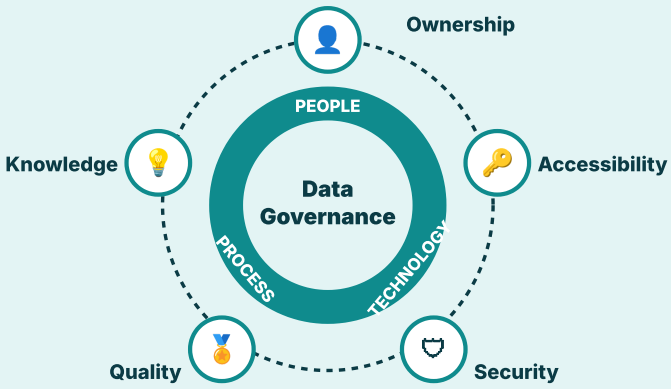
Data governance involves the creation and enforcement of policies, procedures, standards and roles to manage data throughout its lifecycle.

Creation & collection › Storage › Usage › Archival › Disposal

At its core it covers the **people, processes and technologies** needed to manage and protect data assets, so corporate data is **understandable, correct, complete, trustworthy, secure and discoverable**.

RESULT A well-implemented framework keeps data **accurate, consistent, secure** and **aligned with organisational objectives** while meeting **legal, regulatory and ethical** requirements.

Data Governance Framework



Core: **People · Process · Technology** | Pillars: **Ownership · Accessibility · Security · Quality · Knowledge**

1.1 Key Objectives of Data Governance (8)

1 Minimise Risks
Mechanisms to identify & mitigate risks of **data breaches, non-compliance, poor data quality**.

2 Internal Rules for Data Use
Clear guidelines on how data is **accessed, used and shared** within the organisation.

3 Compliance Requirements
Adherence to regulatory & legal frameworks on **data privacy and security**.

4 Improve Communication
Better internal & external communication through **standardised data practices**.

5 Increase Data Value
Maximise value by ensuring **quality and usability** for decision-making & innovation.

6 Facilitate Administration
Streamline data-related processes to **reduce complexity**.

7 Reduce Costs
Optimise data management costs by **leveraging synergies** and avoiding inefficiencies.

8 Organisational Sustainability
Long-term viability through **risk management and process optimisation**.

MEMORY HOOK **"RICC-VACS"** → **R**isks · **I**nternal rules · **C**ompliance · **C**ommunication · **V**alue · **A**dministration · **C**osts · **S**ustainability.

1.2 KEY COMPONENTS OF DATA GOVERNANCE

7 components

Data Ownership & Stewardship

1

Ensures **accountability** by assigning clear responsibilities for managing data. Two **foundational pillars** — closely related but with **distinct roles** in keeping data accurate, secure and well-managed.

DATA OWNERS
Senior-level individuals / business leaders responsible for specific data domains (e.g., customer data, financial data). Have **authority to decide** how data is used; ensure it supports business objectives.

vs

DATA STEWARDS
Operational custodians who ensure data is **accurate, available and properly documented**. Focus on **day-to-day management** and compliance with data policies & standards.

Data Policies & Procedures

2

Documented **rules, standards and guidelines** on how data is handled — consistently, securely, and aligned with legal & business requirements.

Policies define **who** has access to what, **conditions** for sharing, **how long** to retain, how to **dispose securely**.

Procedures give **step-by-step instructions** for policy compliance (e.g., requesting access, reporting a quality issue).

Data Quality Management

3

Monitoring, measuring and improving key **data quality dimensions**:

- **Accuracy** — correct & error-free?
- **Consistency** — same across systems?
- **Completeness** — all required data present?
- **Timeliness** — up to date & available?

Good quality → decisions, efficiency, compliance. **Poor quality → errors, financial losses, reputational damage.**

Data Security

4

Protecting sensitive data from **unauthorised access, alteration or destruction** through:

Technical	Administrative	Physical
encryption, firewalls, access controls	policies, user training	secure server rooms, surveillance

Crucial for **trust**, **regulatory compliance** and preventing breaches → financial loss / legal penalties.

Compliance & Risk Management

5

Data handling meets **legal, regulatory and contractual** obligations:

International	GDPR (EU), CCPA (California)
National	Pakistan's PECA, 2016
Industry standards	ISO 27001, PCI-DSS
Risk management	Identify, assess & mitigate risks of misuse, breaches, non-compliance

Avoids penalties, legal action, reputational damage; builds **accountability & transparency**.

Metadata Management

6

Metadata = "data about data" — describes the **origin, format, definitions, usage and relationships** of data elements. Enables better data **discovery, understanding, classification and governance** (supports analytics, compliance, integration). Involves:

- **Standardised metadata definitions**
- **Metadata repositories / data catalogs**
- **Consistent use & interpretation of data across systems**

Data Lifecycle Management (DLM)

7

Managing data **systematically** from initial creation to ultimate archival or deletion — keeping it **accurate, secure, accessible and compliant**.

① Creation / Acquisition
Generated internally or acquired externally. Ensure accuracy, **define ownership**, apply classification (public / confidential / restricted) — esp. for **PII**.

›

② Storage
Secure environment: databases, data lakes, file systems, cloud. Encryption, access controls & **backups** → availability & disaster recovery.

›

③ Usage
Active phase: accessed, processed, analysed, shared. Only authorised access; tracked & logged. **Cleansing & validation** applied.

›

④ Archiving
Inactive data with business/legal/historical value → **long-term, lower-cost** storage. Typically **read-only**; retention & audits.

›

⑤ Deletion / Disposal
Permanent removal: **shredding, data wiping, cryptographic erasure**. Document destruction to show compliance.

DLM keeps data **efficient and relevant**, **minimises storage costs**, and complies with **retention laws & privacy regulations**.

1.3 LEVELS OF DATA GOVERNANCE

strategic · tactical · operational



A **structured approach** to managing data assets across an organisation. Each level plays a distinct role in ensuring data is **managed effectively**, **aligns with business goals** and **supports decision-making**.

Blue bar = Data Governance Program · **↓ orange** = Decision & Escalation Path · **↑ grey** = Communication Path *(study-text figure)*

STRATEGIC	Provides the **vision & direction**
TACTICAL	Translates vision into **actionable plans & processes**
OPERATIONAL	Executes plans **day-to-day**

① STRATEGIC

Defines the overarching **vision, goals and policies**; aligns governance with broader business objectives & long-term vision.

Vision & Goals — data as a **strategic asset** (on par with financial, human, physical assets); set high-level goals	**e.g.** become a data-driven organisation using analytics for competitive advantage
Data Governance Policies — how data is managed, accessed, protected	**e.g.** privacy regulations, retention schedules, security standards
Aligning with Business Objectives — support mission, vision, priorities	**e.g.** expanding into new markets → accurate, consistent customer data across regions
Governance Frameworks — structure, roles, responsibilities	**e.g.** a data governance council oversees strategic decisions
Executive Sponsorship — senior leadership champions it & secures resources	**e.g.** CEO or CFO sponsors the program for funding & attention

Importance: clear direction · alignment with business objectives · **foundation** for tactical & operational activities.

② TACTICAL

Translates the strategic vision into **actionable plans, processes and roles**; **bridges the gap** between strategy and day-to-day operations.

Implementation Plans — timelines, milestones, resource allocation	**e.g.** steps for a data quality improvement program over 12 months
Roles & Responsibilities — data stewards, data owners, data custodians	**e.g.** a data steward ensures accuracy of customer data in a system
Processes & Workflows — quality monitoring, metadata management, issue resolution	**e.g.** workflow for how quality issues are identified, reported, resolved
Data Standards — naming conventions, data formats, classification schemes	**e.g.** all addresses stored as street, city, state, ZIP
Monitoring & Reporting — track progress of initiatives	**e.g.** regular reports on data quality metrics & policy compliance

Importance: turns strategic goals into actionable steps · consistent implementation across organisation · structured approach.

③ OPERATIONAL

Day-to-day execution of the policies, processes and standards defined at strategic & tactical levels.

Data Quality Management — regular checks for accuracy, completeness, consistency	**e.g.** automated scripts to find & fix duplicate customer records
Access Control & Security — only authorised users access sensitive data	**e.g.** RBAC to restrict access to financial data
Metadata Management — repositories for definitions, **lineage**, usage	**e.g.** update catalog when a field is renamed or schema modified
Issue Resolution — fix inconsistencies, errors, breaches in time	**e.g.** see 6 issue types on next page
Compliance Monitoring · **User Training & Support**	**e.g.** audits to verify compliance · workshops on data privacy

Importance: consistent daily practice · ongoing quality, security & compliance · informed decisions from reliable data.

⚠ Issue Resolution — Common Types of Data Issues

Data Inaccuracy
Incorrect values — manual entry errors, outdated records, integration problems.
Outdated customer contact no.; wrong transaction amount

Data Inconsistency
Conflicting data in different systems — lack of synchronisation / standardisation.
Supplier name differs in procurement vs finance

Duplicate Data
Multiple records for the same entity → confusion & redundancy.
Customer appears twice with different spellings / IDs

Missing / Incomplete Data
Required fields blank or not properly captured.
Employee records missing tax identification numbers

Data Format Issues
Inconsistent formats (dates, currencies) preventing integration & analysis.
DD/MM/YYYY in one system, MM/DD/YYYY in another

Security Breaches · Latency
Unauthorised access / leaks; delays in updates → outdated info used.
Confidential reports accessed; real-time sales dashboard lagging

RBAC — Role-Based Access Control
A security approach where access is granted **based on a user's role** in the organisation — e.g., **finance staff** can access budget reports but **not HR records**.

Data Lineage
The **history or lifecycle** of data — how it **moves and transforms from source to destination**; important for **traceability and trust**.

Interplay Between the Three Levels — one goal, three levels

STRATEGIC — Goal: improve data quality to support analytics-driven decision-making
› **TACTICAL** — Plan to implement data quality tools & assign data stewards
› **OPERATIONAL** — Stewards perform daily quality checks & resolve issues as they arise

2 DATA CLASSIFICATION

by sensitivity, value or intended use

Categorising data based on its **sensitivity, value or intended use**. Proper classification lets organisations **apply appropriate security**, **allocate resources effectively** and **ensure regulatory compliance**.

2.1 Types of Data Classification — sensitivity increases →

Public
Openly shared **without risk**.
e.g. marketing materials, press releases

Internal
For **internal use**, not public disclosure.
e.g. internal reports, employee handbooks

Confidential
Sensitive; **restricted access & robust protection**.
e.g. customer records, financial data, trade secrets

Restricted
Highly sensitive; strict legal / contractual obligations.
e.g. PII, health records, classified govt data

2.2 Common Data Classification Models

① Sensitivity-Based — impact of unauthorised access/disclosure

Level	Description	Example
Public	Freely shared without risk	Brochures, published annual reports
Internal	Internal use only; low risk if disclosed	Internal emails, SOPs
Confidential	Could cause harm; limited access	Financial records, salaries
Restricted	Disclosure could cause serious damage	Trade secrets, legal docs, customer PII

② Role-Based — user roles & need-to-know

Level	Who can access	Example
Executive	Executives only	Strategic plans, merger documents
Departmental	Specific departments	Sales pipeline, inventory data
General Staff	All employees	Company policies, HR forms

③ Compliance-Based — aligned with legal requirements

- **PII** – names, ID numbers
- **PHI** – medical histories
- **Financial** – credit card, tax info
- **Sensitive business** – pricing models, contracts

④ Lifecycle-Based — stage in the data lifecycle

Stage	Classification
Creation	Draft, Temporary
Usage	Active, In Use
Archival	Historical, Archived
Destruction	Expired, Ready for Deletion

⑤ Government Model (example) — public sector / defence

Unclassified	Public information
Restricted	Minor damage if disclosed
Confidential	Damage to national security
Secret	Serious damage
Top Secret	Exceptionally grave damage

2.4 Classification Frameworks

Organisations adopt **ISO/IEC 27001** or the **NIST Cybersecurity Framework** — methodologies for **assessing data sensitivity** and determining **handling procedures**.

EXAMPLE · FINANCIAL INSTITUTION

Public: published interest rates | **Internal: training manuals**

Confidential: client banking details | **Restricted: card numbers (PCI DSS)**

Labels + access controls → only authorised staff → fewer breaches; compliance with **SBP ETGRMF**.

3 DATA STORAGE

accessibility · reliability · security

Methods & technologies to store data. Choose based on **data types, volumes, usage patterns and compliance requirements**.

On-Premises
own physical servers

Servers **owned & managed** by the organisation.

+ Greater **control** over security & customisation

– Significant **capital investment**, maintenance, IT expertise

Cloud
third-party remote servers

Managed by providers (AWS, Microsoft Azure, Google Cloud).

+ **Scalability, flexibility, cost-efficiency**

– Dependence on **vendor security** & **internet connectivity**

Hybrid
on-premises + cloud

Balances **control and scalability**.

Sensitive data → on-premises

Less critical data → cloud

3.2 Emerging Trends

- **Edge Storage** — store data **closer to where it is generated** (e.g., IoT devices) → lower latency, real-time processing.
- **Blockchain-Based** — **decentralised**; enhances security & **immutability** (e.g., supply chain tracking).
- **Data Lakes** — centralised repositories for **raw, unstructured & structured** data at scale; used with big data analytics (Ch 4).

CASE · CLOUD STORAGE IN E-COMMERCE

Millions of transactions in peak seasons (Black/Blessed Friday, Eid sales) → cloud to **scale storage dynamically** + encryption & access controls. Must negotiate robust **SLAs** for data-protection compliance and against **data sovereignty** issues.

4 DATA INTEGRITY

accuracy · consistency · reliability through the lifecycle

Critical for **trustworthy decision-making** and operational efficiency — corrupted or inaccurate data leads to **costly errors**.

Physical

Protect from **physical disruptions** — hardware failures, power outages.

Logical

Consistent & accurate in databases via **constraints, validations, relationships**.

Entity

Each record **uniquely identifiable** (e.g., **primary keys**).

Referential

Consistency **between related tables** (e.g., **foreign key** constraints).

Customers

Customer_ID 🔑 PK	Name
101	Ali
102	Sara

←

Orders

Order_ID	Customer_ID FK
5001	101
5002	102

Primary key — unique identifier for each row; **unique values, no NULLs** (e.g., Customer_ID in Customers).

Foreign key — field that **refers to the primary key** of another table; orders recorded only for **valid customers**.

4.2 Key Practices to Maintain Data Integrity

Data Validation
Checks for accuracy during **entry, updates or imports**.

Auditing & Monitoring
Regularly review processes to find & fix discrepancies.

Access Controls
Limit modifications to authorised users via **RBAC**.

Version Control
Track changes; prevent overwrites; keep **audit trail**.

Data Backups
Regular backups to recover from **corruption or loss**.

4.3 Practice ↓ / Integrity →	Physical	Logical	Entity	Referential
Data Validation	×	Input constraints, range checks	Enforcing unique values for primary keys	Ensuring correct foreign key values
Auditing & Monitoring	Detects physical failures / logs	Identifies logic errors or anomalies	Identifies duplicate or missing keys	Checks relationship consistency
Access Controls (RBAC)	×	Prevents unauthorised changes that break logic	Prevents accidental deletion / duplication of records	Prevents deletion of parent records linked via foreign keys
Version Control	×	Tracks logical changes; rollback capability	Helps trace changes to primary keys	Traces updates to related records
Data Backups	**Protects against hardware failure / data loss**	Restores consistent state	Restores records with valid unique IDs	Restores valid linked relationships

NOTE Only **Auditing & Monitoring** and **Data Backups** map to **all four** types. Validation, Access Controls and Version Control do **not** address **physical** integrity (×).

EXAMPLE · DATA INTEGRITY IN SUPPLY CHAIN MANAGEMENT

Checksums validate data during transfers | RBAC restricts edits | Periodic audits for consistency →

If compromised (corrupted shipment record) → delays, lost goods, strained supplier relationships

5 DATA SECURITY

multi-layered approach

Protecting data from **unauthorised access, breaches and cyberattacks**. Rising frequency & sophistication of threats → organisations need a **multi-layered approach**.

5.1 Data Security Practices

🔐 Encryption
Algorithms encode data; readable only with the **correct key**. **AES-256** for data **at rest**; **TLS** for data **in transit**.

🔢 Multi-Factor Authentication (MFA)
Multiple forms of verification — e.g., password + SMS code + biometric.

🧱 Firewalls & IDS
Monitor & filter network traffic to **detect and block** malicious activity.

💻 Endpoint Security
Secure **devices** (laptops, mobiles) via antivirus & device encryption.

🚫 Data Loss Prevention (DLP)
Tools to monitor & prevent **unauthorised data exfiltration** (e.g., blocking sensitive data emailed externally).

🛂 Zero Trust Architecture
"Never trust, always verify" — access only after **continuous verification** of identity & context.

5.2 Emerging Threats & Solutions

Threat	What it is	Mitigation / Solution
Ransomware	Malware that **encrypts data and demands payment** for decryption	Regular **backups** + employee training on **phishing awareness**
Insider Threats	Risks from **employees or contractors** with access to sensitive data	**Behaviour monitoring** + **least-privilege** access policies
AI-Powered Attacks	AI used for sophisticated **phishing** or vulnerability exploitation	**AI-driven threat detection** systems

CASE · DATA SECURITY IN FINANCIAL SERVICES

AES-256 for data at rest | MFA for employee access | Real-time intrusion detection + PCI DSS & SBP ETGRMF →

Minimised breach risk, customer trust maintained

6 DATA STEWARDSHIP

Complements data governance by assigning **data stewards** to oversee **day-to-day management** of data assets. Stewards act as **liaisons between IT and business units**.

Policy Enforcement
Adherence to policies & standards

Issue Resolution
Duplicates, inconsistencies

Training & Awareness
Educate employees on best practices

Metadata Management
Lineage, cataloguing, discovery

EXAMPLE · RETAIL
Stewards oversee customer data across CRM systems — accurate records, resolve duplicates, train store managers on data entry → better marketing & customer satisfaction.

7 METADATA MANAGEMENT

Metadata — data about data — provides **context** that improves usability & governance. Creating, storing & maintaining it supports **discovery, lineage tracking, compliance**.

Type	What it does	Examples
Descriptive	Describes the **content**	Title, author, keywords
Structural	How data is **organised**	Database schema, file structure
Administrative	Info for **managing** data	Access permissions, creation date, retention policies

✓ **Discoverability for analytics** ✓ **Compliance via lineage & usage** ✓ **Shared understanding across teams**

EXAMPLE · MEDIA
Video assets tagged with **descriptive** (genre, actors) and **administrative** (copyright, expiry date) metadata → efficient retrieval & licensing compliance.

8 COMPLIANCE WITH DATA REGULATIONS

a cornerstone of data governance

Key Data Regulations in Pakistan

SBP ETGRMF — **SBP Enterprise Technology Governance & Risk Management Framework** — for **financial institutions**: data protection, incident response, technology audits.

SECP — **SECP Guidelines for Cyber Security** — directs **corporations** to protect sensitive business & client data.

PECA 2016 — **Prevention of Electronic Crimes Act** — criminalises **hacking, identity theft, cyber terrorism**; empowers investigation & prosecution.

NSCP — **National Cyber Security Policy** — national strategy for cyberspace; **public-private collaboration** & resilience.

ETO 2002 — **Electronic Transaction Ordinance** — legalises **electronic transactions & signatures**.

CASE · SBP COMPLIANCE IN BANKING
A Pakistani bank under SBP ETGRMF:

- Robust data protection
- Regular risk assessments
- Incident response protocols
- Encrypt customer data, **RBAC**, phishing training

Non-compliance → penalties, reputational damage, loss of customer trust.

9 EMERGING TECHNOLOGIES IN DATA GOVERNANCE

AI & ML
Automate **quality checks**, detect **anomalies**, predict governance risks.

Blockchain
Decentralised, **tamper-proof ledger** for lineage & integrity (finance, healthcare).

Governance Platforms
Collibra, Informatica, Alation — policy mgmt, metadata cataloguing, compliance tracking.

Privacy-Enhancing Computation (PEC)
Secure processing **without compromising privacy**.

Homomorphic Encryption
Computations performed **directly on encrypted data** without decrypting; decrypted result = result on raw data. *e.g., a bank runs risk analysis on encrypted customer financial data without seeing actual values.*

Federated Learning
ML model trained across **decentralised devices/servers** holding local data — **no exchange of actual data**; only model updates (gradients) shared. *e.g., hospitals build a shared model without moving patient records.*

EXAMPLE · AI IN DATA GOVERNANCE
A multinational's AI platform monitors data quality across subsidiaries — **flags duplicates, suggests remediation, generates compliance reports** → less manual effort, better efficiency.

EXAM FOCUS · MCQ TRAPS

commonly confused points

1 Owner vs Steward
Owner = **senior leader, decision authority** over a data domain. Steward = **operational custodian**, day-to-day accuracy & documentation.

2 Policy vs Procedure
Policy = **what/who/how long** (rules). Procedure = **step-by-step how** to comply.

3 Which level?
Executive sponsorship, vision, governance council = **Strategic**. Roles, standards, workflows, plans = **Tactical**. Daily checks, RBAC, issue fixing, training = **Operational**.

4 Confidential vs Restricted
Restricted is the **highest** business level — PII, health records, trade secrets (sensitivity model). Confidential = salaries, financial records.

5 Government model order
Unclassified → Restricted → Confidential → **Secret** → **Top Secret** (exceptionally grave damage).

6 Entity vs Referential
Entity = unique records (**primary key**, no NULLs). Referential = links between tables (**foreign key**).

7 Encryption standards
AES-256 → data at rest; **TLS → data in transit**. Don't swap them.

8 Edge vs Data Lake
Edge = store **near the source** (IoT) for low latency. Data lake = **centralised** raw store at scale.

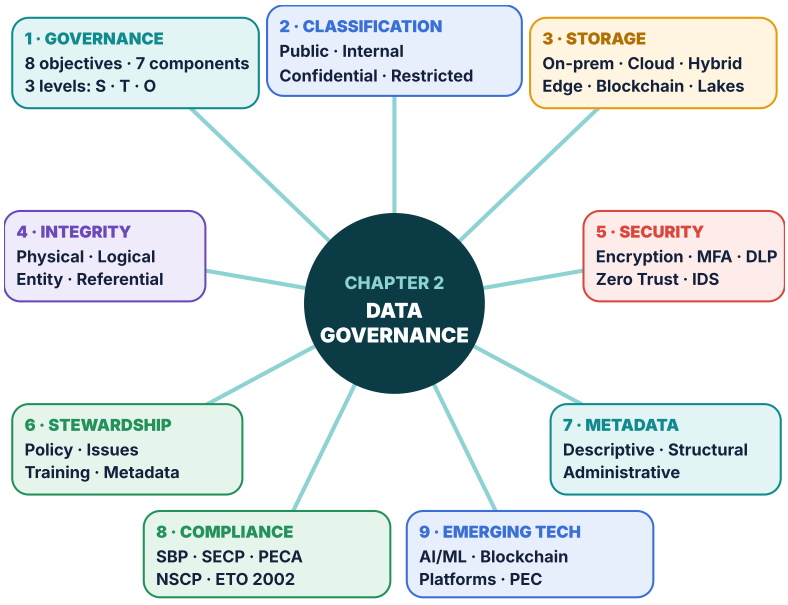
9 Metadata types
Schema / file structure = **Structural**. Access permissions, retention, creation date = **Administrative**. Title, author, keywords = **Descriptive**.

10 PEC techniques
Homomorphic = compute on **encrypted** data. Federated = train model **without moving data**.

11 Who regulates whom?
SBP ETGRMF → **financial institutions**. SECP cyber guidelines → **corporations**.

12 Disposal methods
Shredding, data wiping, **cryptographic erasure** — and **document** the destruction.

REVISION MAP



CLASSIFY IT

1	CEO sponsors the program	Strategic
2	Address format standard	Tactical
3	Scripts fix duplicates	Operational
4	Employee salaries	Confidential
5	Customer PII	Restricted
6	Database schema	Structural metadata
7	Customer_ID in Orders	Foreign key
8	Blocking sensitive emails out	DLP

Governance
Policies, roles & standards across the whole data lifecycle.

Classify
Sensitivity decides protection & access.

Store & Protect
Right storage + integrity + multi-layered security.

People
Owners decide, stewards manage day-to-day.

Comply
SBP, SECP, PECA, NSCP, ETO + new tech (AI, PEC).