Spotify

Data Governance Framework & Implementation Plan

Designed by
Arthur Mbomo : Data governance analyst

1. Current State Overview

Maturity Level: Globally *Proactive* (Gartner Scale: 3)

Strengths:

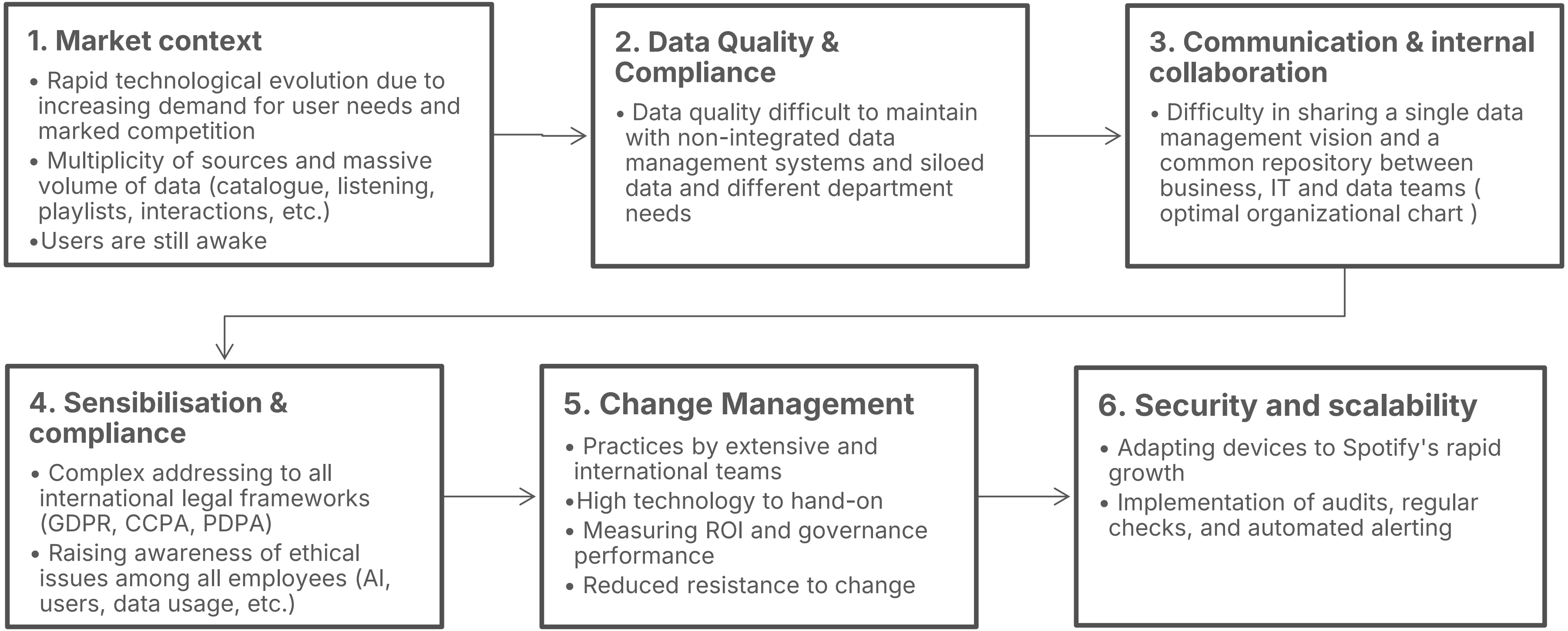
- Robust regulatory compliance (GDPR, CCPA) and data quality practices.
- Advanced data infrastructure (data lakes, databases, cloud) and rapid ingestion and data processing.
- Business & Data operations teams highly specialized and autonomous
- Existing collaboration between DPO, CDO, and cross-functional teams.

Weaknesses:

- Fragmented data and siloed systems.
- Inconsistent security controls and lack of centralized governance.
- Limited automation and AI bias management.

1. Current State Overview

Major challenges of Data Governance at Spotify



1. Market context

- Rapid technological evolution due to increasing demand for user needs and marked competition
- Multiplicity of sources and massive volume of data (catalogue, listening, playlists, interactions, etc.)
- Users are still awake

2. Data Quality & Compliance

- Data quality difficult to maintain with non-integrated data management systems and siloed data and different department needs

3. Communication & internal collaboration

- Difficulty in sharing a single data management vision and a common repository between business, IT and data teams (optimal organizational chart)

4. Sensibilisation & compliance

- Complex addressing to all international legal frameworks (GDPR, CCPA, PDPA)
- Raising awareness of ethical issues among all employees (AI, users, data usage, etc.)

5. Change Management

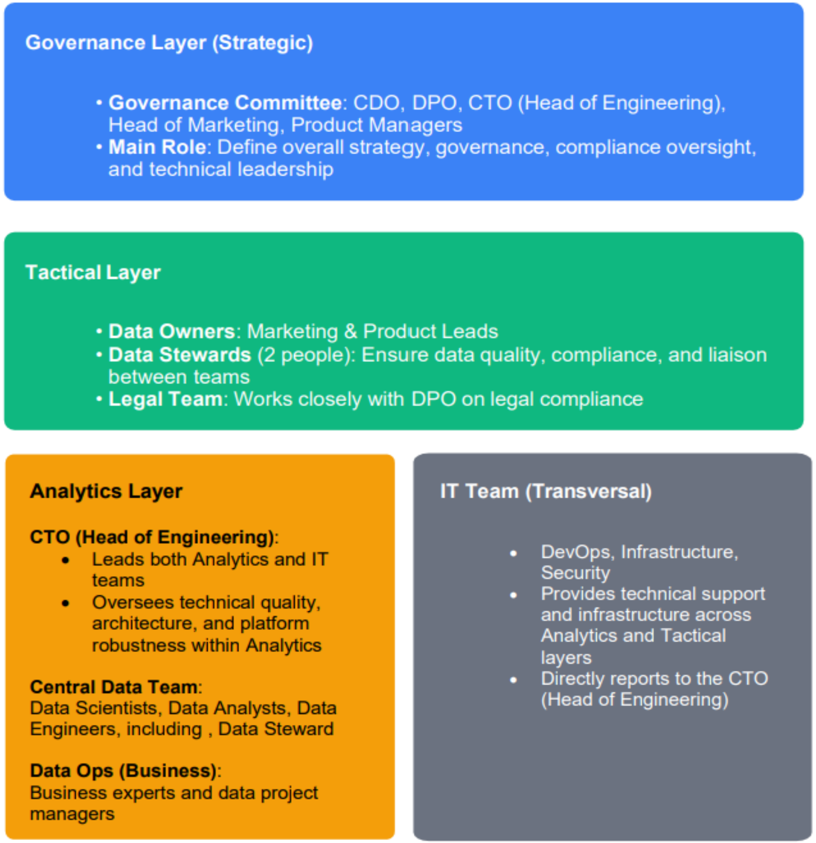
- Practices by extensive and international teams
- High technology to hand-on
- Measuring ROI and governance performance
- Reduced resistance to change

6. Security and scalability

- Adapting devices to Spotify's rapid growth
- Implementation of audits, regular checks, and automated alerting

2. Data Governance Framework

Three-Tier Governance Structure:



Governance Layer (Strategic)

- **Governance Committee**: CDO, DPO, CTO (Head of Engineering), Head of Marketing, Product Managers
- **Main Role**: Define overall strategy, governance, compliance oversight, and technical leadership

Tactical Layer

- **Data Owners**: Marketing & Product Leads
- **Data Stewards** (2 people): Ensure data quality, compliance, and liaison between teams
- **Legal Team**: Works closely with DPO on legal compliance

Analytics Layer

CTO (Head of Engineering):

- Leads both Analytics and IT teams
- Oversees technical quality, architecture, and platform robustness within Analytics

Central Data Team:
Data Scientists, Data Analysts, Data Engineers, including , Data Steward

Data Ops (Business):
Business experts and data project managers

IT Team (Transversal)

- DevOps, Infrastructure, Security
- Provides technical support and infrastructure across Analytics and Tactical layers
- Directly reports to the CTO (Head of Engineering)

Spotify Data Governance Policy

Data is classified into three sensitivity levels, each with specific protection measures:

Red Level (Sensitive Data)

User identification data protected with AES-256 encryption, pseudonymization, and strict least-privilege access

Orange Level (Operational Data)

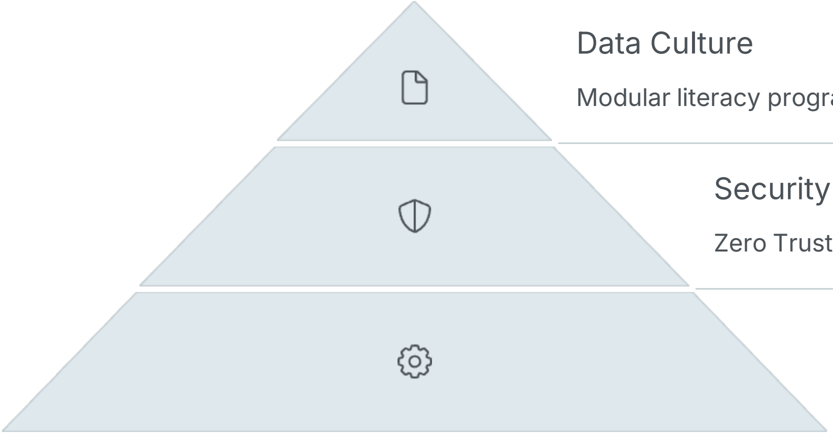
Music preferences and interaction logs managed with anonymization, role-based access control, and Collibra catalog tagging

Green Level (Public Data)

Artist metadata and catalogs with rigorous quality checks and open internal access

Security, Compliance and Data Management

Spotify implements robust security measures and compliance protocols while fostering a strong data culture. The infrastructure leverages multiple specialized tools to ensure data integrity throughout its lifecycle.



Data Culture

Modular literacy programs, workshops, and ambassadors

Security & Compliance

Zero Trust Architecture okta, RBAC, 2FA, ISO 27001, Splunk,Firewall

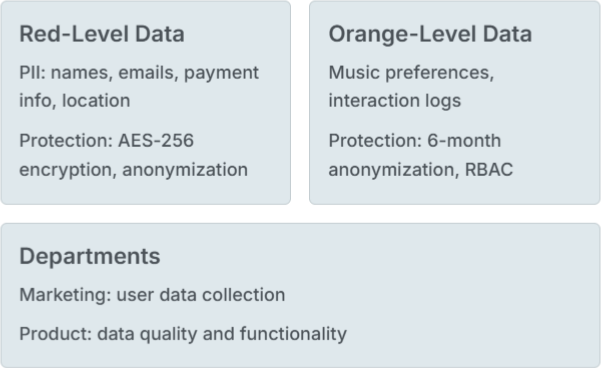
Tools & Infrastructure

Databricks, Collibra, AWS, Informatica, API Spotify

Security is maintained through Zero Trust Architecture with Okta, role-based access control via Splunk, mandatory 2FA, and AES-256 encryption for sensitive data. Regulatory compliance is ensured through VeraSafe and TrustArc, with strict adherence to user rights and 72-hour breach notification.

Data quality is measured through Informatica dashboards with KPIs for completeness, accuracy, consistency, and integrity. The architecture integrates AWS Redshift for master data, Microsoft SQL Server for operational relational databases, and AWS S3 for unstructured data, all connected through APIs to eliminate silos.

3. Implementation Plan



Red-Level Data

PII: names, emails, payment info, location

Protection: AES-256 encryption, anonymization

Orange-Level Data

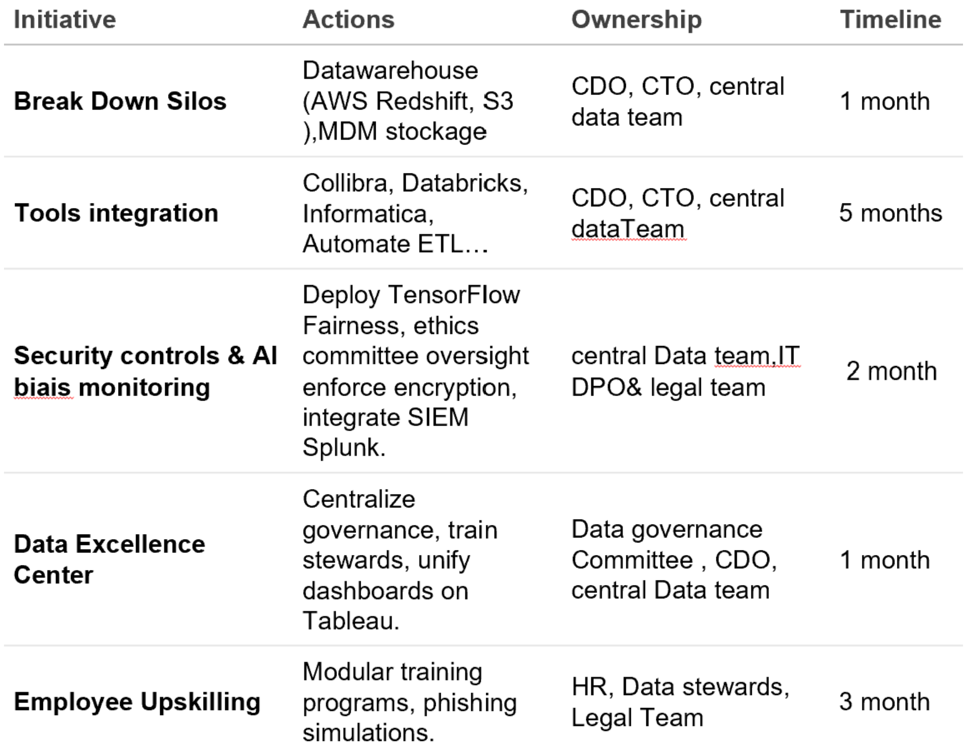
Music preferences, interaction logs

Protection: 6-month anonymization, RBAC

Departments

Marketing: user data collection

Product: data quality and functionality



Initiative	Actions	Ownership	Timeline
Break Down Silos	Datawarehouse (AWS Redshift, S3),MDM stockage	CDO, CTO, central data team	1 month
Tools integration	Collibra, Databricks, Informatica, Automate ETL…	CDO, CTO, central dataTeam	5 months
Security controls & AI biais monitoring	Deploy TensorFlow Fairness, ethics committee oversight enforce encryption, integrate SIEM Splunk.	central Data team,IT DPO& legal team	2 month
Data Excellence Center	Centralize governance, train stewards, unify dashboards on Tableau.	Data governance Committee , CDO, central Data team	1 month
Employee Upskilling	Modular training programs, phishing simulations.	HR, Data stewards, Legal Team	3 month

4. Expected Outcomes

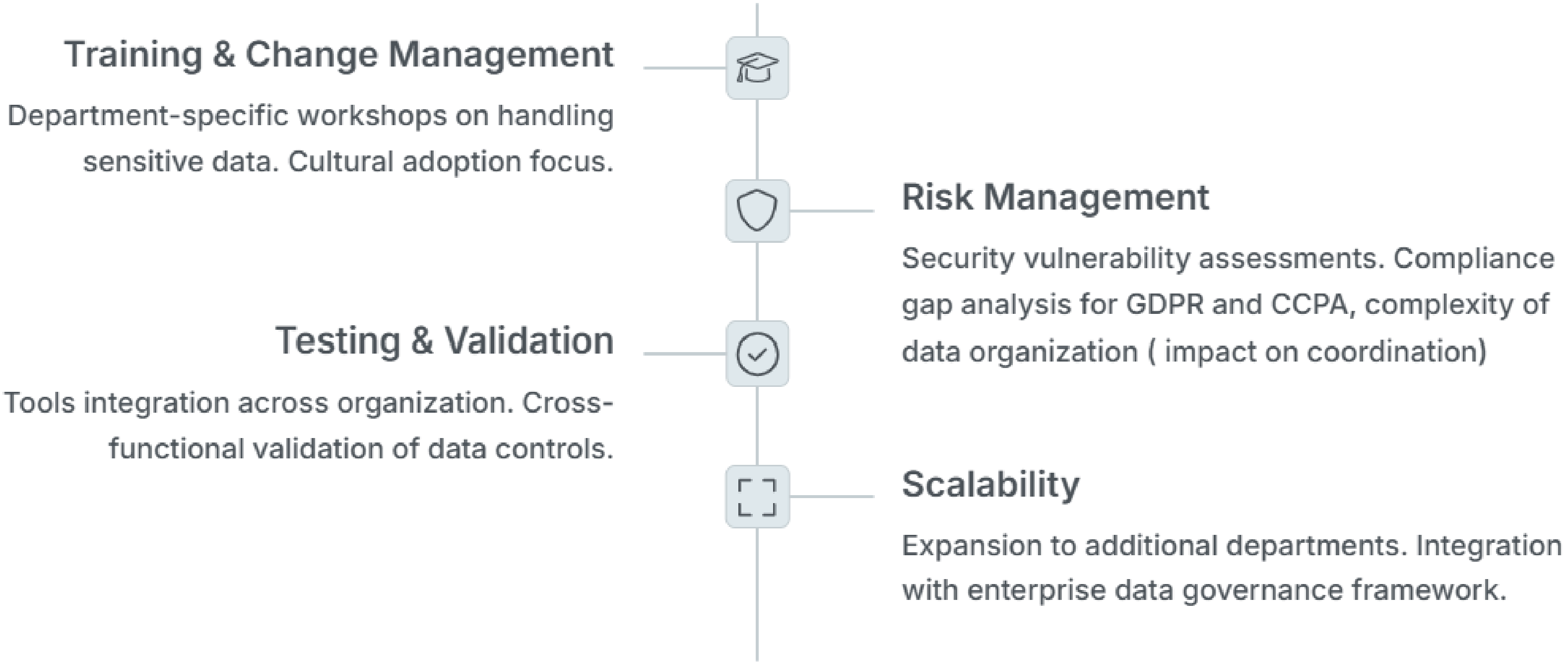
By 2025:

- **Security:** 0 major incidents/year.
- **Quality:** <0.5% data errors.
- **Compliance:** GDPR response ≤10 days
- Data privacy : alert on 72 hours.

Long-Term:

- Unified data ecosystem.
- Trusted AI-driven recommendations.
- Global regulatory alignment.

4. Next Steps



Training & Change Management

Department-specific workshops on handling sensitive data. Cultural adoption focus.

Risk Management

Security vulnerability assessments. Compliance gap analysis for GDPR and CCPA, complexity of data organization (impact on coordination)

Testing & Validation

Tools integration across organization. Cross-functional validation of data controls.

Scalability

Expansion to additional departments. Integration with enterprise data governance framework.

Thanks!

See you in the next course