Spotify Data Governance

Data Governance and Strategy

Context

- **Spotify**: world leading audio streaming platform

- Manages a vast and complex data ecosystem
  user interactions, subscription data, content metadata, ...

- Growing needs demand a robust data governance framework

Context

- **Spotify**: world leading audio streaming platform
- Manages a vast and complex data ecosystem
  user interactions, subscription data, content metadata, ...
- Growing needs demand a robust data governance framework

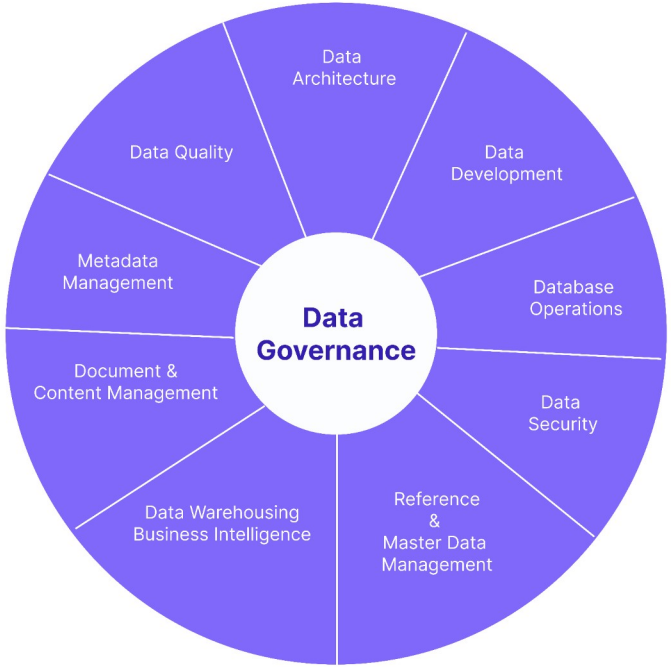
Tasks

- Assess current data maturity of Spotify and identify key challenges
- Propose a data governance plan for Spotify
- Propose an implementation for the governance framework
- Present the work to stakeholders

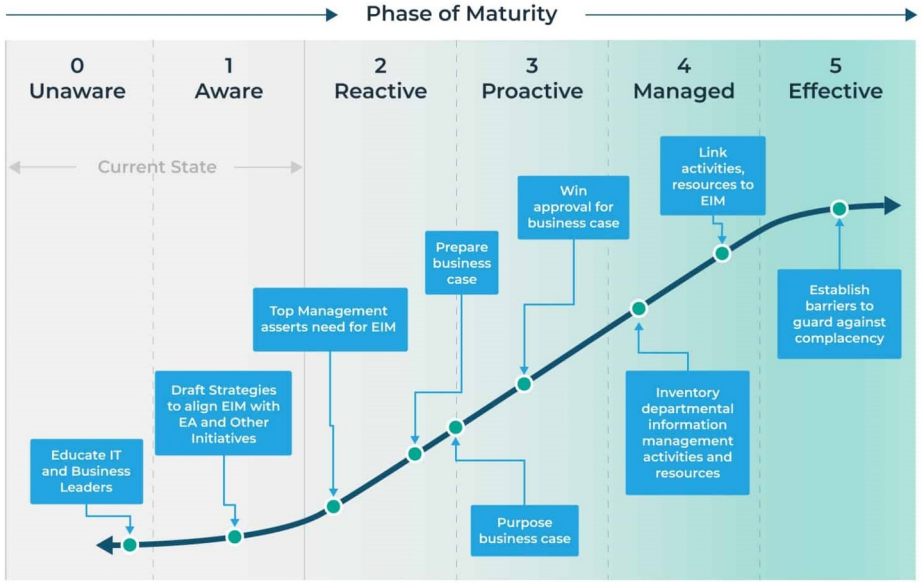
Spotify Data Landscape

Data maturity assessment

DMBoK data management framework



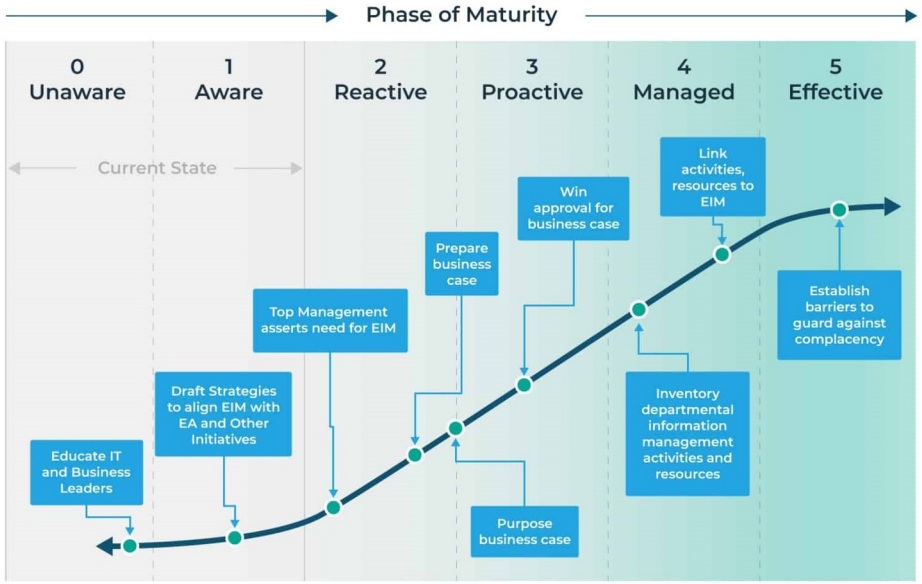
Gartner data maturity scale



Data maturity assessment

- Overall maturity level :
  Reactive to Managed (2 - 4)
  - Strong infrastructure and analytics
  - Uneven governance practices
- Strengths
  - Advanced real-time data infrastructure
  - Robust analytics and ML for personalization
  - High data volume handling
- Need to unify data governance across all departments

Gartner data maturity scale



Key Challenges

- **Data silos and fragmentation**
  Legacy data handling by independent departments. Needs a unified data view.

- **Regulatory compliance complexity**
  Navigate in rapidly evolving regulations across 180+ countries. Need dynamic strategy and stronger ovesight.

- **Low metadata quality**
  Metadata inconsistently applied to various products (e.g. podcast and songs). Impacts systems such as recommendation engines.

- **Access and integration barriers**
  Disconnected systems slow down data access. Need a unified data integration strategy.

- **User privacy and trust**
  Increased user awareness of data rights. Requires better consent management and anonymization protocols..

- **Weak data culture**
  Limited and non-unified data literacy and governance awareness.

Data Governance Framework

Data Governance Principles

- **Accountability**
  Assign clear data ownership and stewardship across teams.
- **Transparency**
  Data usage by Spotify must be clear to users.
- **Data Security**
  Protect sensitive data, safeguard against breaches.
- **Data Quality**
  Ensure data accuracy, consistency and completeness.
- **Data Minimization**
  Collect and retain only necessary data.
- **Compliance**
  Ensure compliance with regulations (GDPR, CCPA, etc).
- **Ethical Use**
  Commit to responsible and privacy-centered data use.
- **User Rights**
  Provide tools allowing for a fine-grained control of their data by users.
- **Continuous Improvement**
  Regularly review and adapt governance practices.

Regulatory Compliance

Requirement	Regulation	Implementation
Data processing principles	GDPR	Transparent, user-friendly notices in every market
User Rights and Consent Management	GDPR/CCPA	Portal for access, deletion, and correction requests ; consent banners, opt-in/opt-out options for cookies and analytics
Data Breach Notification	GDPR	Incident response procedure and breach dashboard overseen by DPO
Secure and Monitor Network & Systems	PCI-DSS	Implement security measures such as firewalls ; assess security through pentesting, ensure proper logging
Protect and Control Access to Data	PCI-DSS	Setup encryption and secure storage ; define access rules for the different data domains and limit access authorized personnel
Vulnerability Management Program	PCI-DSS	Perform regular security audits, ensure systems are up-to-date
Information Security Policy	PCI-DSS	Create and maintain a definite information security policy, ensure regular training of the teams

Roles and Responsibilities

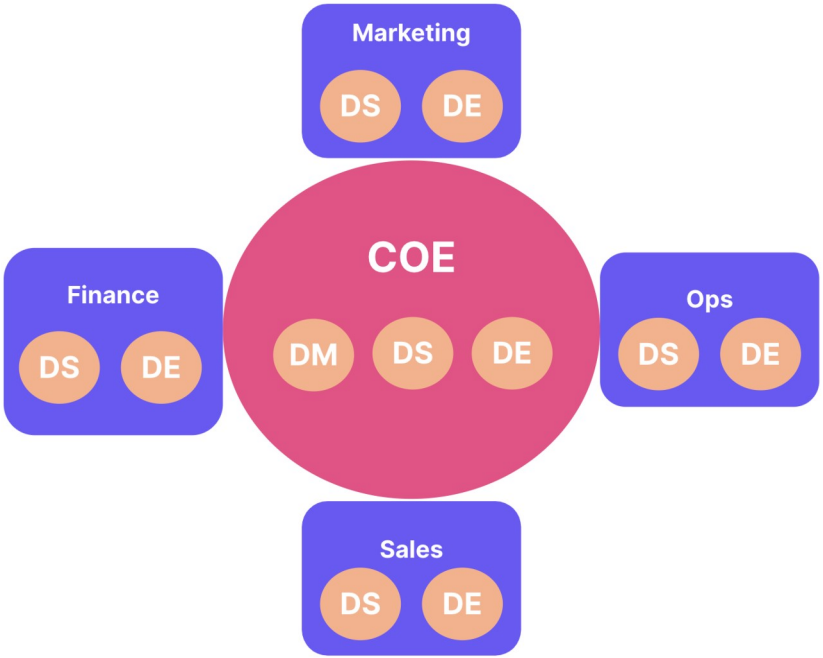
Role	Key responsibilities
Chief Data Officer (CDO)	Leads data governance strategy, policy and KPIs
Data Protection Officer (DPO)	Ensures compliance with regulations (GDPR, CCPA, etc), handles breaches, data subject access requests (DSARs)
Head of Engineering	Builds systems to support secure, governed data access
Marketing Director	Uses the data to drive marketing campaigns or get insight about the platform usage
Data Stewards	Manage data definitions, quality, access control, and metadata within domains
Legal Counsel	Advises on compliance risks and regulatory interpretation
Product & Marketing Leads	Ensure responsible use of user data and accurate analytics

Data Governance Framework Implementation

Organizational Model

Center of Excellence (CoE) Model

- Central data governance team
  - Lead by Chief Data Officer
  - Defines standard and practices for data management
  - Provides training and support to data teams
- Cross-functional governance committee
  - DPO, legal counsel, business leaders, etc
  - Oversights data management and compliance
  - Adjust governance to the needs of the company
- Embedded data teams
  - In each department (finance, marketing, etc)



Technologies and Tools

Category	Purpose	Examples
Data cataloging	Central metadata management, data discovery	Alation, Apache Atlas, Atlan, Collibra
Data quality	Automate data profiling, validation, alerts	Ataccama, Informatica Data Quality, Soda, Talend
Compliance monitoring	Manage user consent, automate DSARs	BigID, OneTrust, TrustArc, VeraSafe
Data security	Protection of sensitive user and payment data	DataGuard, Splunk, Vormetric
Lineage and integration	Track data flow across systems	Apache Atlas, Informatica

Pilot Implementation : Overview

Objective

Test the effectiveness of the proposed Data Governance Framework using the **User data domain** : personal info and interactions with the platform

Pilot Implementation : Overview

Objective

Test the effectiveness of the proposed Data Governance Framework using the **User data domain** : personal info and interactions with the platform

Why the user data?

Critical to key components of the governance framework : **compliance**, **quality** and **security**

Pilot Implementation : Overview

Objective

Test the effectiveness of the proposed Data Governance Framework using the **User data domain** : personal info and interactions with the platform

Why the user data?

Critical to key components of the governance framework : **compliance**, **quality** and **security**

Key Goals

- Improve data quality (accuracy, completeness, consistency)
- Ensure compliance with GDPR/CCPA for user data
- Streamline access to clean data for authorized staff
- Mitigate risks associated with data misuse or poor quality

Pilot Implementation : Timeline & Risks

Timeline Highlights

Milestone	Target date	Responsible
Kickoff meeting	asap	Project Manager
Data assessment and cleansing	+ 1 month	Data Steward
GDPR/CCPA compliance audit	+ 1 month	DPO
Technical setup and integration	+ 2 months	IT Engineer
Mid-project review	+ 2 months	Project Manager
Final review and closure	+ 4 months	Project Manager

Pilot Implementation : Timeline & Risks

Timeline Highlights

Milestone	Target date	Responsible
Kickoff meeting	asap	Project Manager
Data assessment and cleansing	+ 1 month	Data Steward
GDPR/CCPA compliance audit	+ 1 month	DPO
Technical setup and integration	+ 2 months	IT Engineer
Mid-project review	+ 2 months	Project Manager
Final review and closure	+ 4 months	Project Manager

Risks and Mitigations

Risk	Mitigation Strategy
Non-compliance with regulations	Weekly audits and DPO signoff
Team resistance to new processes	Workshops and clear benefits framing
Unresolved data quality issues	Documentation, regular monitoring and review
Technical integration complexity	IT involvement in early scoping

Pilot Implementation : KPIs & Deliverables

Key Performance Indicators (KPIs) : compliance, quality, performance

- 90 % service-level agreement compliance on data subject requests
- 0 security incidents involving personal data during the pilot
- 10 % reduction in missing or inaccurate user interaction logs
- < 3 % data quality error rate in user data domain
- 20 % improvement in data access time

Pilot Implementation : KPIs & Deliverables

Key Performance Indicators (KPIs) : compliance, quality, performance

- 90 % service-level agreement compliance on data subject requests
- 0 security incidents involving personal data during the pilot
- 10 % reduction in missing or inaccurate user interaction logs
- < 3 % data quality error rate in user data domain
- 20 % improvement in data access time

Expected deliverables

- **Data quality report** (metrics for missing/duplicate/inconsistent data)
- **Compliance assessment report** (GDPR/CCPA compliance of data processing practices)
- **Technical integration plan** (data availability across the department)
- **Risk assessment report** (data security and exposure, mitigatoin strategies)
- **Stakeholder feedback summary** (effectiveness of the framework)

Pilot Implementation : Next steps

- Scaling plan for the whole data infrastructure
- Adjustments based on the pilot outcomes
- Refined strategy and timeline for the implementation

Thanks!