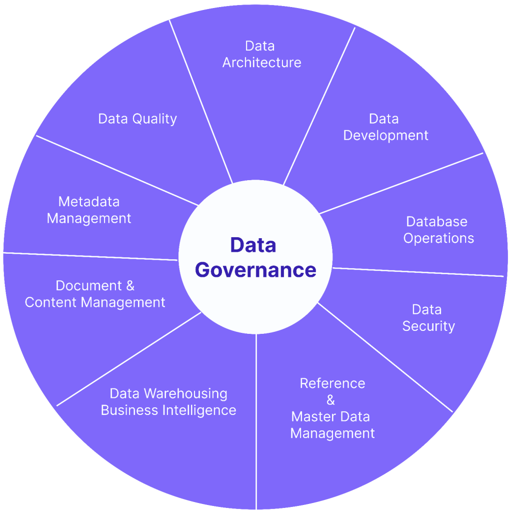
A Proposal for Spotify Data Governance

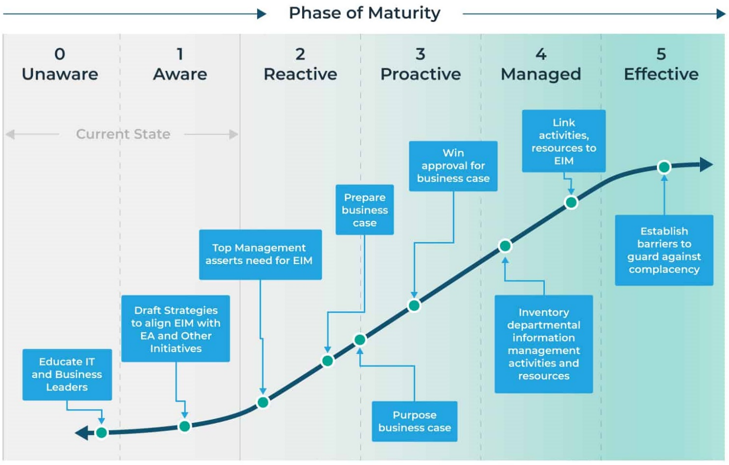
Spotify, founded in 2006 and now serving over 450 million users globally, including 200 million premium subscribers, is the world's leading audio streaming platform. Operating in 180+ countries, Spotify relies heavily on data to deliver personalized experiences, power its recommendation engine, and drive content and marketing strategies. The company manages a vast and complex data ecosystem that includes user behavioral data (e.g., listening habits, search history), content metadata (e.g., song attributes, podcast tags), subscription and billing data, advertising engagement metrics, and localized market insights. This rich data landscape is foundational to Spotify's innovation but also demands a robust governance framework to ensure data quality, compliance with global privacy regulations like GDPR and CCPA, and secure, ethical use of user information.

1 Spotify's Data Landscape

Data Maturity Assessment

Spotify must adapt its data governance to the constraints imposed by its rapid growth. The company can be considered between level 3 and 4 on Gartner's data maturity scale We provide in this section a comprehensive assessment of its maturity for each of the knowledge areas of the DAMA-DMBoK data management framework.




Phase of Maturity
0 Unaware
1 Aware
2 Reactive
3 Proactive
4 Managed
5 Effective
Current State
Educate IT and Business Leaders
Draft Strategies to align EIM with EA and Other Initiatives
Top Management asserts need for EIM
Prepare business case
Purpose business case
Win approval for business case
Inventory departmental information management activities and resources
Link activities, resources to EIM
Establish barriers to guard against complacency

- **Data Governance: level 2 - Reactive**
  No unified governance framework exists; data ownership and stewardship are informal and siloed.
- **Data Architecture: level 3 - Proactive**
  Sophisticated architecture using cloud, lakes, and databases, but lacks enterprise-wide integration and standardization.
- **Data Development: level 3 - Proactive**
  Strong ML and engineering capabilities, but development practices are fragmented with inconsistent documentation.
- **Data Operations Management: level 3 - Proactive**
  Real-time data pipelines are robust, but operational monitoring and SLAs are inconsistent across teams.

- **Data Security Management: level 4 - Managed**
  Advanced controls are in place, but evolving regulatory environments create ongoing compliance risks.
- **Reference and Master Data Management: level 2 - Reactive**
  Artist, content, and user master data are inconsistently managed across departments, leading to duplication.
- **Data Warehousing and Business Intelligence Management: level 3 - Proactive**
  Analytics infrastructure is powerful but suffers from limited cross-team visibility and fragmented key performance indicators.
- **Document and Content Management: level 2 - Reactive**
  No centralized taxonomy or governance for podcast metadata, contracts, or localized content documentation.
- **Meta-data Management: level 2 - Reactive**
  Inconsistent metadata practices and lack of a centralized data catalog hinder discoverability and quality.
- **Data Quality Management: level 2 - Reactive**
  Data quality is reactive; missing validation processes and automated cleansing across data pipelines.
- **Data Compliance: level 3 - Proactive**
  Active GDPR/CCPA efforts, but no unified compliance framework; needs structured oversight and audits.
- **Data Literacy: level 2 - Reactive**
  Data usage varies by team; limited formal training and a strong reliance on technical staff for interpretation.

Key Governance Challenges

Spotify's growth and innovation depend on turning its advanced data infrastructure into a well-governed, integrated, and trusted data ecosystem. This will require formalizing data governance policies, building roles like Data Stewards, and establishing foundational practices in metadata, quality, and literacy — all while maintaining compliance in a complex global regulatory environment.

- **Data silos and frangmentation**
  Independent data handling by departments like marketing, product, and engineering leads to inconsistent, incomplete data views and hinders unified analytics.
- **Regulatory Compliance Complexity**
  Navigating diverse regulations (GDPR, CCPA) across more than 180 countries requires a dynamic compliance strategy and stronger oversight.
- **Low metadata quality and standards**
  Metadata is inconsistently applied, and data quality issues negatively impact core systems like the recommendation engine.
- **Access and integration barriers**
  Disconnected systems delay innovation and product launches; employees lack timely access to needed data.
- **User privacy and trust**
  Increasing user awareness of data rights demands greater transparency, consent management, and anonymization protocols.
- **Weak data culture**
  While data is critical to strategy, there is limited enterprise-wide literacy, governance awareness, or data stewardship.

2 Spotify's Data Governance Framework

In this section we provide a unified approach to managing data across all functions, regions, and systems. This policy applies to all data collected, stored, processed, or shared by Spotify globally, and all Spotify teams, including engineering, marketing, content, legal, and product.

Data Governance Principles

Accountability: Clear data ownership and stewardship across all datasets. Transparency: Documented metadata, data lineage, and access logs. Compliance: All data practices must meet international and local regulatory standards. Integrity: Ensure accuracy, completeness, and reliability of all business-critical data. Security: Protect data confidentiality, availability, and integrity at every lifecycle stage. Collaboration: Encourage cross-functional sharing and integration of data.

GDPR and CCPA Compliance

Roles and Responsibilities

Table 1: Roles and Responsibilities summary table.

Role	Key Responsibilities
Chief Data Officer (CDO)	Leads data governance strategy, policies, and KPIs.
Data Protection Officer	Ensures compliance with GDPR/CCPA, handles breaches, DSARs.
Data Stewards	Manage data definitions, quality, and metadata.
Head of Engineering	Builds systems to support secure, governed data access.
Legal Counsel	Advises on compliance risks and regulatory interpretation.
Product & Marketing Leads	Ensure responsible use of user data and accurate analytics.

3 Implementation of the Data Governance Framework

Organizational Model

Given the size of the company and the amount of data to manage, the recommended approach is that of the center of excellence (CoE) model. In this model, a central data governance team, lead by the CDO, defines the standards and practices for data management. It also provides training and support to the data teams embedded in each department (e.g. marketing, product, engineering). A cross-functional governance committee of data engineers, legal councils, data protection officer and business leaders oversights data management and compliance.

In the context of this organizational model, both full-code and low/no-code solutions will be used.

- **Full-code** technologies are mandatory among the CoE team. They must also be available among the embedded teams to allow for interfacing data manipulation with the CoE team.
- **Low/No-code** technologies are key to the embedded teams in the different department. They provide high data availability for a large volume of non-tech people such as business analysts.

Technologies and tools

It is of course recommended to avoid redundancy by selecting a single tool for a given purpose.

Pilot implementation

Spotify will begin with a pilot implementation in the user data domain. The steps are:

1 Select Stakeholders: CDO, Data Stewards (Marketing, Product), DPO, Legal.

Table 2: Recommended technology tools to support Spotify's data governance framework.

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 data subject access requests (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

2 Audit Existing Data: Identify sources, access logs, quality gaps.

3 Apply Governance Rules: Define metadata, access controls, validation checks.

4 Configure Tools: Deploy data catalog and privacy software for pilot scope.

5 Train Teams: Deliver sessions on new policies, roles, and tools.

6 Monitor Metrics: % of high-quality user records, of DSARs handled on time, user trust/engagement metrics

7 Review & Iterate: Adjust processes before scaling to other domains.

The pilot implementation will have the following target performance indicators:

- 90% service-level agreement compliance on data subject requests
- 80% data literacy training adoption in pilot teams
- less than 3% data quality error rate in governed domains