

Spotify's Data Governance Framework Implementation

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 oversees 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

Table 1: 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 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

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 rationale behind this choice is that user data is central to key components of the governance plan: compliance, quality and security. The steps of this pilot implementation are given below. They should be completed in 3-6 months.

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