POLYMER

7-Step Data Governance Guide

Playbook for IT & Information Security Professionals

What you will learn:

- Scoping Your Program
- Critical Data Elements
- Data Maps
- Data Ownership
- Privacy By Design
- Third-Party Risk
- Data Loss Prevention

What is Data Governance?

Data governance is a framework to define, monitor and implement appropriate usage and disposal of sensitive data within an organization. Any company collecting, processing or storing data requires accountability and a framework to manage the data from acquisition to disposal. Data governance function requires different stakeholders from across the organization to adopt roles, responsibilities and policies for using and protecting PII/PHI or company-specific confidential data assets.

Data governance can look and feel different from one organization to another. Each business has its own unique data flows and tech stacks that require slightly different approaches to how data is protected.

Whether you are a one-person compliance team or part of a data program in a large organization, there are a set of common themes across any data governance implementation. In this article, we look at the principles and framework of implementing a successful data governance program.

1. Scope: Define Your Data Governance Principles

Data governance can be a hazy concept for a large cross section of any company. Defining and publishing data governance goals sets the foundation within any organization to think of the positive outcomes that can be achieved. Examples:

- Limit unencrypted data sharing within the organization
- Disallow public sharing of files via OneDrive or Google Drive

Some of these scopes might be too draconian and will require business input. However, getting the conversation started on these goals is essential to define success criteria. The scope, breadth and depth of any data governance program is dependent on resources, manpower and budget. For example, a one-person data governance team can structure a program based on acknowledgments and acceptance criteria from business, while a larger team can probably implement a fully functional real time monitoring and audit system.

2. Definition: What is Sensitive Data for Your Business?

What you consider sensitive data can vary based on your business industry. However, a good starting point is to bucket up broad data grouping. The most common data groups are:

- Customer Data
- Transaction Data
- Patient Data
- Employee Data

Data groupings are essential to build an organization-level lingua franca where not everyone is necessarily technical or data savvy.

3. Data Mapping: Critical Data Elements of Your Business

Data mapping is a core construct in implementing a data governance program.If you don't know the critical data elements of your business, you don't know what you risk losing. Data mapping refers to categorization and linking data categories to actual data found in systems or databases. There are two levels of data mapping:

Logical Data Elements: Refers to a grouping of data elements with a business connotation.

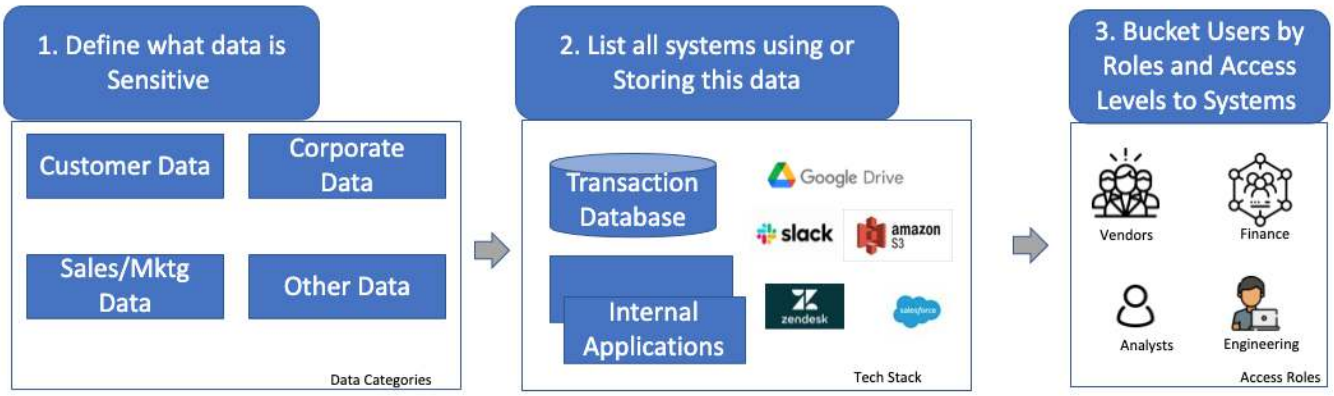
Customer Data→ Customer Name, Customer Address, Customer SSN

Physical Data Elements: Physical manifestation of the data elements like name or credit card data found as database columns or within documents.

Customer Name→ Postgres1.people_table.{customer_name}
Customer Address→Snowflake1.analytics_db.{customer_name}
Customer SSN→ GoogleDrive.contracts_folder.{acme_corp_signup.pdf}
Some broad categories of data are:

- Customer Data
- Corporate (Internal) Data

- Sales/Marketing/Operations Data



Applying these labels and constructing a mapping of the Logical→ Physical data elements across all data within a company's environment might be the hardest part of implementing any governance program.

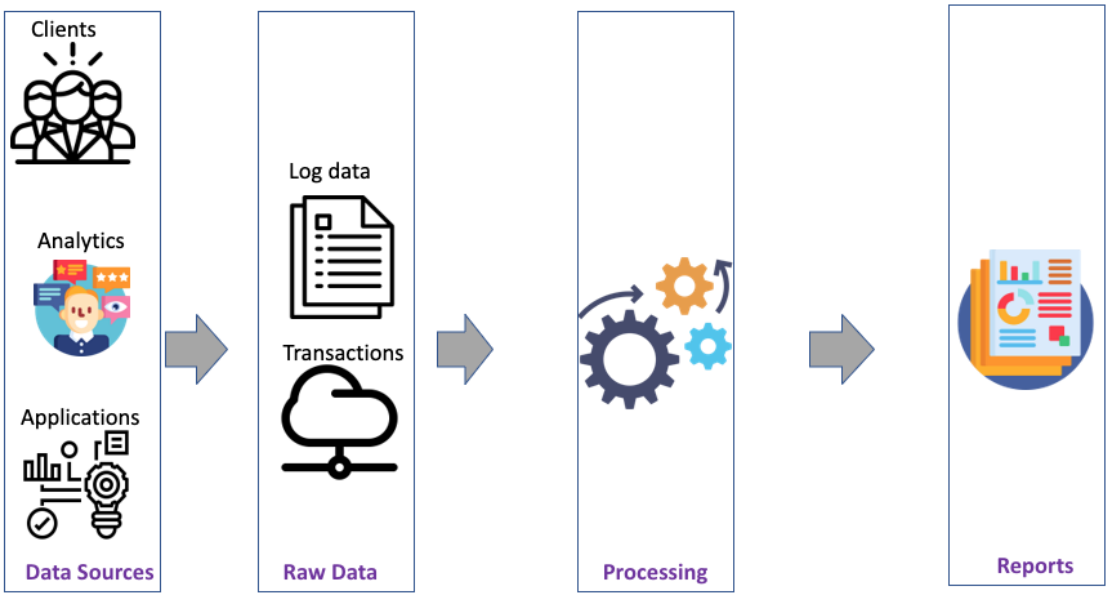
Data Quality: For organizations with a well-defined data understanding, knowing the accuracy of the data is also an important consideration when considering what physical data elements to use.

Keep the data categories narrow and start by asking the engineering teams about the inventory in the most-used systems and databases. This can reduce the amount of mappings required and help save engineering time.

4. Data Ownership: Who Owns Sensitive Data?

It is important to define who is in charge of protecting and maintaining data within an organization to have a data governance program that can be maintained and be effective.

Data ownership structure can be i) federated or ii) centralized. Both have their pros and cons, but in general for fast-growing companies, federated models where each business unit/area is in charge of



data under their preview is best. The centralized approach might be appealing for large organizations and/or slower changing tech stacks.

As multiple data inputs get transformed and reported on, who has the responsibility of keeping this data secure?

A best of both worlds might be a data ownership format in a dual-ownership structure where business and technical teams each own their assets. Example of how this ownership structure might break down:

- Business
    - Finance
    - Marketing
    - Operations
- Technical
    - Database Operators
    - Engineers

Of course there are challenges to define this ownership structure when data is moving from one system to another or is changing in form. Demarcation of where data is found

gets tricky when you have states of data that are 'in-between.'.For example, log data from applications and ETLs might be an output which has business or technical ownership.

5. Privacy By Design: Operational Data Controls

Privacy by design is a simple yet powerful concept to bring the privacy conversation to the forefront of any new product release, vendor engagement or operational changes. Basically, you make the business unit responsible for considering data controls with a 5-10 point privacy questionnaire that must be filled out with any new initiative. Based on the responses to the privacy questionnaire, it might trigger a review where legal and engineering along with the product team advise on the privacy implications of the new initiative.

Following are three simple questions to start your privacy by design journey:

1. **What type of data is being used in this new initiative?**
   Any IT or information security project needs to start with an understanding of the data that will be consumed, transformed and outputted. This data analysis can also provide a data sensitivity rating on a per project basis.

2. **What is the retention policy of this data?**
   Storing information for auditing purposes is the norm for financial services. However, in other industries, there should be a well laid-out reason for storing data beyond a reasonable amount of time. More raw data stored only increases the risk of theft or exfiltration and can be a big liability. Reduce the data retention period in your business if you can.

3. **What infrastructure will be leveraged (or needs to be built) to store and protect this data?**
   As part of your processes, understanding key servers, databases, tables, storage locations and SaaS systems where sensitive data is stored is important to flag high-sensitive-traffic workflows.

Security Assessments (SOC2, HIPAA, ISO27001)

Security self-assessments can be misleading in explaining the rigidity of security controls at the vendor or partner side. These controls are non-uniform and do not tell the full story of security risks. Having evidence of controls as mentioned above is absolutely essential. A good data governance documentation should be used to cover a decent chunk of the data controls sections of most assessments.

Software Development Lifecycle

Adding a Privacy by Design checklist to any product or major software release can align technology with the data principles. This is a very low-touch method to bring engineering processes under the purview of data controls. Shifting security principles left, closer to development, also makes compliance more automated.

6. Third Party Vendor Evaluation

Contracts with vendors or third parties is a critical data handoff that requires special attention. A thorough look at and evaluation of data governance programs with vendors and partners is necessary in an era of Solar Winds where supply-chain attacks can put any data sharing mechanisms at risk.

Areas of focus are:

1. **Verification of data controls:** As mentioned in the SOC 2 or ISO report provided by the vendor *("Is the vendor doing what they say they are doing in their SOC self-assessment?")*
    a. **Effect of data protection:** Practices related to HIPAA, GDPR, CCPA or other regulations if customer or patient data is being handled at all
2. **Analyzing any new relationship:** Considering that if the counterparty is breached what risk does it pose on the company

Asking the tough questions up front is the best policy. There are some basic security and technical hygiene requirements that are no-brainers when purchasing a product. A vendor product should have:

1. SAML Connectivity
2. MFA
3. SOC 2
4. Data Processing Agreement

A vendor assessment might also include a privacy assessment that can be customized depending on what kinds of data is being shared/collected. A vendor/partner accessing customer data will have a higher hurdle to pass in this evaluation. DPA agreement or sub-processor agreements are also very important when defining the terms of engagement, especially within the EU.

7. Data Security: Data Protection on SaaS & Infrastructure

An increasing amount of cloud-based workflows is creating a data governance blind spot. Cloud applications have opened new methods and channels of data misuse. Data governance principles allow information security to be effectively applied and measured. Examples of this might be:
i) Data maps can highlight the most important assets to protect
ii) Misuse of sensitive data can be measured from the perspective of the data owner vs data subject
iii) Effectiveness of data protection can be quantified with the egress traffic from various endpoints

Role of Compliance and/or Legal

Small legal, compliance and IT teams responsible for data governance need to assign one or two individuals in each area of business. This "data point person" can sit within engineering, product, human resources, security or IT.

A regular dialogue between compliance, IT, InfoSec teams and the "data point person(s)" is important to verify efficacy of controls and progress towards success criteria.

Conclusion

A good data governance posture reduces your company's risk against cyber attacks. Scaling organizations with a secure data foundation is only possible when everyone in the organization is accountable for data. With upcoming global and state level privacy regulations, data governance will become table stakes when it comes to attracting customers and working with partners. The ideas laid out in this guide should lead to net positive financial outcome in the long run.

If you are looking for more information on the topics covered in this guide or want to learn more about Polymer, please reach out to Avery Neims at aneims@polymerhq.io or visit Polymer's website https://www.polymerhq.io.

References

Polymer DLP for SaaS Blog has a rich selection of articles focused on security, governance, legal, HIPAA, SaaS data protection and other related data governance topics.

About Polymer

Polymer DLP for SaaS is a data loss prevention platform for cloud hosted applications. Polymer jump starts a data governance program in one click to protect data over collaboration platforms like Slack, OneDrive, Google Drive, Teams, Github, Jira and others. The platform embodies the 7 Step Data Governance highlighted here to allow monitoring, effectiveness, risk measurement and data protection within cloud hosted applications.