Whitepaper

How to Build a Sustainable Data Minimization Strategy

by **Ritu Khanna**

Table of Contents

Abstract

With data becoming the new currency, the natural tendency has been to accumulate vast amounts of consumer data by one and all. This proliferation has made it possible to multiply the use cases of data and bring speed in converting hypothesis to predictions and forecasting.

Excited by the proposition of being able to grow and create new demand purely on the basis of data fueled hunger for data. And the world slipped smoothly into a circle of more data and more possibilities.

But the other side of this data strategy is fraught with risk of accuracy, integrity, and breach of sensitive consumer information. Do you know the degree to which consumer data held in your organization is stale, incorrect, and unused? This can lead to increase in threat surface area, poor data process output, longer and inadequate response to privacy requests, not to mention storage costs for data that should not be in the system in the first place.

Data Minimization

Most regulatory enforcements and frameworks in areas of privacy have evolved from the OECD principles, which highlight importance of privacy and data minimization. As per GDPR regulation and specified in ICO Article 5 (1) (c) data minimization principle says, “Personal data shall be: (c) adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed (data minimization).”

This means organizations should identify the minimum amount of personal data needed to fulfil business purpose. Business should hold that much information, and no more.

Furthermore, ICO also highlights below areas-

- **What is adequate, relevant, and limited** – Though these are not defined under GDPR, it is important for organizations to be clear as to what and why the data is being collected. If it involves storage of data to abide by surveillance laws in the country or national security, organizations need to secure and segregate that portion of data too.

- **Too much personal data** – Most organizations require additional data for analytical purposes or to monetize on the collected data. It's important to understand all aspects of data processing and collection before storing and collecting any additional data, which differs from the current business purpose or lawful basis of processing.

- **Checklist** - ICO also provides certain questions which help organizations identify if they are collecting and storing what data is necessary and discarding the remaining data in secure manner.

Regulation Impact

As stated above, GDPR is one of the regulations that explicitly highlights data minimization. There are several other regulations which stress the importance of data minimization across different sectoral (PCI, DSS), federal (HIPAA, GLBA), and state laws (CCPA, MA, NYPD).

Consequences of non-compliance

In 2018, Denmark's data protection authority, performed an audit of a taxi company. Authorities found that the company had implemented a data retention policy, but had failed to follow it. Furthermore, it was observed that personal data relating to about 9 million individual taxi rides was being preserved beyond the lawful two-year retention policy. The company erased the name and address of each customer, but retained their phone number. It claimed that the phone number was used as an "account number" and so the company had a legitimate purpose to retain it. Admittedly, the phone number itself was not required — an anonymized number would fulfill the purpose. However, its computer systems were unable to convert the phone number into a new unique ID that would not be classified as personal data. The Danish DPA stated, "[O]ne cannot set a deletion deadline, which is three years longer than necessary simply because the company's system makes it difficult to comply with the rules." In March, the Danish DPA fined the taxi company 1.2 million kroner (USD 180,000), its first fine under the GDPR.

In its ruling, the Danish DPA found that the taxi company had violated Article 5 of the GDPR in three ways: purpose limitation, data minimization, and storage limitation.

Article 5(1)(c) requires personal data be adequate, relevant, and limited to what is necessary in relation to the purposes for which it is processed. The company argued that it had met minimization requirements by removing the names associated with the phone numbers. But its systems were not capable of transferring the anonymous data about the taxi ride from a phone number to a unique ID. The Danish DPA stated, in no uncertain terms, that costs associated with migrating personal data to a new anonymous data structure do not justify continued use of the personal data like the phone number in this case beyond the retention policy.

Data Minimization: An Inside View

Let's take a closer look at the what is data minimization all about and how can organizations deliver on this concept. Let's take an example of insurance life cycle and associated data value chain. Data and information are collected about the customer while applying for insurance and a sample can include:

- **Health information:** pre-existing diseases, blood group, blood pressure, heart condition, other health vitals.
- **Lifestyle habits:** smoking, drinking, fitness, etc.
- **Financial:** credit score, annual income, net worth.
- Employment details, education details.
- **Nominee and family data:** age, relationship, income, etc.

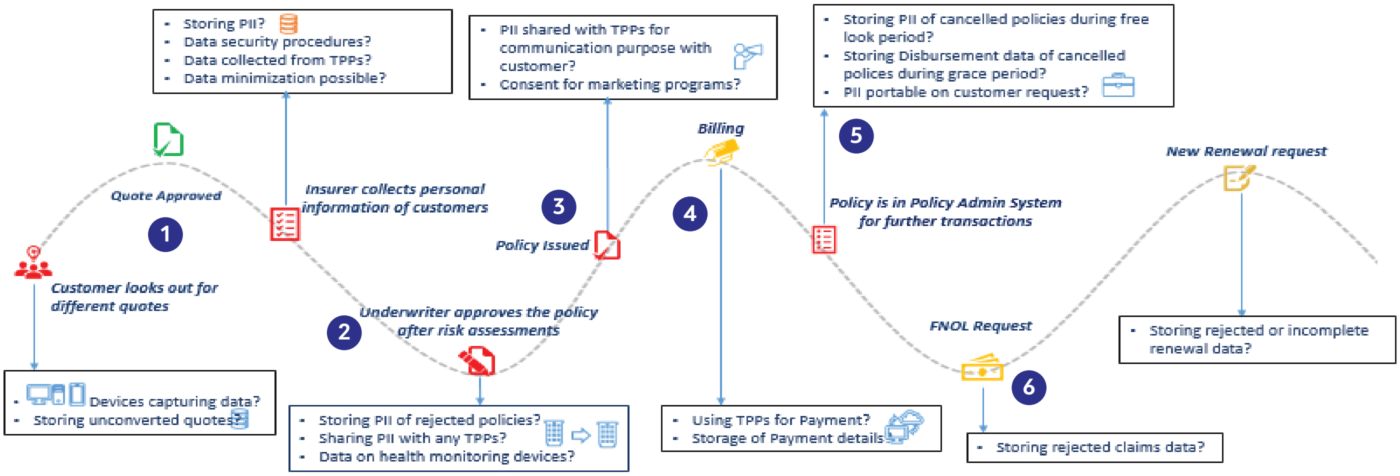
While all the data collected is important for underwriting, issuing policies, and servicing customer claims and renewals etc., for all applications that get rejected, there is a huge amount of data that is now in the system belonging to individuals who are not customers of the organization. Similarly, for unconverted quotes, there is huge amount of sensitive personal data in the system.

All this personal information poses a huge data risk while serving little or no business use. Typically, since this data is not useful in business transactions, applications which hold this data do not receive regular housekeeping, thus becoming vulnerable to internal and external breach.

Apart from data collected directly from the individual, there are other kinds of data that is collected from third-parties about the individual, or data generated based on business models, which provide some form of business decision making about the individual, thus exploding the data that gets accrued for individuals. One of the most common woes of an IT director is the lack of visibility they have on all data that exists in their system. Because of its complex nature, there is no tracking of where the data resides, how it flows through the organization, who has access to it, and where does it meet its end of life. While it is intuitive that a data catalog and inventory is the starting point of any form of data management, it is still a holy grail, and very few tools have been able to get it right.

Opportunities to minimize data

Sample data life cycle: Insurance



Storing PII?
Data security procedures?
Data collected from TPPs?
Data minimization possible?
PII shared with TPPs for communication purpose with customer?
Consent for marketing programs?
Storing PII of cancelled policies during free look period?
Storing Disbursement data of cancelled polices during grace period?
PII portable on customer request?
Billing
5
New Renewal request
Quote Approved
1
Insurer collects personal information of customers
3
4
Policy is in Policy Admin System for further transactions
Policy Issued
Customer looks out for different quotes
Underwriter approves the policy after risk assessments
2
FNOL Request
6
Storing rejected or incomplete renewal data?
Devices capturing data?
Storing unconverted quotes?
Storing PII of rejected policies?
Sharing PII with any TPPs?
Data on health monitoring devices?
Using TPPs for Payment?
Storage of Payment details
Storing rejected claims data?

1 **Privacy Notices and Consent**
Collect only what is required

2 **Data Discovery and Cataloguing**
Identify and classify data

3 **Third Party Assurance**
Audit & evaluate compliance risks

4 **Record Keeping (ROPAs)**
Process only what is required

5 **Data Protection**
Anonymization & Restricted Access

6 **Data Decommissioning**
Discarding stale & redundant data

Case Study

Data deletion, depersonalization, and minimization can be achieved using multiple approaches.

Example of data deletion using a Policies-based systematic approach

Company ABC Corp. has operations across different European countries. They have a large number of contractual and temporary staff. The organization has more than 200 IT systems which have data on employees. Some of this data is collected from individual, some from third-parties, and some of it is generated while in employment, like performance data. All this data is residing in different structured and unstructured data sources. For example, data in data lakes, SQL, AWS web services, etc.

The data deletion policy for each data set varies by country, which makes data categorization (what data is available, which systems they are in, business definitions of data; e.g. benefits, performance data) a critical aspect of implementation. They are different applications across technology landscapes, which pose a challenge to maintain data quality and consistency with implemented policies. For example, retention policies of data would differ based on different regulations and would need to address on preceding guidelines for retention and deletion.

Steps taken to implement the framework and adhere to minimization guidelines -

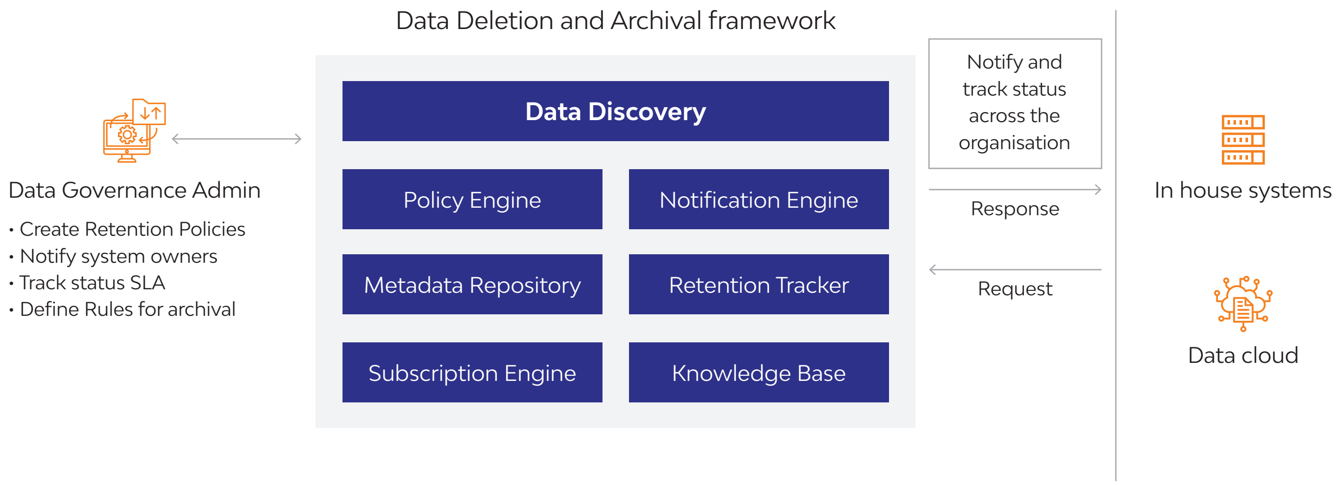
- Develop a metadata repository based on the data schema and categories that could be optimized for future downstream reporting.

- Define a schedule where the identified Event Triggers would scan the data in data warehouse at defined intervals, raise an event for processing. Events can include termination of employment, transfer outs, end of retention periods, etc.

- Develop a policy warehouse and the set of rules that defined the hierarchy between policies based upon the nature of events, and kind of data.

So, in summary, organizations can develop a system to work through various policy categories, build rules around making deletion decisions, and embed the privacy regulations that impact the data. Building a technology solution helps a company scale up a standard process across all their operating countries globally, bring in a centralized view and control of how data deletion will be dealt with in a compliant way. The diagram below displays how the end-to-end data minimization works:

Approach towards data minimization



Data Deletion and Archival framework

Data Discovery

Policy Engine

Notification Engine

Metadata Repository

Retention Tracker

Subscription Engine

Knowledge Base

Data Governance Admin

- Create Retention Policies
- Notify system owners
- Track status SLA
- Define Rules for archival

Notify and track status across the organisation

Response

Request

In house systems

Data cloud

- Data Discovery Scanner - To discover, classify, and catalogue data
- Policy Engine – Repository of all data retention policies
- Metadata Repository - List of record identifiers that need to be tracked and acted on
- Subscription Engine - Subscriptions to track and identify applicable policies
- Knowledge Base – For creating automated data classification and tagging
- Notification Engine - For Data Deletion/Anonymization to meet regulatory guidelines

Most businesses are adopting an automation-first approach to consumer rights management solution to address the nature of compliance, the SLAs, and the impact of consumer requests on application and data spread within the company. Different approaches exist to address fulfilment of the request to 'forget me' or the data deletion request. Processes to de-identify the personal data belonging to the consumer by anonymizing the data has worked well to fulfil the consumer request.

Solving for event-based data deletion requires an automated solution that includes receipt of consumer request, consumer identity and request validation, identification of impacted applications, discovering personal data, and applying anonymization or deletion rules to the data.

Moreover, the solution should provide SLA tracking, audit traceability and extensive reporting and analysis of received requests. Based on the examples discussed, the key components that must come together to implement an enterprise-wide data deletion and archival framework are as follows:

- **Data discovery** – knowing what data you have, how much of personal and sensitive data you collect and generate, and where does this data reside.
- **Metadata repository** – business-wise metadata information laying the foundation of information life cycle management from data collection to disposal.
- **Rules-based policy engine** – develop a hierarchy-based decision engine bringing together regulatory, industry, and organization policy to define the rule of data deletion and archival.
- **Event notification** – identify roles and actors in the process, automate workflow, and notify events and changes.
- **Additional considerations** – automation opportunities using intuitive decision making, enable an easy-to-implement hub and spoke model for application subscription throughout the organization, resulting in a scalable solution to address data minimization.

Conclusion

In summary, organizations are beginning to take **data minimization** seriously and understand that addressing this area has multiple benefits beyond meeting compliance. Building for minimizing data meets the requirement of reducing data risk and the impact of a data breach or fraud. It reduces the cost of storing and managing data, which is of no use to the company.

LTIMindtree has niche solutions for privacy rights management. These solutions are intuitive, address global regulatory rights, have an end-to-end reach from consumer to application owner, and are ready to integrate with other technologies in use in the business.

LTIMindtree has helped clients solve data minimization & anonymization use cases through its data minimization solution. Global health insurer, eCommerce and cloud giant, largest technology company are some of the clients who have implemented LTIMindtree's data minimization solution addressing their business needs.

Author Profile

Ritu Khanna

Data Privacy Practice Head, LTIMindtree

Ritu leads the Data Privacy practice at LTIMindtree, working with clients to adopt "privacy by design" in their data management and governance operations. Ritu's global experience spans across consulting, technology, and operations across BFSI, Fintech, Consumer, and Life Sciences.