Data Governance

Data governance plays a crucial role in identifying the right people who need access to the correct data at the right time. It helps us deal with multiple issues such as Multiple versions of the truth, No Data Owners, lack of meaning or context of data, and absence of structured documentation. We plan to address some of these issues in our project's data governance plan. We plan to follow the framework below:

1. **Discovery:** Identifying and understanding the different types of data stored. In our case, we have ticketing, plane maintenance and baggage tracking details. Identifying the type of data, we deal with will make classifying them easier.

2. **Classification:** The process of understanding what type each data element belongs to and categorizing them accordingly.
    - Business Terms
    - Data Classes

    For example, details such as the baggage, last location or current location should be primarily stored in the baggage tracking OLTP system. Similarly, details regarding reservation, customer age, and other such information should be stored in the ticketing system.

3. **Policy:** Set of guidelines to manage data effectively. Every company needs to follow regulations such as GDPR, CCPA, HIPAA, ePrivacy, etc. These guidelines can be observed during the different phases:
    - **Data capture**: Capture minimum data, just what is necessary—consent management, permissions to allow a user to either opt-in or opt-out. Channel preference, ask if the user needs any communication from the organization via text or email.
    - **Data storage and usage:** Data sensitivity or PII (Personally identifiable information) fields such as SSN should be identified and stored more securely.
        - Identify or detect
        - Flag/Classify or label
        - Mask or Pseudonymize
    - **Data Retention**: retain data only if it is needed. This needs justification from the organization to prove why they might have to store certain information for a given period.

    We intend to move ahead with privacy by design policy. Therefore, we make sure every step aligns perfectly with the regulations.

4. **Rules**
    - **Data Owner/Sponsor:** Individual who has ultimate accountability for the data to be current and UpToDate. Usually, these roles are assigned to higher authority members. Depending on the organization's size, the team can have one data owner for sale and another for marketing data. For Southwest, we want dedicated data owners for marketing, sales, business, etc.
    - **Data Stewards**: These roles are assigned to people with a more hands-on approach to the data. They have a complete understanding of the contents of the data. They also are referred to as subject matter experts or Data Champions. These roles would be assigned to developers at southwest data team.

5. **Document Data Sources** – help create a knowledge repository, trace back requirements and quality assurance.
    - **Data Mapping**

      For example, ticketID in the ticket table can be linked with ticketID in the baggage table. These unique IDs help us identify each ticket separately. We will have an entry in Business name and definition called Passenger Name record; this would link the ticketID fields in both the ticket and baggage Feed/Table.

    - **Data Cataloging**

      Identifying key or critical data elements such as credit card details, social security numbers, or ticket prices to perform particular operations. These fields must be recorded and reported to a regulator.

    - **Metadata**

      A catalog that helps us describe 'what/where/how' of data.

      For example, Flight ticket price.

      Technical metadata – Data model, schemas, Data lineage and transformation.

      Data Type: Float

      Column: Ticket_amount

      Table: Customer_Ticket

      Schema: Ticketing

      Server: Ticketing.southwest.net

      Business Metadata – Glossary, Business Definition or rules and domain

      Data Domain: Ticketing

      Line of Business: Airline

      Business Name: Flight ticket amount

      Business Definition: Price amount of the ticket. This could vary depending on the class of flight or date of purchase.

      Operation Metadata – DQ metrics and runtime data

      Null Check

      Negative amount check

6. **Data Lineage:** Tracing or tracking the movement of data along with recorded the history of data. This process can be automated. Any new data can be tested with data lineage to ensure it doesn't impact the existing flows. It can help us detect underlying problems and provide transparency. However, with complex systems, this may get more challenging. Stakeholders may see this as a regulatory system rather than benefits beyond it. This is a task for us as data managers to convince the stakeholders to invest in data lineage.