Legacy system table - SAP sales and distribution

This table represents the data in a legacy system before validation and reconciliation.

Patient ID	Name	Age	Phone Number	Date of Birth	Address
001	John Doe	30	123-456-7890	01/01/1993	123 Main St
002	Jane Smith	28	234-567-8901	02/20/1995	456 Park Ave
003	John Doe	30	123-456-7890	01/01/1993	789 Maple Rd
004	Mary J.	200		03/10/1950	101 Pine St
005	Jane Smith	28	234-567-8901	02/20/1995	456 Park Avenue
006	Alex Johnson	25	345-678-9012	05/25/1998	89 Oak Dr

ERP - Salesforce

This table represents the data in the ERP system after initial data entry (potentially with some errors).

Patient ID	Name	Age	Phone Number	Date of Birth	Address
001	John Doe	30	123-456-7890	01/01/1993	123 Main Street
002	Jane Smith	28	234-567-8901	02/20/1995	456 Park Ave
003	John D.	30	123-456-7890	01/01/1993	789 Maple Rd
004	Mary Johnson	98	345-678-9020	03/10/1950	101 Pine St
005	Jane Smith	28	2345678901	02/20/1995	456 Park Avenue
006	Alex J. Johnson	25	345-678-9012	3/25/1998	91 Oak Drive

Data Validation Process

Step 1

Assessment of Current Data Quality

Analyze the current data in both systems.

Example

Identify duplicates: Two entries for John Doe and Jane Smith.

Find inconsistencies in date formats and missing phone numbers.

Step 2

Define Data Governance Framework

Action: Establish standards for data entry.

Example: Create a table outlining the required formats.

Field	Standard Format	Example
Name	First Last	John Doe
Age	Numeric (0-120)	30
Phone Number	XXX-XXX-XXXX	123-456-7890
Date of Birth	MM/DD/YYYY	01/15/1993
Address	Valid US Address (Street Style)	123 Main St

Stage 3

Data Validation Rules by Field

Legacy System: Requires **one or more uppercase letters** (A-Z+): This means that multiple uppercase letters at the beginning of either name is acceptable, making names like "JOHN DOE" valid.

The ERP system expression uses a **NOT** function, which implies that it is checking for names that do **not** match the given regex pattern. This means the validation logic is inverted, which could result in an error if a name does match the format.

Field	Validation Rule & Example	Legacy Implementation	ERP System Implementation
Name	Must contain first and last names, only letters/spaces	Regex: `^[A-Z]+[A-Za-z]+$`	Formula: `NOT(REGEX(Name, "^[A-Z]+$"))`
Age	Must be numeric, between 0 - 120	If check: `If age < 0 or age > 120`	Formula: `Age < 0`
Phone Number	Must follow format XXX-XXX-XXXX	Regex: `^\d{3}-\d{3}-\d{4}$`	Formula: `NOT(REGEX(Phone, "^\d{3}-\d{3}-\d{4}$"))`
Date of Birth	Must be a valid past date	If check with datetime comparison	Formula: `DOB > TODAY()`
Address	Must not be empty	Simple string check	Formula: `ISBLANK(Address)`

ERP System: Requires **exactly one uppercase letter** (A-Z): This means that names like "JOHN DOE" would be invalid in the ERP system.

Stage 4

Regular Data Reconciliation

Action: Regularly compare the data in both systems.

Example: Compare the patient ID and name fields between both tables.

Patient ID	Legacy System Record	ERP Record	Status
001	John Doe	John Doe	Match
002	Jane Smith	Jane Smith	Match
003	John Doe	John D.	Mismatch
004	Mary J.	Mary Johnson	Mismatch
005	Jane Smith	Jane Smith	Match
006	Alex Johnson	Alex J. Johnson	Match

Stage 5

Data Cleansing Initiatives

Action: Clean the data based on validation results.

Example:

Consolidate entries for John Doe, correcting the name and address. Fix Mary J. to Mary Johnson.

Stage 6

Enhance User Training and Documentation

Action: Train staff on the new data entry processes.

Example: Develop a training session focusing on how to correctly enter patient details based on the governance framework.

Stage 7

Deployment of Monitoring Tools

Action: Set up dashboards to track data quality.

Example: Create a dashboard that shows the percentage of missing phone numbers and the number of duplicates over time.

Metric	Count
Missing Phone Numbers	1
Duplicates Found	2
Invalid Ages	1

Stage 8

Plan for Regular Updates

Action: Schedule periodic reviews and updates to data processes.

Example: Conduct quarterly reviews of data quality and make necessary policy adjustments based on employee feedback.

Stage 9

Leverage Technology

Action: Utilize software tools for data management.

Example: Implement tools that automatically highlight duplicates within both the legacy system and ERP.