

Data Cleansing Checklist for Legacy Systems

1. Data Assessment

- Identify all data sources in the legacy systems
- Assess data quality and completeness
- Categorize data by type and sensitivity
- Document known data issues

2. Data Profiling

- Analyze record counts and structure
- Detect and document data inconsistencies
- Check for duplicate records
- Review metadata definitions

3. Data Standardization

- Define standard formats for key data fields
- Normalize values for date, number, and text formats
- Standardize naming conventions across datasets
- Apply uniform units of measurement

4. Data Cleansing

- Remove duplicate entries
- Eliminate obsolete or deprecated data fields
- Correct data entry errors and typos
- Fill missing required values where possible
- Flag or remove invalid records

5. Data Validation

- Schedule validation rules for key data fields
- Cross-check with authoritative sources
- Review outliers and edge case values
- Perform user/stakeholder review where needed

6. Documentation & Review

- Document all cleansing steps and changes
- Preserve original data backup
- Share results with project stakeholders
- Obtain sign-off on cleansed dataset