Data Quality Assessment Template

Comprehensive AI Implementation Guide

EXPANDIA.CH

Contact: hello@expandia.ch

DATA QUALITY ASSESSMENT TEMPLATE

Comprehensive Framework for Evaluating Data Quality

Executive Summary

This template provides a systematic approach to assess data quality across six key dimensions: Accuracy, Completeness, Consistency, Timeliness, Validity, and Uniqueness. Use this framework to establish baseline data quality metrics and identify improvement opportunities.

Data Quality Dimensions

1. Accuracy Assessment

Definition: Data correctly represents real-world entities or events

Measurement Criteria: - [ ] **Syntactic Accuracy:** Data conforms to defined formats and patterns - [ ] **Semantic Accuracy:** Data values are correct and meaningful - [ ] **Reference Data Validation:** Data matches authoritative sources

Assessment Questions: - Are addresses valid and deliverable? - Do phone numbers follow correct formats? - Are email addresses syntactically correct? - Do customer names match official records? - Are product codes valid in the system?

Metrics to Track: - Accuracy Rate = (Accurate Records / Total Records) × 100 - Error Rate = (Records with Errors / Total Records) × 100 - Validation Success Rate = (Records Passing Validation / Total Records) × 100

Target Thresholds: - ■ Excellent: >98% accuracy - ■■ Good: 95-98% accuracy - ■ Poor: <95% accuracy

2. Completeness Assessment

Definition: All required data elements are present

Measurement Criteria: - [ ] **Field Completeness:** Individual fields are populated - [ ] **Record Completeness:** Complete records exist for entities - [ ] **Dataset Completeness:** All expected records are present

Assessment Questions: - Are mandatory fields populated? - Are there missing customer records? - Is historical data complete for the required time period? - Are all product attributes present? - Do we have complete transaction histories?

Metrics to Track: - Completeness Rate = (Complete Records / Expected Records) × 100 - Missing Value Rate = (Fields with Missing Values / Total Fields) × 100 - Record Coverage = (Existing Records / Expected Records) × 100

Target Thresholds: - ■ Excellent: >95% completeness - ■■ Good: 90-95% completeness - ■ Poor: <90% completeness

3. Consistency Assessment

Definition: Data is uniform across different systems and datasets

Measurement Criteria: - [ ] **Format Consistency:** Same data types and formats used - [ ] **Value Consistency:** Same values represented identically - [ ] **Cross-System Consistency:** Data matches across systems

Assessment Questions: - Are date formats consistent across systems? - Do customer records match between CRM and billing systems? - Are product names standardized? - Do financial figures reconcile across systems? - Are business rules applied consistently?

Metrics to Track: - Consistency Rate = (Consistent Records / Total Records) × 100 - Format Standardization = (Records Following Standard Format / Total Records) × 100 - Cross-System Match Rate = (Matching Records / Total Records) × 100

Target Thresholds: - ■ Excellent: >98% consistency - ■■ Good: 95-98% consistency - ■ Poor: <95% consistency

4. Timeliness Assessment

Definition: Data is current and available when needed

Measurement Criteria: - [ ] **Currency:** Data reflects the current state - [ ] **Freshness:** Data is recently updated - [ ] **Availability:** Data is accessible when required

Assessment Questions: - How current is the customer contact information? - Are inventory levels updated in real-time? - Is financial data available for timely reporting? - Are market prices current? - How quickly is new data processed and available?

Metrics to Track: - Data Age = Current Time - Last Update Time - Update Frequency = Number of Updates / Time Period - Availability Rate = (Time Data Available / Total Time) × 100

Target Thresholds: - ■ Excellent: <1 hour data age for critical data - ■■ Good: 1-24 hours data age - ■ Poor: >24 hours data age

5. Validity Assessment

Definition: Data conforms to defined business rules and constraints

Measurement Criteria: - [ ] **Domain Validity:** Values fall within acceptable ranges - [ ] **Format Validity:** Data follows required patterns - [ ] **Business Rule Validity:** Data satisfies business constraints

Assessment Questions: - Are numeric values within expected ranges? - Do categorical values match predefined lists? - Are relationships between data elements valid? - Do calculated fields produce correct results? - Are business rules properly enforced?

Metrics to Track: - Validity Rate = (Valid Records / Total Records) × 100 - Rule Violation Rate = (Records Violating Rules / Total Records) × 100 - Constraint Compliance = (Records Meeting Constraints / Total Records) × 100

Target Thresholds: - ■ Excellent: >99% validity - ■■ Good: 97-99% validity - ■ Poor: <97% validity

6. Uniqueness Assessment

Definition: No duplicate or redundant records exist

Measurement Criteria: - [ ] **Record Uniqueness:** Each entity represented once - [ ] **Field Uniqueness:** Unique identifiers are truly unique - [ ] **Cross-System Uniqueness:** No duplicates across systems

Assessment Questions: - Are there duplicate customer records? - Do unique identifiers have duplicates? - Are there redundant product entries? - Do transaction records have duplicates? - Are master data entities unique across systems?

Metrics to Track: - Uniqueness Rate = (Unique Records / Total Records) × 100 - Duplicate Rate = (Duplicate Records / Total Records) × 100 - Identifier Uniqueness = (Unique IDs / Total IDs) × 100

Target Thresholds: - ■ Excellent: >99.5% uniqueness - ■■ Good: 98-99.5% uniqueness - ■ Poor: <98% uniqueness

Assessment Methodology

Step 1: Data Profiling

Automated Analysis: - [ ] Run data profiling tools on all datasets - [ ] Generate statistical summaries - [ ] Identify patterns and anomalies - [ ] Create data distribution reports

Manual Review: - [ ] Sample data for visual inspection - [ ] Validate business rule compliance - [ ] Check cross-system consistency - [ ] Review data lineage documentation

Step 2: Quality Measurement

Quantitative Metrics: - [ ] Calculate quality scores for each dimension - [ ] Establish baseline measurements - [ ] Track metrics over time - [ ] Compare against industry benchmarks

Qualitative Assessment: - [ ] Document data quality issues - [ ] Assess impact on business processes - [ ] Prioritize improvement opportunities - [ ] Estimate remediation costs

Step 3: Root Cause Analysis

Common Causes: - [ ] **Data Entry Errors:** Manual input mistakes - [ ] **System Integration Issues:** ETL process problems - [ ] **Business Process Gaps:** Lack of data governance - [ ] **Technical Limitations:** System constraints - [ ] **Organizational Issues:** Unclear responsibilities

Analysis Framework: 1. Identify the quality issue 2. Trace data lineage to source 3. Analyze contributing factors 4. Determine root cause 5. Develop remediation plan

Data Quality Scorecard

Overall Quality Score Calculation

```
Overall Score = (Accuracy × 25%) + (Completeness × 20%) + (Consistency × 20%) + (Timeliness × 15%) + (Validity × 15%) + (Uniqueness × 5%)
```

Quality Rating Scale

Grade A (90-100%): Excellent - Ready for advanced analytics **Grade B (80-89%):** Good - Suitable for most business uses **Grade C (70-79%):** Fair - Requires improvement for critical uses **Grade D (60-69%):** Poor - Significant remediation needed **Grade F (<60%):** Unacceptable - Major overhaul required

Sample Scorecard Template

Dataset	Accuracy	Completeness	Consistency	Timeliness	Validity	Uniqueness	Overall	Grade
Customer Data	95%	92%	98%	85%	97%	99%	94.5%	A
Product Catalog	88%	85%	90%	95%	92%	96%	90.1%	A
Transaction History	97%	98%	85%	98%	95%	99.8%	95.1%	A
Inventory Data	75%	70%	65%	60%	80%	95%	70.5%	C

Improvement Action Plan

Priority Matrix

High Impact, High Urgency: - [ ] Critical data quality issues affecting operations - [ ] Regulatory compliance violations - [ ] Customer-facing data problems

High Impact, Low Urgency: - [ ] Strategic data quality improvements - [ ] Process optimization opportunities - [ ] Technology upgrades

Low Impact, High Urgency: - [ ] Quick wins and easy fixes - [ ] User experience improvements - [ ] Reporting accuracy issues

Low Impact, Low Urgency: - [ ] Nice-to-have improvements - [ ] Long-term optimization - [ ] Future-state considerations

Remediation Strategies

Immediate Actions (0-30 days): - [ ] Fix critical data errors - [ ] Implement data validation rules - [ ] Establish monitoring alerts - [ ] Train data entry personnel

Short-term Actions (1-3 months): - [ ] Improve data collection processes - [ ] Enhance system integrations - [ ] Implement data governance policies - [ ] Deploy data quality tools

Long-term Actions (3-12 months): - [ ] Redesign data architecture - [ ] Implement master data management - [ ] Establish data stewardship program - [ ] Upgrade technology infrastructure

Monitoring and Maintenance

Ongoing Monitoring

[ ] **Daily:** Critical data quality metrics [ ] **Weekly:** Comprehensive quality reports [ ] **Monthly:** Trend analysis and improvement tracking [ ] **Quarterly:** Full assessment and strategy review

Key Performance Indicators

[ ] Data Quality Score (Overall) [ ] Issue Resolution Time [ ] Data Freshness Metrics [ ] User Satisfaction Scores [ ] Cost of Poor Data Quality

Governance Framework

[ ] **Data Stewards:** Assigned for each domain [ ] **Quality Standards:** Defined and documented [ ] **Review Process:** Regular assessment cycles [ ] **Escalation Procedures:** Issue resolution paths [ ] **Training Program:** Ongoing education

Tools and Technologies

Recommended Data Quality Tools

Enterprise Solutions: - Informatica Data Quality - IBM InfoSphere QualityStage - SAS Data Management - Talend Data Quality

Open Source Options: - Apache Griffin - Great Expectations - Deequ (Amazon) - OpenRefine

Cloud-Based Solutions: - AWS Glue DataBrew - Google Cloud Data Prep - Azure Data Factory - Snowflake Data Quality

Implementation Checklist

[ ] Tool evaluation and selection [ ] Pilot implementation [ ] Integration with existing systems [ ] User training and adoption [ ] Performance optimization

Success Metrics

Business Impact Measures

[ ] **Decision Quality:** Improved business decisions [ ] **Operational Efficiency:** Reduced manual effort [ ] **Customer Satisfaction:** Better customer experience [ ] **Compliance:** Regulatory requirement adherence [ ] **Cost Reduction:** Lower operational costs

Technical Performance Measures

[ ] **Data Quality Score:** Overall improvement [ ] **Issue Detection:** Faster problem identification [ ] **Resolution Time:** Quicker issue remediation [ ] **Prevention Rate:** Reduced new quality issues [ ] **Automation Level:** Increased automated processes

This template is provided by Expandia.ch - Your Partner in Building Practical, Scalable AI Solutions.

Contact Information: - Website: https://expandia.ch - Email: hello@expandia.ch - Phone: +1 (555) 123-4567

Next Steps: 1. Customize this template for your specific data 2. Conduct initial assessment 3. Establish baseline metrics 4. Develop improvement plan 5. Implement monitoring processes