# Data Quality Sample Rules Guidebook

Data quality is the foundation of effective government operations and public service delivery. For state agencies, maintaining high-quality data is not merely a technical requirement but a critical governance responsibility that directly impacts policy decisions, resource allocation, citizen services, and public trust. This document contains sample data quality rules that state agencies can implement as part of a comprehensive data governance framework.

## **Completeness Check**

Required fields must not be left empty.

* A tax filing system must verify that all required tax identification fields (SSN/TIN) are populated before allowing submission
* A driver's license application must ensure all mandatory fields (name, date of birth, address) are completed before processing
* A property record system must confirm that parcel identification numbers are always present in land transaction records

## **Format Validation**

Data must conform to specified formats.

* Email addresses must follow standard format with @ symbol and valid domain (user@domain.gov)
* Vehicle identification numbers (VINs) must be 17 characters with the correct manufacturer code, year code, and check digit
* Case numbers must follow the agency's standardized format (e.g., ABC-2025-123456)

## **Range Validation**

Numeric values must fall within expected ranges.

* Citizen age values must be between 0-120 years for healthcare enrollment systems
* Income reported on benefit applications must be non-negative values
* Speed measurements in traffic violation systems must be between 0-200 mph

## **Code Value Validation**

All coded values must match official reference tables and code lists.

* Occupation codes in workforce systems must match Bureau of Labor Statistics standard occupation codes
* County codes entered in property records must exist in the official county code reference table
* Department IDs in employee records must match valid departments in the organizational structure table

## **Unique Identifier Rule**

Primary keys and unique identifiers must not contain duplicates.

* Each citizen must have only one state ID number in the central identity management system
* Each tax parcel must have a unique parcel identification number in the property assessment database
* Each business entity must have a unique business registration number in the business licensing system

## **Consistency Across Systems**

Identical data elements must have the same values across different systems.

* A citizen's legal name must match exactly between voter registration, driver's license, and tax systems
* Address information for a property must be consistent between property tax, utility, and emergency services databases
* Business entity information must be consistent between licensing, tax, and regulatory compliance systems

## **Timeliness Rule**

Data must be updated within specified timeframes.

* New birth registrations must be entered into vital records systems within 5 business days of receipt
* Address changes must be processed and reflected in all connected systems within 3 business days
* Court dispositions must be reported to criminal history systems within 24 hours of judgment

## **Business Rule Validation**

Data must comply with specific business rules.

* A dependent child's age on tax benefits must be under 19 (or under 24 if a full-time student)
* Professional license renewal dates must be set based on birthdate and license type according to regulatory requirements
* Unemployment benefit calculations must follow state-specific formulas based on work history, income, and claim type

## **Statistical Anomaly Detection**

Values should not deviate significantly from historical patterns without explanation.

* A 50% increase in benefits applications in a single county requires verification and explanation
* Significant deviation from seasonal patterns in tax collections triggers automatic review
* Unusual spike in license transactions at a specific DMV office compared to historical averages requires investigation

## **Standardization Rule**

Text entries must follow standardization rules.

* All street designations must use approved USPS abbreviations (St., Ave., Blvd.)
* Agency names must be stored using official full names, not acronyms or shortened versions
* Professional credentials must be recorded using standardized nomenclature from the official credentials reference table

## **Cross-field Validation**

Related fields must be logically consistent.

* Program enrollment end date must be after the enrollment start date
* Child's date of birth must be after parent's date of birth in family benefit applications
* Employment end date must be blank if employment status is "Currently Employed"

## **Logical Value Check**

Field values must make logical sense in context.

* Pregnancy status should not be marked for male patients in health systems
* Retirement benefits should not be calculated for employees under minimum retirement age
* Concealed carry permit applications should not be approved for individuals with qualifying criminal convictions

## **Duplicate Record Prevention**

Systems must identify and prevent duplicate citizen or case records using defined matching criteria.

* New voter registrations must be checked against existing records using name, DOB, and SSN matching algorithms
* Business registration systems must prevent duplicate business entities by checking name, address, and tax ID
* Human services case management systems must identify potential duplicates using household address and member information

## **Geographical Validation**

Address data must conform to USPS standards and verified against official address databases.

* Residential addresses in voter registration must validate against USPS delivery point validation system
* Business locations in licensing databases must match official municipal addressing systems
* Service delivery locations in health and human services systems must verify against official address databases