



# Factsheet Data Quality Rule

## Content

Introduction.....	2
Definition.....	2
Synonym.....	2
Purpose.....	2
Procedure to manage Data Quality Rules .....	2
Characteristics and requirements .....	2
Relationships .....	2
Classification A.....	3
Classification B.....	4
Examples A .....	5
Examples B .....	5
Tips .....	5
References.....	6
Authors.....	6



Stichting DAMA-NL, The Netherlands, 2021. This work is licensed under the Creative Commons Attribution 4.0. To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>.

Version 1.1  
Date 16 May 2021

## Introduction

This factsheet describes knowledge about DQ Rules in a nutshell. DQ Rules are highlighted from different angles in a structured way.

## Definition

A Data Quality Rule is a constraint used to validate data values or records.

Note: Non-compliances with Data Quality Rules can be regarded as a Data Issues.

## Synonym

Data validation rule

Data quality rule specification

## Purpose

The purpose of a DQ Rule is to prevent, monitor or report insufficient quality of **data values**.

## Procedure to manage Data Quality Rules

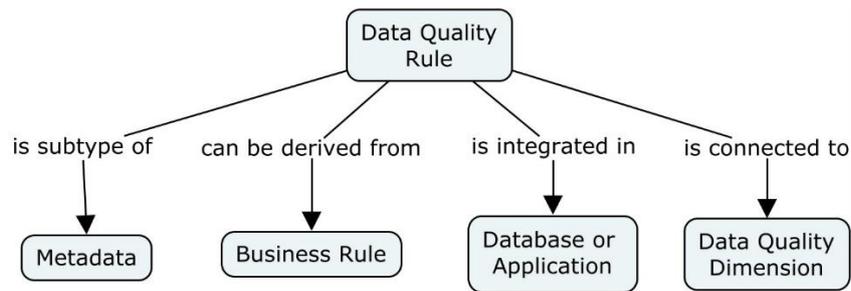
Phase	Activity
Plan	<ul style="list-style-type: none"><li>• Derive DQ Rules from <b>Business Rules</b></li><li>• Collect/identify DQ Rules</li><li>• Specify DQ Rules</li><li>• Document DQ Rules as <b>metadata</b></li><li>• Establish DQ rules</li></ul>
Do	<ul style="list-style-type: none"><li>• Implement/apply DQ Rules in a <b>database</b> or <b>application</b>.</li></ul>
Check	<ul style="list-style-type: none"><li>• Test DQ Rules</li><li>• Evaluate DQ Rules</li></ul>
Act	<ul style="list-style-type: none"><li>• Adapt/Maintain DQ Rules</li></ul>

## Characteristics and requirements

Characteristic	Requirement
Accessibility	DQ Rules are accessible to data consumers. It ensures that data consumers can ask questions about it and provide feedback on rules.
Unambiguity	DQ Rules are unambiguous. It ensures that they can be implemented in a database or application.

## Relationships

- A DQR is a kind (subtype) of **metadata**.
- A DQR is associated with **data quality dimension** regarding data values and records, i.e., accuracy, completeness, consistency, uniqueness.
- A DQR is integrated in a **database** or **application**.
- A DQR can be derived from a **business rule**.



## Classification A

In table 1 DQ Rules are classified in three categories, and their subcategories.

**Table 1:** Category, subcategories, and examples.

Category and subcategory	Example
<b>Simple data element content rules.</b>	
These are considered “simple” because you only need to inspect the contents of a single data element and check to see if the content meets the rules.	
Valid values, range, data type, pattern, and domain.	
Optional versus mandatory (evaluates completeness).	
Reasonable distribution of values.	In a customer database, you would expect a fairly even distribution of birthdays; a much larger number of birthdays on a given day of the year probably indicates a problem.
<b>Cross data element validation rules.</b>	
The rules require inspecting values in multiple data elements (typically in a single data file) to determine whether the data meets the quality rules.	
Valid values that depend on other column values	An overall list of location codes might pass the simple-column content rules, but only a smaller list of locations is valid if, for example, the region code is set to “West.”
Optional becomes mandatory when other columns contain certain data.	The <i>Value of Collateral</i> field may be optional, but if the loan type is “mortgage,” a positive value must be filled into the <i>Value of Collateral</i> field.
Mandatory becomes null when other columns contain certain data.	The <i>Writing Insurance Agent Name</i> field might normally be mandatory, but if the <i>Origination Point</i> is “web” (indicating the customer applied for the policy online), the <i>Writing Insurance Agent Name</i> field must then be blank.
Cross-table validation rule. As the name suggests, these data quality rules check columns (and combinations of columns) across tables.	An example cross-validates the name of a city with the name of a state in an address table—that is, Minneapolis is <i>not</i> in Wisconsin.
<b>Cross data files validation rules.</b> As the name suggests, these data quality rules check data elements (and combinations of data elements) across data files.	
Mandatory presence of foreign-key relationships	If an account must have a customer, then the account table must have value in the <i>Customer ID</i> column that matches a value in the <i>Customer ID</i> column of the Customer table.

Category and subcategory	Example
Optional presence of foreign-key relationships depending on other data	If the <i>Loan Type</i> is “mortgage” in the Loan table, there must be a matching value for <i>Loan_ID</i> in the Collateral table. On the other hand, if the <i>Loan_Type</i> is “unsecured,” then there must <i>not</i> be a matching value for <i>Loan_ID</i> in the Collateral table because “unsecured” means there is no collateral for the loan.
Columns in different tables are consistent	If the <i>Collateral_Value</i> column contains a value above a certain level, the <i>Appraisal_Type</i> must be “in person” because of the high value of the property.

## Classification B

In Table 2 DQ Rules are classified in ten categories.

**Table 2:** Category of DQ Rule, description, and example

Category	Description	Example
Domain List	A domain list rule defines a list of values that a data element is allowed to have.	The Gender data element can have 'M' or 'F'.
Domain Pattern List	A domain pattern list rule defines a list of patterns that a data element is allowed to conform to. The patterns are defined in the regular expression syntax.	An example pattern for a telephone number is as follows: (^[[::space]]*[0-9]{ 3 }[[::punct :space:]]?[0-9]{ 4 }[[::space]]*\$)
Domain Range	A domain range rule defines a range of values that a data element is allowed to have.	The value of the salary data element can be between 100 and 10000.
Common Format / Pattern	A common format rule defines a known common format that a data element is allowed to conform to.	This rule type has many subtypes: Telephone Number, IP Address, SSN, URL, E-mail Address.
No Nulls	A no nulls rule specifies that the data element cannot have null values	The department_id data element for an employee in the Employees table cannot be null.
Functional Dependency	A functional dependency defines that the data in the data object may be normalized or derived	
Unique Key	A unique key data rule defines whether a data element or group of data elements are unique in the given data object.	The name of a department should be unique.
Referential	A referential data rule defines the type of a relationship (1:x) a value must have to another value.	The department_id data element of the Departments table should have a 1:n relationship with the department_id data element of the Employees table.
Name and address	A name and address data rule evaluate a group of data elements as a name or address	
Custom	A custom data rule applies a SQL expression that you specify to its input parameters.	VALID_DATE with two input parameters, START_DATE and END_DATE. A valid expression for this rule is: “THIS”.“END_DATE” > “THIS”.“START_DATE

## Examples A

Table 3 shows how a DQ rule is derived from a business rule.

**Table 3:** Example of a DQ Rules derived from Business Rule

Business Rule	DQ Rule
The marital status code may have values of single, married, widowed, and divorced. It may not be left blank. A value must be picked when entering a new customer. The values for widowed and divorced are tracked separately from single because risk factors are sensitive to whether the customer was previously married and is not married anymore.	<p>“Customer.Mar_Stat_Cd” may be “S,” “M,” “W,” or “D.”</p> <p>Blank is considered an invalid value.</p>

## Examples B

Table 4 shows DQ Rules of two fields.

**Table 4:** Example of DQ Rules

Data element	Data quality rules
Email	<ul style="list-style-type: none"> <li>• An email must contain the '@' sign.</li> <li>• '@' must be used only once.</li> <li>• An email must contain any or all of the following: letters, digits, non-alphabetic characters, such as, ! # \$ % &amp; ' * + - / = ? ^ _ ` { .</li> </ul>
Customer name	<ul style="list-style-type: none"> <li>• The 'Customer's full name' field an email refers to must not be 'Null'.</li> <li>• Customer's full name must consist only of letters; no other characters allowed.</li> <li>• Only first letters in customer name, middle name (if any), and surname must be capitalized.</li> </ul>

## Tips

1. Involve subject matter experts from various department.
2. Be moderate with the number of rules. Define rules that really matter according to the subject matter expert. Don't create rules which already are implemented in the sources system as input check.
3. Favour a step-by-step approach.
4. Treat each field of the database individually and creates rules accordingly.
5. Decide between a centralized and local storage for your data quality rules.

## References

DAMA (2017). DAMA-DMBOK. *Data Management Body of Knowledge*. 2<sup>nd</sup> Edition. [Technics PublicationsLlc](#). August 2017.

DAMA Dictionary of Data Management.

*Data migration checklist: Planner + template for effective data migration planning — Data migration pro*. (n.d.). Data Migration Pro. <https://www.datamigrationpro.com/data-quality-rules-management-repository>

*Data quality - Data rules*. (2020, March 31). Datacadamia - Data and Co. [https://datacadamia.com/data/quality/data\\_rule](https://datacadamia.com/data/quality/data_rule)

*Data quality and business rules explained: Expert interview with Ronald G. Ross — Data quality pro*. (2020, February 29). Data Quality Pro. <https://www.dataqualitypro.com/blog/business-rules-for-data-quality-ronald-g-ross>

*Data quality rule*. (n.d.). ScienceDirect.com | Science, health and medical journals, full text articles and books. <https://www.sciencedirect.com/topics/computer-science/data-quality-rule>

Reeb, B. (n.d.). *Data quality rules*. IData Insights Blog. <https://blog.idatainc.com/data-quality-rules>

*Tips to create effective data quality rules*. (2016, April 29). TDAN.com. <https://tdan.com/tips-to-create-effective-data-quality-rules/24525>

*What are data quality rules? How do they look, and why are they needed in data management? Our expert provides the answers!* (2020, September 3). Data Quality Solutions: Data Quality Software & DQaaS | CDQ. <https://www.cdq.ch/data-sharing/data-quality-rules>

## Authors

Members of the Working Group of Data Quality of DAMA-NL.