DATA QUALITY DISCIPLINE

How to manage data quality for business value?

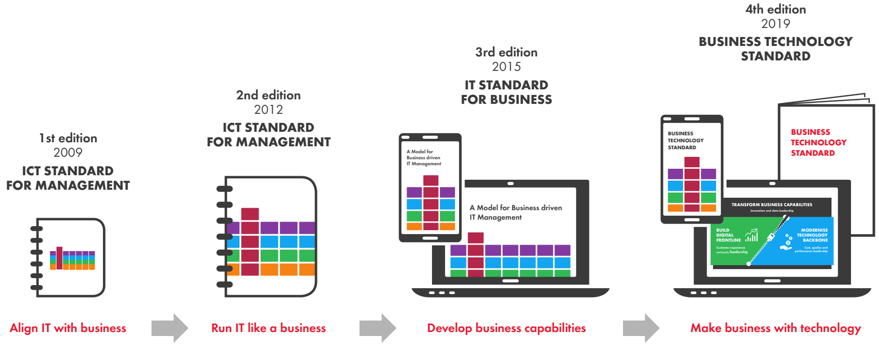
ARTICLE

By Business Technology Forum

About Business Technology Standard

The Business Technology Standard (or BT Standard) is an open-source management framework to plan, build and run information technology in today's technology-driven business world. It has been constantly developed and renewed during the past 10 years with global companies and public organisations. It is recognised today as one of the leading best practices and used in hundreds of globally operating companies and public organisations, especially in Nordic countries.

The fourth edition has been completely rewritten and upgraded, and the scope of technology management has been extended from information technology to business technology.



The Business Technology Standard has been developed by the Business Technology Forum, a community of forerunner companies and organisations collaborating based on a platform economy model where every company can benefit from each other's development input and efforts.

Authors and acknowledgements

This article has been contributed to the Business Technology Standard by the following authors:

Markus Sipilä (Sofigate)
Pertti Tuisku (CGI)
Jouko Eronen (Tikean)
Mikko Salomaa (Tikean)
Jaakko Kekäläinen (Sofigate)

The contribution of the following individuals who offered their time to review this article is highly appreciated:

Tuula Koppanen (Konecranes)
Ville Ikonen (Konecranes)
Ian Johns (King's College London)
Petri Hassinen (Valmet)
Lasse Rantalainen (Valmet)
Antti Pikkusaari (Aedon)
Panu Arvila (Sofigate)
Tobias Eckstein (Sofigate)
Hanna Heiskanen (Sofigate)
Elena Van Leemput (Sofigate)

CONTENTS

1 Introduction And Structure Of This Article

Data is one of the most valuable resources for any organisation. It is a strategic business asset with value that is dependent on the quality, relevance, and scope of the data.

As illustrated in the figure below, data quality is one of the seven aspects of data asset management in Business Technology Standard's "Operating Model for Data Governance and Management". This article is a deep dive to one of these aspects, data quality management.

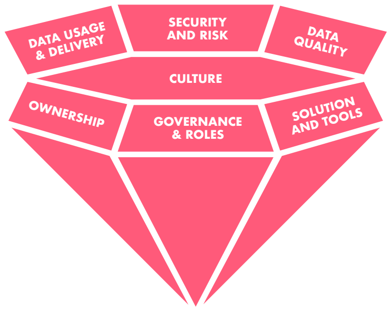


Figure 1 *Data quality is one of the aspects of data asset management*

Poor data quality has direct cost implications. It also introduces risks, which may become even more expensive than the direct costs if these risks realize. Chapter 2 of this article focuses on these costs and risks.

Targeted one-off quality improvement actions can be done whenever a need arises but if an organization wants to approach data quality systematically, it needs to have fundamental elements of data management in place. Chapter 3 gives an overview of the data governance and data management aspects that are required for systematic data quality management.

The actual data quality assurance starts from planning the data quality control activities and defining the data quality rules. Chapter 4 focuses on this topic and chapter 5 continues with how corrective and preventive actions should be managed and made part of business governance.

Based on the experience of the authors and reviewers of this article, some best practices of data quality management have been documented to chapter 6.

The world around us is changing faster than ever before. Data quality that might be considered good today, might not be sufficient at all tomorrow. Chapter 7 discusses the challenges that evolving business needs introduce to data quality.

Implementing data quality management practices can take very different forms depending in the context. Chapter 8 discusses implementation considerations when implementing data quality management to ongoing business operations, in data migration and cleansing projects and as part of business transformation programs.

2 Why Data Quality Matters: Costs And Risks Of Poor Data Quality

Poor data quality will hit customer and employee satisfaction and company bottom line and will in the worst case jeopardize the development projects that are dependent on good quality data. Making the costs related to poor data quality transparent and evaluating business risks caused by poor data quality provides information for data related decision making. Bringing the data quality aspects part of the agenda of the business process governance structures allows business leaders to identify the most critical problems, plan improvement actions and follow up the results of those actions.

Let's consider the following aspects of data quality:
- Profitability aspects: What is the cost of poor data quality?
- Risk management aspects: What kind of risks poor data quality can introduce (negative risks)? And on the other hand, what kind of opportunities (positive risks) good data quality can bring?

2.1 Cost Of Poor Data Quality

Cost of inefficient business processes
Let's consider two business processes where the latter is dependent on the data coming from the first one. The latter process should

1. Be able to trust the data coming from the previous process is correct
2. Be able to assume that all necessary data is present

A simple example would be a sales process which results to a sales order. The sales order data is input for invoicing process.
- Let's say that the sales process didn't have the customer's purchase order number available at the time when the sales order was created.
- Let's say the salesperson forgets to follow up with the customer to get their purchase order reference and thus the purchase order number is missing from the invoice when we send it to the customer.
- When the invoice is received, the customer complains that the invoice is missing their purchase order reference and is unwilling to pay the invoice.
- The salesperson needs to reach out to the customer's contact person and request them to create a purchase order for the deal that they had agreed.
- When the salesperson eventually receives customer's purchase order reference, he/she will update it to the sales order, the corrected invoice can be sent, and the customer finally pays the invoice.

This simple example illustrates several steps of process inefficiency. All this hassle could have been avoided if
- The customer master data would have a properly filled information that this customer always requires a purchase order reference
- The system would have warned upon the sales order creation that the customer's purchase order reference is empty
- The salesperson would have requested the purchase order reference in the beginning.

In a McKinsey article published in June, 2020, the amount of time employees waste on non-value adding tasks due to bad data quality can be significant, up to one third of their working time. The leading firms that have paid attention to the topic had 75% less of this overhead compared to the average.

Re-work cost of correcting and cleansing of data
A classic example is duplicate customer master data records which are created over the years because the data creation processes have not been clear, they have not been followed, or both. While it might have been possible to execute the existing business processes, it might become a bottleneck for new data uses such as customer facing digital frontline solutions or an ERP transformation.

Depending on the magnitude of the dataset, the clean-up activities can take significant amount of time and effort.

Cost of customer unsatisfaction and claims
Let's consider an example where a company is sending e-mail notifications to the customer when service orders have been completed. These notifications are automatically generated based on the input collected from the service technician's during their site visits. If the descriptions the technicians are writing do not make sense to the customer, the good intention of these notifications may turn against the company and result in unsatisfied customers which might not renew their contract.

In some context there might even be penalty costs that are exercised if the company is not able to accurately report what has been done.

2.2 Risks Of Bad Data Quality And Opportunities With Good Data Quality

Risk of not meeting strategic business goals
Data can easily become a major bottleneck in big strategic transformation programs. In the worst-case bad data quality can cause the whole transformation program to fail. It is quite typical that the expectations for data increase in major transformations and thus, data should be taken to the agenda as early as possible.

Business Technology Design is a method for systematic mapping of value streams and required capabilities, including data. The earlier the new expectations for data are identified in transformation programs, the better because the data preparations might take significant amount of calendar time and effort and might even require a separate sub-project. On the other hand, modern technologies combined with high quality data can bring huge business benefits. See chapter 7 for further considerations on this topic.

In a study published in MIT Technology Review in April 2021, only 13% of the participating organizations considered themselves performing high on delivering on their data strategy. According to the study, the common nominator for these high-achievers is their attention to the foundations of sound data management and architecture.

Risk of basing business decisions in inaccurate data or incomplete picture
Companies are using more and more data and analytics as a basis for decision-making. However, if the data is not consistent or some data is missing from the analysis altogether, the business decisions might be based on inaccurate, incomplete, or even completely false assumptions which can have disastrous results.

On the contrary, if the organization is using well defined and quality assured datasets and there is high data literacy in the organization, modern analytics can make a huge difference and bring competitive edge.

Reputation risk and risk for losing business
A classical (but unfortunately common) example of bad data quality is duplicate records in customer master data. If part of the sales is being booked to customer 123 and part to customer 234 when these are actually one and the same customer, it might be very embarrassing to present only customer 123's data in a customer portal. If the internal mess is exposed to customers like this, it might decrease their trust in the company and trigger them to seek for other alternatives.

Risk of diminished regulatory compliance
Regulatory compliance reporting was quite common driver in the early data warehouse implementations. Even though modern data platforms are used to more and more sophisticated purposes, the compliance reporting is still a valid use case. It goes without saying that bad data quality decreases the trustworthiness of reporting.

Risk to show/publish wrong or unauthorized data
The effects of data quality to GDPR compliance and also cyber security can be tremendous. There are multiple aspects related to this, consider the following examples:

- Is the collection and processing of personal data compliant with the regulations?
- Is the data correctly associated to different customers? If not, one customer might see other customers' data.
- Are the access control principles clearly defined and does the Data Owner take the ownership of this aspect of data as well? If not, it might be that unauthorized users have access to data that they shouldn't.
- The technical aspects of securing data is a topic of its own but it is essential to understand that the technical data protection measures are of little use if the aspects described above are not in good shape.

3 Building A Foundation For Data Quality

When talking about data quality, the very first thing is to understand
- what data we are talking about
- what does good quality mean for this data?

It is usually quite easy to identify some aspects to what good quality means for a given data asset. For example, the customer master should not contain duplicates or that all customers should have an official business ID. Some targeted one-off quality improvement actions can be done based on these and there is nothing wrong in doing such quality clean-up exercises. However, if we raise the bar and want to start managing the data quality systematically, we need to have fundamental elements of data management in place.

3.1 Definition For Good Data Quality

Data quality can be considered good when it meets the expectations of the data consumers in the intended data use. DAMA has defined data quality management in a similar way in the Data Management Body of Knowledge. Data quality management is defined in DMBOK as " The planning, implementation, and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers." (DAMA-DMBOK 2, second edition, 2017). According to this definition, data is of good quality when it is "fit for use".

Our definition in this article is a bit more specific so let's consider it a bit closer. When defining and assessing data quality, we need to know and understand:

- Intended data use
- Data consumers (users of the data in the intended data use)
- Expectations of the data consumers in the intended data use

When considering the corrective and preventive actions, understanding the processes for data creation and consuming is vital.

Dimensions of data quality

Different data quality dimensions exist in literature (DAMA-DMBOK, first edition, 2010 & Sebastian-Coleman, 2013). We have chosen to use the following data quality dimensions in this article.

Dimension	**Description**
Completeness	All records exist, all required fields are filled.
Validity	Data conforms to a set of business rules. This can be a simple syntax validation, but the validity might also be derived from the context. For example, a planned delivery date cannot be earlier than the date when an order was created.
Consistency	Data is consistent across systems and over time. For example, if a data is mirrored from one system to another, we can compare if they have the same values. There is also a time aspect to the consistency: is the data created this week comparable to the data created last week?
Integrity	Data conforms to the data relationship rules as defined by the data model. Example 1: there are no orphan records where a sales order is referring to a customer that has been deleted from the database as a duplicate. Example 2: All sales orders have a reference to a customer.
Uniqueness	There are no duplicate records. The process of eliminating duplicates is called deduplication. It is good to understand that the same real-world entity might appear as two separate records with different IDs with slightly different name, but they are still meaning the same thing.
Timeliness	Data is available when it is needed. For example, does the customer record exist in the system where sales orders are created at the time the order is created?

Accuracy	Data represents the real world. Note that the accuracy might be very difficult to evaluate. For example, it is relatively easy to define a validity rule for a phone number syntax, but we cannot necessarily know if the phone number is correct. Another note is that the accuracy has a time aspect. The phone number can change so even if the data was known to be accurate a year ago, it does not necessarily mean that it's still accurate. This is why it is quite common that quality checks for data accuracy are built-in to the business processes. For example, when registering an appointment with a doctor, there might be a check that the phone number is still correct.

3.2 Identify Strategic Data Domains And Data Assets

When establishing data management and data governance, one of the very first things is to identify the strategic assets the organization has. The approach used in Business Technology Standard Operating Model for Data Governance and Data Management is to use a map of Data Domains and Data Assets as the highest abstraction level for this purpose, see figure below.

Figure 2 An example map of data domains and strategic data assets.

Once the data domains and strategic data assets have been identified, each data asset needs to be managed to maximize its value. The Business Technology Standard's Operating Model for Data Governance and Data Management defines seven perspectives to data asset management. Data quality is one of these aspects. See figure 1.

3.3 Data Standard Defines the Data Asset

At minimum, a data standard contains two things:

- Definition of what the data represents
- Definition of the data structure and format

Conceptual level data modelling: Definition of what the data represents
When starting to work with a given data asset, the very first thing is to ensure that the whole organization shares the same understanding what the data represents in the real world. This is usually referred as conceptual data modelling. (DAMA-DMBOK, first edition, 2010). Consider the following example.

An organization in machinery industry manufactures equipment and provides service for them. For the service business, the term "customer" means the physical customer premise where the service visits are done. For the finance department, the term customer refers to a legal entity that we have a contract with. If one customer company has 10 physical locations, there is a fundamental mismatch; with the interpretation of the service business there would be 1+10 customers but from the finance department's point of view there is only 1 customer. The solution to this mismatch is to make a distinction and define two entities for example like this:

- **Customer** is a legal entity that we do business with.
- **Customer Location** is a location where the equipment under service are located.

In the example above, it was possible to define Customer and Customer Location as two separate entities with different definitions. However, changing the name of an entity might not always be possible due to established conventions. For example, in higher education and in the context of employees, the term FTE means “full-time-equivalency” for the purpose of a work year. In the context of students, FTE means students that are studying full time (full-time-enrolled students). Having a definition of FTE in different contexts and documenting this to the data dictionary helps the organisation to understand what the data represents in the real world.

Enterprise logical data modelling: Definition of the data structure and format
The target of data quality assurance is to ensure that data conforms to the data standard. After the conceptual level data modelling, the next level of data modelling is to define a data structure and the format of the data. The structure defines the attributes (fields) that the entity has. The format defines the expected format for each attribute. (DAMA-DMBOK, first edition, 2010) Consider the following simplified example.

Attribute	Type	Required	Definition
Customer ID	Numeric	Yes	Unique identifier that is used to identify the customer in all IT solutions and data storages.
Name	Text	Yes	The official name of the customer as present in an official register.
Business ID	Text	Yes	The ID of the legal entity in an official register. The Syntax is in a format xxxxxxx-y.
Industry	Select list	Yes	Customer's industry used in customer segmentation. Possible values: X, Y, Z.

To summarize, at simplest the data standard contains the definition of the entity and the structure and format of it. It is a good idea to build a business data glossary, data catalogue and/or data dictionary that contains a list of all data assets, their definitions, their business owner, data domain lead from data organization, information in which systems the data is created and used, association to business process descriptions and links to data standard documents.

Reference data is an important aspect of data modelling
In the example above, the attribute for customer's industry is an example of reference data. Reference data is used to categorize or classify other data. If the data is used in multiple different systems, it is quite typical that there are mismatches in the allowed values if reference data is not managed properly.

- Let's say that the industry is a drop-down / select list with possible values X, Y and Z in one system. Let's say that in a second system there are only two possible values, X and Y. This will cause problems if you try to send data from the first system to the second via an integration because the value Z doesn't exist in the second system.
- If the second system has values A, B, C and D instead of X, Y and Z, we have slightly bigger problems because the possible values for Industry are completely different.

These two examples illustrate the importance of data modelling and data standards. The data standard defines the allowed values and all systems in the enterprise must conform to this data standard.

When integrating different IT systems, it is inevitable that the integration layer needs to transform the format of the data being sent, but it's alarming if business logic like "apples in system A mean oranges in system B" start to creep into the integration layer. These kind of mapping rules are very difficult to maintain, and it may decrease how well the data can be utilized in analytics and other use cases. Furthermore, the data standard itself must be under change control so that changes are done in a controlled manner.

Data standard allows to define what good quality means for this data asset
When the data standard is defined, it is possible to plan data quality assurance controls by considering the data standard and the data quality dimensions described earlier in this article. With the simplified data standard used in our example earlier, we could define the following simple quality controls:

- **Completeness:** Are all required attributes present for each customer? Or do we have customers that are missing for example the business ID?
- **Validity:** Is the Business ID syntactically correct for all customers?
- **Consistency:** Does the product data have the same weight in different systems? Do different systems have the same possible values for reference data fields?
- **Uniqueness:** Do we have duplicate customers i.e., is each customer present only once in our systems? Do we have a unique identifier to match the same customer between different systems?
- **Accuracy:** When the official business ID is known, it is possible to cross-check the spelling of the customer's name against an external data source.

The simple example above illustrates how the data standard is used to define what good quality means for one data asset. Chapter 4 discusses this topic in detail.

3.4 Define Ownership For Data

Data quality management is next to impossible if the data does not have business owners who see the value of good quality and who ensure that the whole organization take the quality seriously.
The Business Technology Standard Operating Model for Data Governance and Data Management defines the roles

between business, data function and IT, see the figure below. Business has the data ownership, including the quality of the data. The cross-functional data function is responsible for defining and maintaining the data standards according to the business needs. The solution architects in IT design the systems.

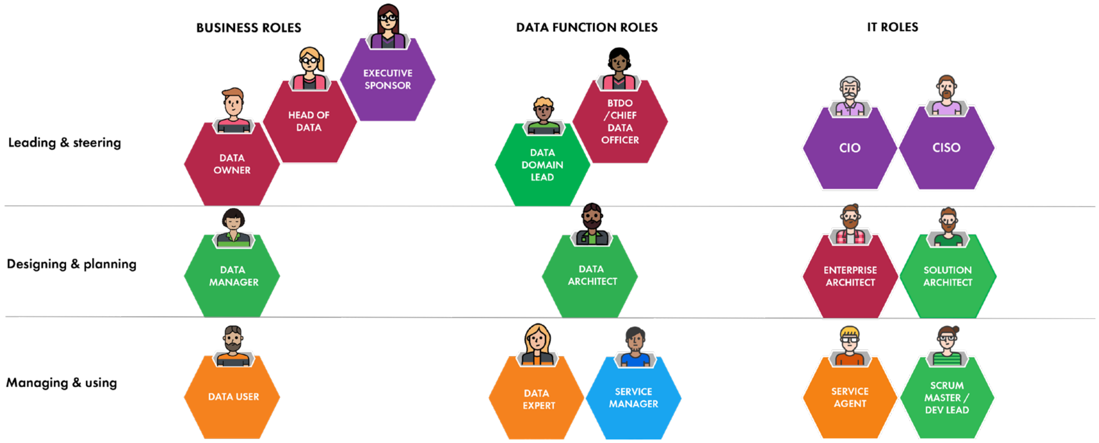


Figure 4 *Data roles in business, data function and it.*

Data owner is a critical business role for data quality. Each data asset needs to have a nominated data owner who is responsible for planning data use cases and outcomes and who is accountable for the data quality at the end of the day.

Data managers have the best knowledge of the data content and quality needs for each business purpose. Data managers follow closely data quality reports for their data assets, coordinate corrective actions and implement changes in processes which are targeting to improve the data quality. In large organizations there might be several data managers for one data asset, for example each plant could have a named data manager for item data. The way the roles, especially the business roles, are implemented in an organization can have quite wide variance depending on the structure of the company.

3.5 Ensure That Data (And Quality) Is On The Agenda Of The Relevant Governance Bodies

The aspects of data governance defined in the Operating Model for Data Governance in the Business Technology Standard are illustrated in the figure below. When considering data quality, the most important aspects are:

- Governance structures for Enterprise Architecture need to ensure that the data architecture standards and data modelling principles are followed. When the data standards are defined in the same manner for all data assets, it is easier to define the data quality controls for each asset.

- Governance structures for development need to ensure that the development projects and small enhancements follow the established principles, policies and processes. For example, new fields should not be added to data entities without proper change control.

- It is essential that the governance structures for business processes take data aspects, and especially data quality, to their agenda.

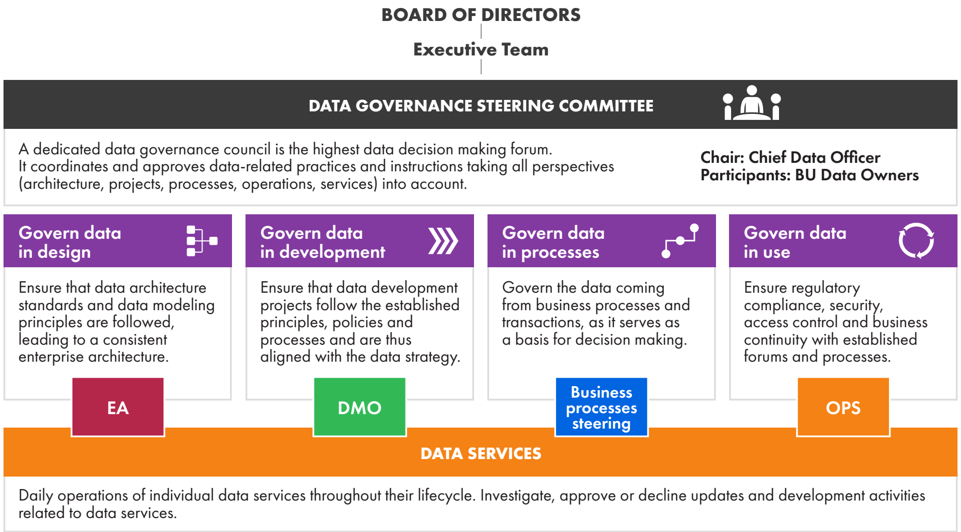


Figure 5 *Aspects of the data governance in the Operating Model for Data Governance in Business Technology Standard*

4 Defining What Good Quality Means For A Given Data Asset

4.1 Introduction To Data Quality Assurance

The purpose of data quality assurance is to ensure that data conforms to data standard. Planning of data quality assurance follows the two principles:

- First time right: Data should be created and later maintained so that it conforms to the data standard without any further actions. This is called the "first time right" principle. This saves rework and enables our second key principle.
- Data from the previous process step can be trusted: This principle will guide us in defining roles and responsibilities for data quality assurance. For example, if we find a problem in the data, we will ask the original creator (or role) of the data to correct the data. This ensures feedback to the creator of data. It also ensures that data in the source system is corrected.

The control activities need to be planned and prioritized so that the most important data assets are covered first. As discussed in chapter 3.2, the first step is to identify the strategic data domains and data assets. Once this is done, the control activities can be started from the prioritized data assets and once those are in place, the next data assets will follow.

Master data forms the foundation for business activities, so it is logical to start the data quality assurance activities from them. Another rationale for starting from master data is that transactional data is typically referring to the master data records so it is not possible to achieve good quality in transactional data without ensuring good quality of master data first. Typical master data assets include

- Product data
- Customer data
- Supplier data
- Customer assets
- Customer contacts

When planning the corrective and preventive actions, an understanding of the end-to-end business process and data flow is essential, see figure below.

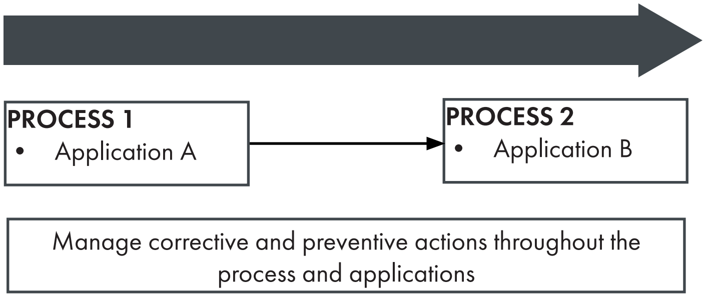


Figure 6 *Corrective and preventive actions should be managed throughout the process flow*

The corrective and preventive actions may include for example the following aspects:

People:
- Clearly defined roles and ownership
- Sufficient time allocated
- Support from data team
- Training

Data processes:
- Understanding of data end-to-end data flow
- Instructions
- Review and approval workflows

Applications and user interfaces:
- Reference and master data management
- Real-time validations and use of external data sources
- Data creation friendly user interfaces

End-to-end flow lineage:
- Control data consistency across systems

4.2 Plan Control Activities For A Given Data Asset

Clarify data quality requirements for quality assurance purposes

Data standards do not necessarily define the quality requirements on a sufficient level for measuring whether certain

field in data conforms to the specification. Data quality is much more than "does the field contain a value" and "is the value syntactically correct". The data quality dimensions accuracy, completeness, consistency, timeliness, validity, integrity and uniqueness can be used as a checklist when defining the data quality rules.
Typically, we need to:

- Review the data standard and other available documentation to understand the business requirements for data quality
- Check the requirements against data quality dimensions to identify possible unclear or missing requirements
- Clarify identified unclear requirements with data owner

It is also important that data standards and other key documentation are under change control. This enables proactive modification of data quality assurance activities in case of changes in data standard.

Profile data
"Inventory" of the existing data, also known as profiling, forms a baseline for data quality assurance. Profiling gives an understanding of the order of magnitude for needed resources and tools to manage the quality. After sufficient understanding has been achieved, it is recommended to scope the data quality control activities to the most relevant areas.

The example below illustrates what profiling could mean for analysing customer data. The sufficient level of details obviously varies case by case.

- Each attribute is profiled (e.g. length, data type, what kind of values we have) to identify for example missing values or inconsistent use of values.
- Integrity checks are done to identify possible orphan records
- Checking when the data has been last updated
- Lifecycle status checks, inactive / active
- Number of sales orders and sales value during a defined period to understand which customers active and which ones are inactive. This helps to focus the corrective actions to the most important customers.
- Initial duplicate analysis

If the data is used in multiple systems, the inventory should also take into account a consistency check across systems.

Content of data to be investigated
Define which data attributes are most relevant for the quality control of the selected data sets. Define attribute level requirements for the data accordingly:

- Data types (for example, date, boolean, etc.)
- Range of values (for example, weight must be greater than 0, but smaller than 100 kg)
- Dependencies (for example, if the field has value Y, then other field must be Z)
- List of values (for example, Company A, Company B, Company Z)

Select suitable implementation mechanisms and the tooling in accordance with the requirements from previous steps.

4.3 Define Data Quality Rulesets And Data Quality Error Reports

Converting the business requirements to data quality rulesets
While the data itself may seem logical and straight forward, the business logic beyond the data might be more com-

plicated. When the business requirements for data objects are understood, these requirements can be converted to concrete data quality rulesets.
The rules can be

- Simple attribute validations that check for example that required attributes are not empty.
- Attribute specific rules that validate that some key attributes contain expected values.
- Business logic level rules containing multiple logical rules ("if this or that, then ...")

In large organizations there might be a need to different rules on global level, on sales organization and on local unit level. In addition to differentiating the rules, ownership for these rules must be set in place in order to keep them up to date and managed properly. Documenting the rules and making them available for stakeholders plays important part in transparency of the process.

Defining data quality rules in a data quality assurance tool
When the amount of data is manageable and data quality rules are simple, we don't necessarily need any special tools and simple reports or pre-defined SQL queries may be sufficient.

However, when the amount of data grows, the amount of data quality rules grows or we want to take a more structured approach to data quality assurance, tools designed for data quality assurance come into play. These kinds of tools allow data managers to systematically start addressing the quality situation by creating rulesets, get summarized views of the quality status and filter down to relevant data.

The example below illustrates product data specific rule configuration in Tikean Data Quality Platform which allows rules to be created and maintained with logical rules without coding.

- This example rule validates that the global product data records contain a valid unit of measure (KG, M2 or M3).
- The rule is limited only to those records where the version status is "accepted" so that we don't get false alerts for example for those historical records which are marked as "obsolete".
- The rule is classified as critical because the business processes cannot operate if this information is missing or contains an unexpected value.
- Pro-tip: use a naming convention for the rules
- Pro-tip: Use error messages that are self-exploratory

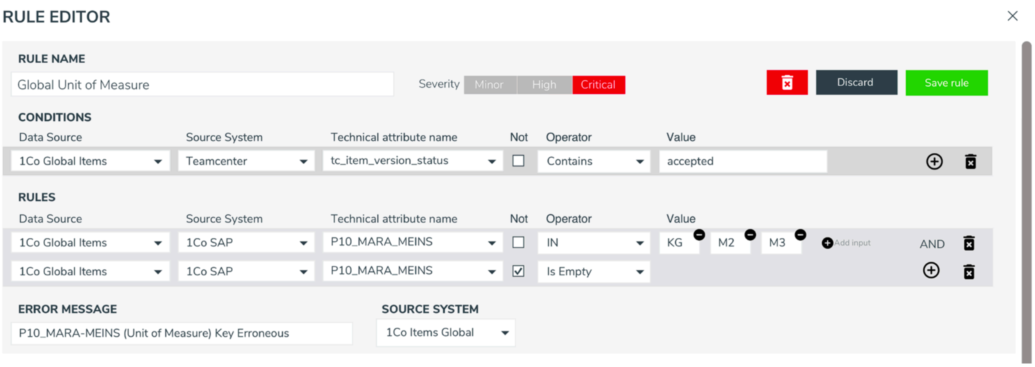


Figure 7 *Example of a data quality rule*

Prioritize and classify data quality errors

Data quality errors need to be prioritized so that the corrective and preventive actions can focus on the most critical ones. It also gives a better overall understanding when there are lots of errors. The situation might be completely acceptable even if there are thousands of minor errors as long as the number of critical errors is small. The errors can be categorized for example in three categories:

- Critical: Data errors which cause severe issues in business processes, stop or significantly delay projects or deliveries or cause for example compliance issues. These errors need to be corrected urgently.
- High: Data errors which cause additional work in business processes, but the amount of extra work is manageable. These are generally listed as backlog corrections.
- Minor: Data errors which are considered as lower priority items, which can be summoned and corrected without immediate effect to business operations. It's important to understand that errors which are currently classified as minor might become big issues in the future when there will be new use-cases for the data.

The example below shows a list of data quality errors in supplier data. The errors are classified as critical, high and

Supplier Data Quality

Basic information rules (Supplier Key attributes)

Critical Severity Error
High Severity Error
Minor Severity Error

	Character rules and restrictions	Length limitation	Additional information
SBM ID	System generated. If applicable, check if SAP account exists	40 numeric	If filled, SAP ID must exist within 7 days
SAP ID	System generated. If applicable, check if SDM account exists	40 numeric	If filled, SDM ID must exist within 7 days
VAT ID	No blank spaces in front or end	12 (14) numeric	Must not be empty
Supplier Name	!, “, #, €, %, &, /, (,),[,,±,],[,|,§,‸,$, blank spaces in front and end	40 char	Must not be empty
Supplier Address 1	!, “, #, €, %, &, /, (,),[,,±,],[,|,§,‸,$, blank spaces in front and end	40 char	Must not be empty
PO Box	No blank spaces in front or end	10 numeric	Must not be empty
City	No blank spaces in front or end	40 char	Must not be empty
Tax Codes	No blank spaces in front or end	12 (14) numeric	
Payment terms	Must be a dropdown list value	40 char	Must not be empty
Delivery(Inco) terms	Must be a dropdown list value	40 char	Must not be empty
Currency	Must be a dropdown list value	10 char	Must not be empty
URL	In format: www.address.suffix	90 char	
Email	Must contain txt before and after @-char	40 char	
Telephone number	Must be in international format with +prefix	20 char	Must not be empty
Bank account	Must be in international IBAN format with country prefix	40 numeric	Must not be empty
Capabilities	Must be a dropdown list value	x char	Must not be empty
Supplier Type	Must be a dropdown list value	x char	Must not be empty
Supplier Profile	Must be a dropdown list value	x char	Must not be empty
Hierarchy Level	Must be a dropdown list value	x char	Must not be empty
Hosting Main Contact	No blank spaces in front or end	30 char	Must not be empty

Figure 8 *Example of classified data quality errors*

Flexible error reports help in corrective actions

The data error reports must contain enough information to correct the data. The figure below shows an example of an error report where the errors can be filtered, grouped by tags and assigned for correction.

Severity	ErrorField	ErrorMessage	Tags	Product ID	Owner	Owning group
Normal, High, ...						
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545341	John Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87445342	Jane Doe	STG Mech
!	TC Item Weight	Item weight cannot be blank or zero	Teamcenter	87445343	John Doe	STG Mech
!	TC Item Weight	Item weight cannot be blank or zero	Teamcenter	87545344	Jane Doe	STG Mech
!	TC Item Weight	Item weight cannot be blank or zero	Teamcenter	87545345	John Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545346	Jane Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545347	John Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545348	Jane Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545352	John Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545356	Jane Doe	STG Mech
!	TC Purchase Group	Item purchase group must the from list of values	Teamcenter	87545364	John Doe	STG Mech

Figure 9 *Example of data quality error report.*

Data consistency checks across systems
Data consistency across systems is another data quality dimension. The example below illustrates two consistency checks between the product master (TC) and SAP.

- The TC status "Accepted for Production" corresponds to SAP status Z4. This mapping is defined in the consistency validation rules and mismatches are highlighted on this report.
- The description comparison is validating that the item description must be equal in both systems. Mismatches are highlighted on this report.

Severity	TC Item type	TC - SAP Status Comparison	TC Item status	1CO SAP X-plant status	TC - SAP Description Comparison	1CO SAP English description	TC Item description
All		All			False		
✓	DesignItem	✗	Accepted for Production	Z6	✗	GEAR ASSEMBLY	GEAR REDUCER
✓	StdDefItem	✓	Accepted for Production	Z4	✗	KEY	KEY BAR
✓	DesignItem	✓	Accepted for Production	Z4	✗	GEAR	GEAR REDUCER
✓	DesignItem	✓	Accepted for Production	Z4	✗	GEAR	GEAR REDUCER
✓	DesignItem	✓	Accepted for Production	Z4	✗	GEAR	GEAR REDUCER
✓	StdDefItem	✓	Accepted for Production	Z4	✗	WASHER	TAB WASHER
✓	DesignItem	✓	Accepted for Production	Z4	✗	GIRDER; MAIN GIR...	MAIN GIRDER
!	DesignItem	✗	Accepted for Production	Z1	✗	CUBICLE	ASSEMBLY
✓	DesignItem	✓	Accepted for Production	Z4	✗	GIRDER; MAIN GIR...	MAIN GIRDER
✓	DesignItem	✓	Accepted for Production	Z4	✗	GIRDER; MAIN GIR...	MAIN GIRDER
✓	DesignItem	✓	Accepted for Production	Z4	✗	SIGN	LOGO SIGN
✓	DesignItem	✓	Accepted for Production	Z4	✗	GIRDER; MAIN GIR...	MAIN GIRDER

Figure 10 *Example of a data consistency error report*

Rule setup is iterative by nature

Data quality assurance is an iterative process, which often requires patience and multiple refined re-takes. It is a common practice to iterate with the data quality rules as the initially defined rules are not usually giving a complete picture of the data quality. As the understanding of the specific data asset grows, the ruleset might require for example:

- additional logical clauses to take into account special conditions
- focusing specific rulesets only to specific data subsets
- refining the accepted list of values

Consolidated / Non-consolidated data

An enterprise logical data model data model is an abstract model that organizes elements of data and standardizes how they relate to each other and how they relate to the properties of real-world entities. Products, vendors and customers are all typical examples that are used in multiple business processes and multiple IT systems.

A consolidated data model shows which fields or attributes are synonyms for each other in different systems and allow quality management across the systems by for example verifying the consistency of the data.

The example below illustrates a consolidated data model for product data. The PDM system is the master for product data and the data is shared to the ERP and a third receiving system. When considering the consistency quality dimension, there could be for example the following data quality rules:

- Validate that all relevant records exist in all three systems
- Validate that the values of key attributes are consistent in all three systems. The consolidated data model helps in this. The PDM system has for example an attribute called "Product Type". The same attribute is called "Material Type" in the ERP and the third system. This reference data is well managed and there are 5 allowed values regardless which system we are talking about.

Systematic management of reference data is one of the fastest ways to improve data quality across systems.

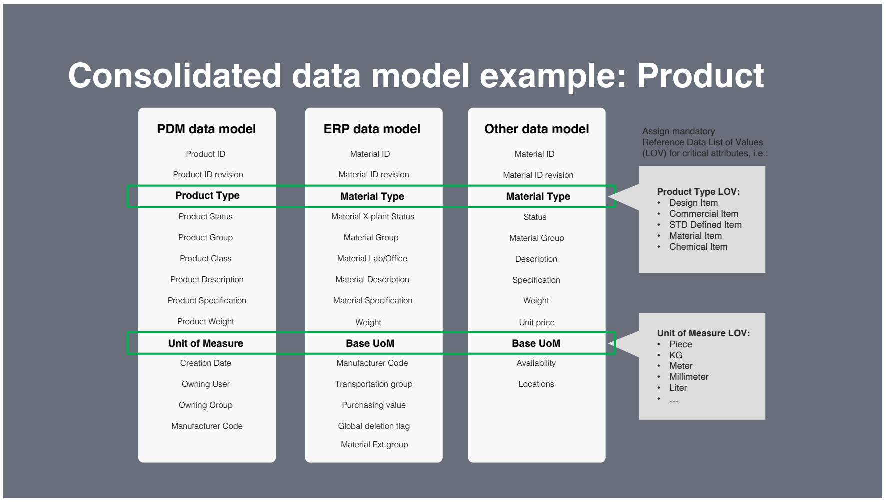


Figure 11 *Example of consolidated data model for product data*

Defining data validation cycles
After the data quality rules and the scope for which they are applied are defined, data validation cycle needs to be defined. Data validation refers to the process of applying the data quality rules to selected data sets whereas data validation cycle refers to how frequently the validation is repeated. The validation cycle depends on the adjacent data processes of the company and the following aspects should be considered:

- Data refresh cycles. The amount of data that needs to be validated will have impact on the validation cycle. If the data is transferred to external quality monitoring systems, a typical approach is to do an initial data load and subsequent delta transfers where only changed records are transferred.
- Data correction cycles. The data to be investigated needs to be validated in synch with appropriate data correction cycles.
- Data validation needs to be synchronized with other possible resource intensive activities in related systems.

The validation process should be automated and produce the needed understanding of the quality status.

5 Manage Corrective And Preventive Actions

5.1 Key Principles

Our key principles for data quality management are "first time right" and "data from previous process step can be trusted". If we detect an error in data, either by business user during business process or by data quality control, it means that we have failed to deliver according to these principles. Data errors cause unnecessary costs and often hurt customer satisfaction. All detected data errors can also be signs of hidden problems that can cause or have caused many similar errors.

Therefore, a systematic corrective and preventive actions management is a critical to maintain and improve data quality. A well implemented and documented process will also ensure regulatory compliance which is required in many businesses and especially when handling personal data.

In many companies the data user either corrects the error themselves or calls the person who created the data and gets the data corrected.

- If the error is corrected by data user in a downstream system / process only, the source data will remain as-is and the problem continues hiding under the radar. No systematic preventive actions are implemented, and the company continues to lose money due to rework.
- Therefore, it is important to correct the errors where the data was originally created. On top of ensuring a sustainable correction, it is important that the original creator gets feedback from the data users in the downstream process steps.

5.2 Processes, Roles And Governance

For the corrective and preventive actions to be effective, we need to have clearly defined processes, roles and governance. Also, a supporting tool to log and manage errors and defined actions are needed. The follow-up of corrective and preventive actions belongs to the business governance bodies.

Important roles

Data owner is an important business role. The data owner is typically ranked high in the organizational hierarchy. The data owner is responsible to have data managers nominated in their organizations and that sufficient time is allocated to the data error management. Data owners ensure the needed resources, the proper implementation of the corrective and preventive processes and enforce the ways of working so that the data creators and users accept these processes in practice.

Data managers in the business have a key role in the process. They are responsible for the analysis of errors in their own area, process or unit and also for the planning and implementation corrective and preventive actions. They will manage the corrective and preventive work of data users and data creators.

Data domain lead from the centralized data function facilitates and coordinates data error management in his or her own data domain and acts in close collaboration with the data owner and data managers from business. The data architect is responsible for the data modelling and data standards. The exact role split between the data domain lead and data architect varies in different organizations and sometimes the same person has both roles for a given data domain.

Chief data officer ensures that the processes for corrective and preventive actions are defined and that feasible supporting tools are available. The process for corrective and preventive actions should be same for all data domains, but with domain specific implementation guidance. While a common process enables effective error management across domains, the domain specific implementation guidance takes the variations in data processes into account.

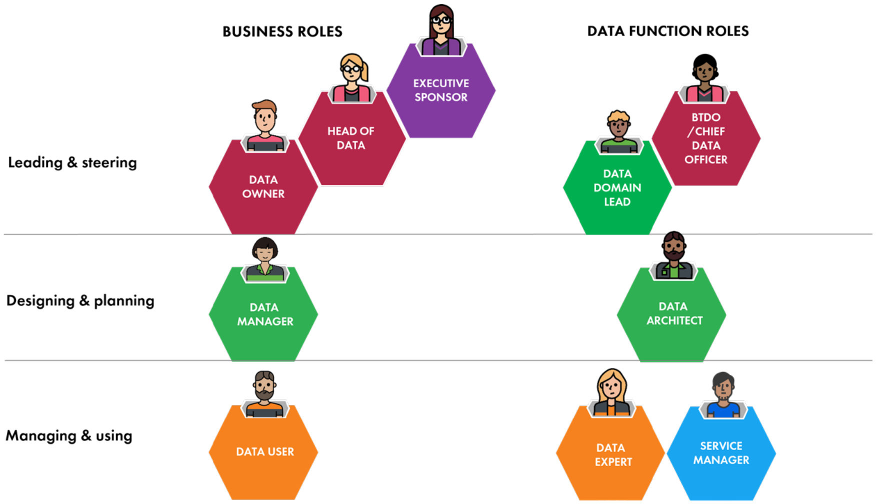


Figure 12 *Business and data function roles*

It is also important to clearly allocate enough authority to each role in respect to data error management. For example, what kind of decisions for preventive actions can be made by data manager and what the escalation paths are. Otherwise, the process will be too slow.

Data quality belongs to the agenda of business governance

The follow-up of data quality belongs to the agenda of business governance bodies. As discussed above, the data owners come from business and are ranked high in the organisational hierarchy and are naturally members of business governance bodies.

It is the data owner's responsibility to make the status and trends of data quality transparent in their business governance bodies. To enable this, the relevant KPIs and dashboards need to be defined so that they give an overview of what the quality is at the moment and how it is developing over time.

In addition to the current quality status, the business governance bodies should of course follow the actions that are targeting to improve the data quality. These actions can include improvement projects and continuous improvement.

The figures below show some examples of data quality KPIs and trends.

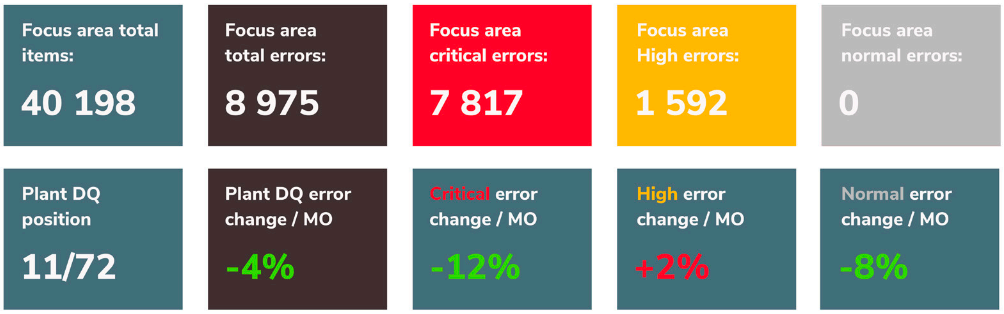


Figure 13 *Example of a data quality KPI describing the current status*

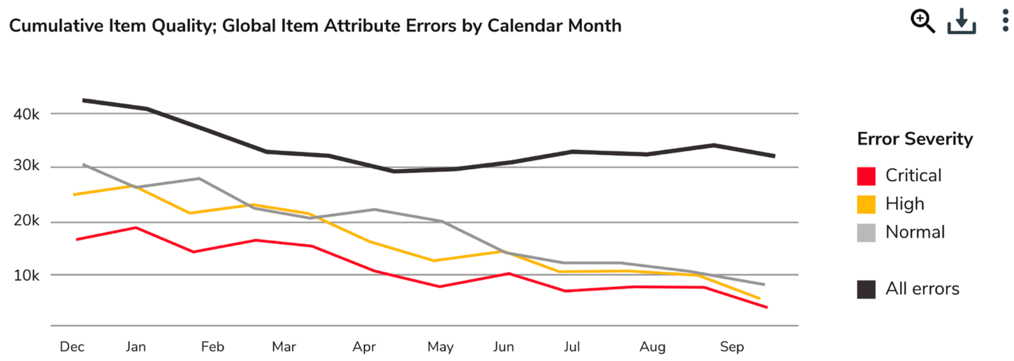


Figure 14 *Example of a trendline report demonstrating positive development*

5.3 Corrective Actions

When we have identified errors in data, we first need to minimize the negative business impact and to correct the errors as soon as possible. If the errors seem to be repetitive, we need to also plan for preventive actions.
The key tasks to be covered in corrective action process

- Log and assign identified errors
- Minimize negative business impacts
- Review errors and assign correction of data
- Confirm correction
- Close case or continue to preventive actions

Log and assign identified errors

Errors can be identified by data users or by data managers monitoring data quality reporting. In both cases first thing is to log the found data errors to the selected task management tool and to assign the errors to the data manager in charge of the data in question.

Some advanced data quality monitoring tools support assigning of errors directly in the monitoring tool and automatic notifications to the persons who should coordinate the corrective actions (or execute corrections by themselves). The figure below illustrates an example of weekly error report subscriptions.

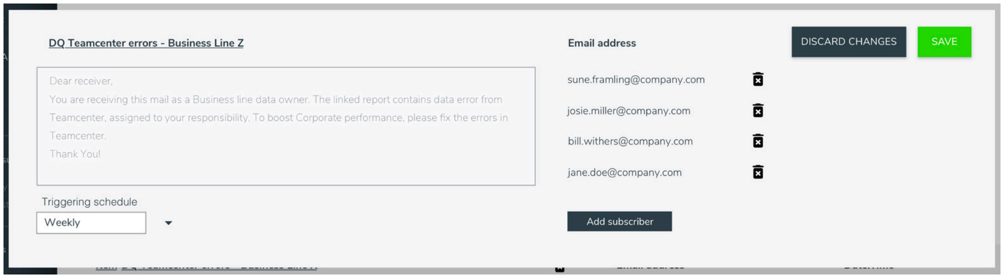


Figure 15 *Example of a weekly error notification configuration*

Whatever tool is selected, the most important thing is to ensure prompt actions to any identified errors. This can only be ensured if data managers have been allocated sufficient time for their role. The importance of business assuming ownership of good data quality cannot be emphasized enough.

Documentation of errors and their volumes, error correction and preventive actions provide valuable information about data quality and data management performance.

Minimize negative business impacts

When there are errors in data, the data user and data manager should evaluate the possible impact on business. If there is a risk, the data in question should be marked "under investigation" or similar and all transactions using the data should be put on hold to avoid more damage to the business. As an example, purchasing puts on hold orders for components until engineering has completed all needed information. If somebody else is using the same data from engineering, they should also understand that there are data missing and should be at least very cautious when using the data. That is why we need to also mark the data records with identified errors.

Review errors and plan and assign correction of data
The errors need to be reviewed by data manager. This is to understand the scope of correction needed and the need for later preventive actions. Scope can vary from adding a single missing value to an individual record to correcting hundreds of rows of data. Data manager plans and assigns the correction.

If there are only a few records to be corrected, the work can be assigned to the team that originally created the data, if it was a human error or to for example global data service team member who can correct the other types of errors (integration etc.).

If we have hundreds (or even more) of records of data to be corrected, we typically need to plan a batch correction. This could contain automated corrections combined with manual correction and review. With larger cases, a formal project may be needed.

Confirm correction
Important step in the process is that data users confirm the correction of errors they have identified. That also gives confirmation that the corrected data has reached the system where errors were identified.

In the case of batch correction, we need to plan and run automated data quality checks to the batch and in major cases let data users validate the data in non-production environment before releasing the corrected batch to production environment.

Close case or continue to preventive actions
If the error was type single error, case is closed after correction has been confirmed. When there are repetitive errors, preventive actions should be taken to avoid the issue from occurring again.

5.4 Continuous Improvement & Preventive Actions To Tackle Repetitive Errors

Organisations can achieve amazing results with minimum investment by doing continuous improvement systematically. Ideas are typically initiated by data creators and data users, but they can also result from a root cause analysis of repetitive quality issues. Data managers need to review the reporting on regular basis and identify areas for preventive actions.

These ideas and tasks should be collected and followed up systematically. Categorization, prioritization and formulating these into actionable and manageable pieces of work is essential and give an overall picture of needed actions and their size

Examples of repetitive errors:

- integration and erroneous transformation issues can be identified, when data in the source and target systems are being compared
- constantly repeating data errors or empty values can be caused by human error. Data managers will analyse the root cause. In these cases, corrective action could be additional trainings or improved work instructions.
- repetitive delivery delays may refer to missing or erroneous data, which can be laborious to cover in later phase due to missing knowhow or missing access to relevant systems. One of the actions could be to enforce the importance of data quality in the data creation phase.

The key here is to facilitate and support teamwork and build a systematic approach for continuous improvement. The problem analysis, design and implementation of these improvements is a collaboration between business, the data function and IT. The roles involved in this typically include data managers from business, data domain lead from the centralized data function and the solution architects from IT.

There are well documented problem-solving methods like 5-why, fishbone diagrams and cause-and-effect diagrams. Regardless of the method, the following tasks should be covered:

- Identify and evaluate repetitive or systematic data errors or occurrences where data is completely missing
- Analyse the root causes and select preventive actions
- Implement actions, monitor results and adjust if needed
- Document the case in the selected task management tool

Identify and evaluate repetitive or systematic data errors

Repetitive errors are mainly identified in data quality monitoring, or they come from the corrective actions process. Sometimes data users also prompt data managers about repetitive errors. In both cases, the data manager has to first define the problem for the next steps. This can be done by answering questions like what, where, when, by whom, etc. Once the problem has been defined, the business impact of the error should be evaluated.

Analyse the root causes and select preventive actions

Very often the root cause analysis is somewhat neglected, and we end up taking aspirin instead of curing the illness. Sometimes root causes are quite easy to find and to evaluate. If new employee is creating errors, induction, training, instructions, or support might need attention. In some other cases we might need systematic testing to ensure we have found the root causes. If we have a software bug or integration error, it's sometimes difficult to identify.

Once the root causes are identified, candidates for preventive actions need to be planned. Typically, we can find many potential candidates. We need to prioritize based on the business impact of the error and the cost and effectiveness of the proposed actions.

If the elimination of root causes requires wider changes to processes, systems or organizations, the issue should be escalated to relevant data governance body for starting a separate data quality project. Also in this case, some of the actions could be done inside the corrective and preventive action process.

Some examples of typical preventive actions:

- Process and instructions updated to give better guidance in data creation
- Training plan for data creators updated to better cover for induction training of newly nominated data creators
- User interface modified to validate data
- New reference data brought under change control
- Review and/or approval steps added to data process
- Sometimes introduction of completely new systems may improve data quality. It could be for example possible to collect some data automatically with a new tool so that a manual data entry could be avoided altogether.

Implement actions, verify results, and adjust if needed

As a part of planning the implementation of the selected actions, we need to estimate the targeted results and set up monitoring to ensure that we reach them. If we fail to reach the target, we need to go back to see if our root causes are valid, if the selected preventive actions were correct or if the implementation was good enough and adjust where needed. Then monitor the results again.

Document
Preventive actions need to be documented from identification to verified results in the selected task management tool.

Improvement projects
If the improvement requires significant resources, a separate improvement project can be established. The goal of the project can be to increase data quality to the next level and typically these changes include changes to the existing systems (or even completely new systems), processes and changes to the role descriptions.

Examples of these kind of projects include projects where item data is enriched for spare parts sales by for example photographing big number of items, different kinds of data cleansing projects or data harmonization projects where lots of data is updated to meet the data standards.

These data quality improvement projects follow the normal project practices and project governance of the company.

6 Proven Best Practices For Data Quality

The key principle in data management is first time right. We need to do everything to ensure that data is created according to standard.

In this chapter we will provide a set of proven best practices for ensuring first time right in data operations. These best practices are applicable to any data management environment and business.

Key best practices described
- Build and enable active data teams
- Manage reference and master data centrally
- Design applications to support successful and efficient data creation
- Ensure quality of end-to-end data integrations
- Beware of conceptual mismatches when reusing existing data for new use-cases

6.1 Build And Enable Active Data Teams

The best person to create first time right data is the person who has the best knowledge of the subject. In the case of customer data, it is the account manager working with the customer. In the case of part data, it is the engineer designing the part of a purchaser who is buying the commercial part. Then there will be additions to the data along with the data flow from originator to the final user of the data. High-quality spare part data is a result of many persons creating, extending and modifying the part-related item data along its flow from engineering to spare part business. These people represent various business lines, disciplines, geographical areas and work according to many business processes.

The key success factor for high data quality is to ensure that all these people understand how important their input is for the other users of data and for the overall business, not only for their own unit. To help in this, it is important to invest in teaming data management people across the company. When people know each other, quality will improve through natural co-operation. Data domain lead and corporate data owner for the data domain are the key roles to drive this end-to-end understanding of the importance of data. Data domain lead is typically the leader of the corporate wide virtual data team.

Some critical tasks in enabling virtual data teams:

- Organize and nominate virtual teams, let people know they are part of a team
- Invest in team meetings and social happenings, virtual and face to face
- Inform and coach virtual teams systematically
- Report data quality and improvements, cases etc. to team
- Engage team members in improving data quality

Leading a team requires typically a person or small team of persons to lead and facilitate the team, otherwise it loses direction and momentum quite fast. We also highlight a few other success factors, time, induction, support and availability of related instructions.

Allocate time for creating quality data

In most of the organisations, recognition is based on target settings. It is a good idea to have data management activities part of target setting for all involved people. In the target setting cross-functional / cross-process targets should be included to allow using time and effort.

Induction to the team

Induction and training for all data creators and data users should provide an end-to-end understanding of where and by whom the data is created and where, for what and by whom it will be used afterwards. The other obvious aspect is the practical instructions for data creation. Introduction of the virtual data team is also part of induction. Data team will also provide continuously training when there are for example changes that affect working instructions.

New IT systems and other major transformation require planning and training in advance to ensure that there will not be data quality problems. A team approach helps there too, as we know the people involved.

Support for virtual data team members

As the team members are often spread around the organization, it is important to invest in support. Types of support include:

- Support from other team members
- Subject matter expert support, e.g. hydraulics component expert in part data creation
- Data modelling support, e.g. Modelling of data assets that have not yet been modelled or explanations on data attributes
- Tool usage support, how to e.g. add inventory related data in ERP

The support often requires local data admin resources to be included in the team to support data creation according to standards.

Make data processes, instructions, data models and data standards available for data team

Data architecture, data flows, data processes and instructions, data management roles and data standards are daily references and should be made available for virtual data team members easily. In the best-case, the person who is creating data gets guidance directly in the user interface (wizards or additional online guidance tools).

6.2 Manage Reference Data And Master Data Centrally And Automate Distribution To Target Systems

Create and manage reference data centrally
Reference data is the data that we often get from the drop-down lists. Examples of reference data include units of measure, product groups, customer segments or customs tariff codes.

Reference data should be created and managed centrally to ensure that they are consistently used across different IT systems, including reference data history. This is typically one of the first parts of implementing master data management. Master data management tools can help in managing reference data and can be used to distribute it to all consuming systems. This way we can be sure that the possible values in drop down lists are consistent in various systems.

New technologies and available external data enable widening the scope for reference data. For an example, street addresses can be selected from a list using search engines like Google maps. This also ensures that the city, postal code, and country are correct.

Manage master data
Master data is the shared data about customers, vendors, products, and other key data objects that are commonly used in many systems. It is important to centrally ensure that master data is created and maintained properly and then distributed to all relevant systems using the master data. This way we can ensure that the core of the data is of good quality and consistent across systems. It is unfortunately common that the processes for creating new records and maintaining the existing ones are unclear or inefficient and result in poor data quality. These processes and workflows should be critically evaluated when establishing master data management. Also, history of master data needs to be stored to enable utilization of older transactional data and older documentation.

As we discussed earlier, the creators of modifiers of data are typically not part of one organization or in one location. Therefore, it is worth noting that "managed centrally" means shared processes, standards and that the creators of master data have to be part of a virtual data team.

For the most critical data like supplier approval or supplier bank account, we need to establish review and approval workflows to enable companywide control.

Effective management of master data in a globally operating company requires global master data system. Many companies have already product data management (PDM) system to manage item data and product bill-of-materials and changes to this data. PDM system supports engineers in creation of the data and it distributes the data to ERP and other systems. We should have similar master systems for other master data entities as well.

6.3 Include Data Reviews And Approvals In Data Creation And Update Processes Where Applicable

In many cases we need to have reviews and approvals in our data processes. In some cases, this is the way to ensure that the content represents the reality and suits the company targets. As an example, if we are defining commercial parts in engineering, it is a best practice to review the selection by subject matter expert in the technology to ensure that the data accurately describes the part.

As another example, the process for creating new and updating existing supplier master data records needs to be designed so that it prevents frauds. Bank accounts must be approved by a few nominated persons to assure that the money will go to the supplier and not some other bank account.

In some cases, we need an audit trail for example for regulatory compliance reasons, and the reviews and approvals must be documented.

6.4 Design And Use Applications That Support Successful Data Creation

Provide reference data and core shared master data about customers, products
We have already covered the importance of centrally managed reference data and core master data such as customers, vendors, employees or products. These data should be provided automatically for all systems needing that data. As an example, core item data is provided for ERP system and the local data creator only needs to add the needed additional data for the logistics purposes.

Select applications that provide modern support for data creation
Modern application architecture enables flexible support for the creation of the data. User interfaces can for example suggest values in the same way like search engines do, they can provide flexible drop-down lists, autocomplete fields and so on. Good user interfaces also include data validation capabilities before saving. We can even add machine learning and other apps to into the user interface layer to support data creation and validation. It is important to review the data creation aspect when selecting applications.

6.5 Ensure Quality Of End-To-End Data Integrations

Data integrations play a big role in data quality. They transfer data between operational systems and to data platforms or data warehouses from where the data can be consumed to different use cases.

This flow through systems and many integrations needs to be carefully planned. It is very important to document and describe these data flows and share the documentation to different stakeholders. There is often need for different levels of integration documentation, for example a conceptual documentation which describes the overall flow and a separate technical document that describes for example the detailed field mapping and transformation rules.

We have seen many times that this end-to-end design is missing when applications in the middle of the flow are changed causing severe and systematic data problems. For example, in a major ERP project the flow of product data from product data management system to a sales configurator, from there to a CRM and finally to an ERP system was not properly defined. Because of this unclarity, there would not have been a single successful customer order or delivery when the new ERP was about to go-live.

The key is to plan the flow of data. The "physical" flow of data must be consistent so that data elements like part ID flows from system to system and the content is not corrupted but it’s not sufficient if the logical or conceptual flow is broken.

Understand the limitations of the data
When we are using data along the data flow to new uses, like in analytics, it is important to understand the limitations

of the data. Even if the data might look good and valid, we cannot be sure that it is created to take our new use case into consideration.

One company had manufactured their products in their own factories. They decided to move to outsourcing big part of their manufacturing. Soon they found out that the quality of products was not at all what it used to be. The drawings and the other specifications had been sufficient for own factory, as there was a lot of tacit knowledge about the products. The same drawings and specifications (data) were not sufficient for third party suppliers.

6.6 Beware Of Conceptual Mismatches When Reusing Existing Data

Sometimes instead of using the same attribute (e.g., Item specification) in multiple business processes, it's better to create a separate attribute for different purposes. By separating the data, it is easier to find out the ownership for each of the data elements and it is clearer for business department and IT to understand, what kind of changes are targeted into each attribute.

7 How To Cope With Changing Business Requirements For Data

7.1 Problem statement

The definition of data quality is built on an assumption that we know the intended use of data, the data consumers and their expectations for the data. To manage the quality, we turn these expectations into data standards.

When the business evolves or is going through a transformation, it is quite typical that there will be new data consumers with new expectations and the data no longer meet these expectations.

Let's consider a company that delivers process plants to customers. Each plant is engineered to order. Most of the equipment are bought with specifications. When designing a new plant, engineering defines parts with requirement specification documents and some item data. In the case of purchased part, the item data contains for example item number, item description and reference to a specification. This information is intended for use in engineering, purchasing and assembly line. Part can be ordered from vendors and identified on assembly line with the item data and the specification.

As long as the company is only delivering process plants, the item data is not needed after the project. New projects can design new parts and create new items for them. Data consumers, engineering, sourcing and assembly line, can do their work based on the project specific data. Data meets the expectation of data consumers in the intended uses.

When new data consumers and new intended uses arise

Let's say that the company decides to expand its business into spare part sales for already delivered plants. We have now two new data consumers and two new uses for the item data:

- Customer wants to identify the correct part and find price & availability for the part

- Spare parts sales need to identify what part do the customer need and provide price and availability for the part

Identifying the right part is the first big challenge. Most of the information for finding the right part is in the requirement specifications of the delivered plant. However, the installed equipment might have been updated and the information in the original engineering specifications might not be valid anymore. It takes the spare part sales time to dig out the part needed from the specifications and for double checking with customer to ensure the part is right.

If there is a desire to build an online solution where customers can order spare parts as a self-service, the situation is quite hopeless with existing data. We would need to have:

- Customer equipment structure (after possible modifications)
- Available spare parts linked to the above installed equipment structure so that the correct part is easy to identify. The online spare part portal would additionally need new attributes like pictures and item descriptions which are intended for customers.

Some data is still missing to satisfy the needs of the customer: price and availability. Once the spare part sales have identified the part, they can contact sourcing to get the cost and delivery time for the part and then calculate the selling price. This kind of manual process is not feasible for the online spare part solution as it takes way too long compared to the expectations of the customers who would like to order the part with a couple of clicks and know how much the part will cost and when it will be delivered. Data does not meet the expectations of the data consumers in the new intended use.

These kinds of challenges are also very typical in large business transformation projects, for example when implementing a new ERP. It is very typical that there will be many new expectations and requirements for data. To make the situation more challenging, changing big amounts of data will typically require a major investment and can take months of calendar time.

This raises a key question: Could we prepare for the changes in business in advance?

7.2 Ways To Prepare For Changes In Business

There are several ways preparing for changes in business.

- Use business strategy as a starting point
- Use Business Technology Design as a method to define business value chains and consider product or service life cycles
- Implement consistent data architecture & standard for critical data

Requirements for data should be considered when revising business strategy. Implementing new strategies typically require new capabilities and at least updated data to be available. If data is considered during strategic planning as a prerequisite, it is possible to address data aspects before they become bottlenecks.

Use Business Technology Design to understand business value chains and consider product lifecycle
Business Technology Design is a systematic method for understanding business value chains and to identify the capabilities, including data, that are needed in different parts of the value chain.

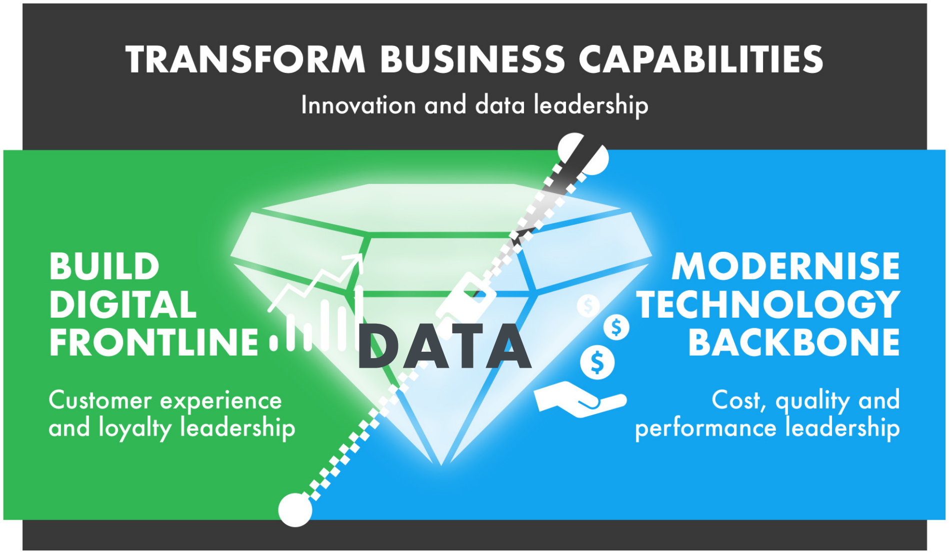


Figure 16 *Data is in the center of business technology*

Let's continue with the example described above. The company could have considered the complete lifecycle of the plant and identified that there are business opportunities in the operate & maintain phase even if the company would not yet enter into the spare part market. Keeping this in mind, it would have been possible to invest a little bit more in the data and start to create the structures of the plants with spare parts linked to the structures instead of creating new items for each project.

This would have made the life easier for the data users even in the project delivery business, but it would have been a major advantage when entering in the spare part business. Had it been properly planned into the process and tools, it would not have added much (if any) additional work compared to the way data was created separately for each project.

Another example: product lifecycle of a car
Let's think about the product lifecycle of a car. The car is

- first designed
- then manufactured
- then sold to a customer
- after that it is used and maintained by a customer
- finally, it is recycled at the end of the lifecycle

Data about the car is created and used by engineering, production engineering, sourcing, manufacturing, marketing, sales and delivery. Once the car is sold, the user will need some data, maintenance will use and create data. Recycling also requires data about the car. We would need different views and content of data for each data consumer, for example:

- engineering has an "as-designed" -view with all technical details and CAD models

- customer need user instructions for using and maintaining the car
- maintenance would need e.g. spare part data and maintenance instruction
- recycling needs specific information about removing and recycling the battery

Considering the complete lifecycle of the product helps to plan the data so that it is as future-proof as possible. If we do not want to consider the possible future data needs, the minimum approach is to manage data in a way that there is a data standard and data quality management that ensures data conformance to the standard. If data is similar consistently across the company, it provides a much easier starting point for any additions and adjustments later on.

8 Implementation Tips For Data Quality Management

This chapter summarizes the key steps for establishing data quality management in three different scenarios:
- Introducing data quality management to ongoing business operations
- Data quality management in data migrations and cleansing projects
- Data quality in business transformation programs

8.1 Introducing Data Quality Management To Ongoing Business Operations

Company wants to improve data quality in ongoing business operations. There is existing data and it is being created and updated by many people in many different systems. Some documentation and work instructions might exist. In these kinds of situations, it's typically not possible to start with a clean table approach, because there might be some critical and urgent data issues to solve.

The steps
- Secure management and business commitment by implementing data quality aspects into the business process governance structure and strategy.
- Map the data assets and select the ones with critical issues or which needs biggest improvements.
- Nominate and train data owner and data managers.
- Gather documentation and interview key experts to fill in the gaps in the documents to understand the data flows, processes, and business requirements for this data. Document these as minimum viable data standard document.
- Profile the data; what data do we have, how much, in which systems? What is the initial data quality?
- Prioritize the data quality requirements, plan and implement quality control activities and manage the corrective and proactive activities.
- Proceed to monitoring, assessment and continuous improvement, and/or widen the scope to next data assets.

8.2 Data Quality Management In Data Migrations And Cleansing Projects

Cleansing and migrating data is a special case for data quality assurance. For example, in the case of a merger or an acquisition, there is a need to quickly cleanse and migrate data from one system to another to enable efficient operations.

Again, we need to have the minimum set of foundations in place:

- Data standards (both source and target).
- Nominate data owner and data managers.
- Profile the data to understand the quality of the data against the data standard.
- Analyse the needed changes to the data and the needed cleansing activities.
- Plan the processes, instructions, roles and quality assurance with controls.
- Perform user acceptance testing, especially in case of larger amount critical business data.

8.3 Data Quality In Business Transformation Programs

Too many transformations have encountered serious delays and suffered from negative business impacts for not having the foundations of data management in place. If the quality assurance is not implemented for data migrations and the new or modified data processes are not properly considered, it's more than likely that there will be major problems.

When preparing a transformation project:

- Use business strategy as a starting point
- Use Business Technology Design as a method to understand business value chains and the needed capabilities. Consider the whole product or service life cycle.
- Implement consistent data architecture, data flows, data standards, processes & working instructions
- Define data related roles, nominate data owners and data managers

Include the planning and implementation of data quality assurance in the scope of the program.

Who We Are

The Business Technology Forum (or BT Forum) is a non-profit professional organisation consisting of a community of forerunner companies, and public organisations collaborating according to platform economy model.

The BT Forum provides business and technology leaders with an open-source technology management framework called the Business Technology Standard. The BT Standard consists of best practices, models and tools developed together with the BT Forum community in order to plan, build and run information technology in today's technology-driven business world.

The BT Forum coordinates the development work within the community members and publishes an upgraded version of the BT Standard twice a year. In addition the BT Forum also organises events and conferences, publishes educational materials and offers training courses to advance the business technology management profession.

Follow Business Technology Forum on Linkedin to stay updated with the latest news!

Contact us: info@btforum.org

BUSINESS
TECHNOLOGY
STANDARD

www.managebt.org