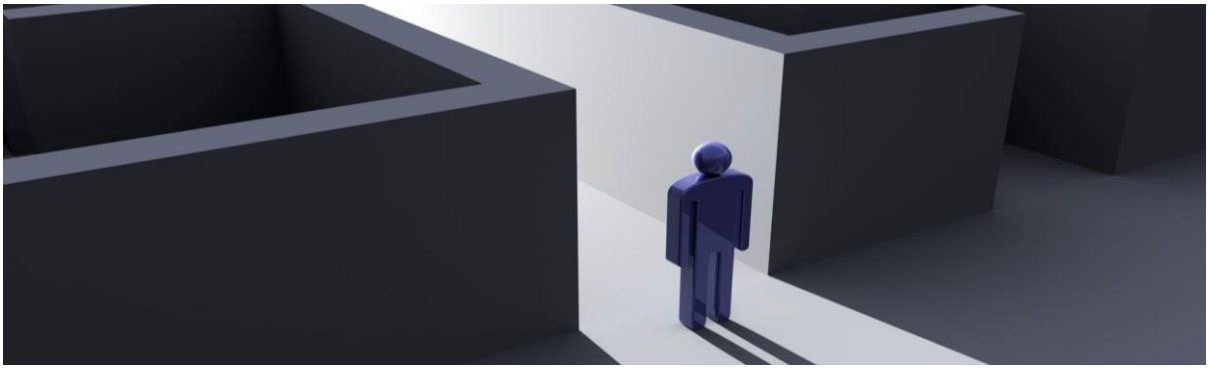

Identifying Perceived Barriers and Enablers to manage Data Quality Awareness

DAMA-NL Data Quality Working Group

Aris Prins, Kamal Ahmed, Peter van Nederpelt, Laurens van der Drift, Marco Heij, René Wiertz, Merijn Douwes, Dirk Hueting, Andrew Black, Aylmer Hansen

data quality awareness, perceived barriers, enablers

Introduction

In today's era, many of the services, products, and processes we interact with, revolve around data: from financial transactions when we purchase an item on the internet, to receiving instant recommendations to help us expand our social networks. It is therefore not surprising that data are considered by many as the 'new oil' of the 21st century, thereby referring to the value data can have. By further enriching these data and using insightful data analyses, for example in the form of predictions, the value of data can increase to generate profits, or save costs considerably. However, if these data are incorrect, they can be(come) harmful to an organization's reputation, impact processes and affect customers or society in general.

So, what is the damage an organization may suffer from if the quality of the data does not meet the requirements? Generally speaking, this boils down to determining the associated risk: the probability occurrence of an event multiplied by the severity impact of the event. For example, an incorrect, too low pricing of a popular product in a web shop will immediately lead to a measurable loss of turnover. Similarly, incorrect stock registration will result in unknown delivery times, possibly resulting in both a loss of turnover and reputational damage. In the case of repeated writing to deceased customers, this will mainly result in reputational damage and emotional distress to family members of the late customer.

Incorrect data thus can be harmful and have limited or far-reaching consequences. To manage the associated risks, data quality may need to be maintained or improved, for which the data quality must be evaluated, which in turn requires the organization to 1) become aware of its data quality, 2) remain aware of its data quality, or 3) raise the level of awareness. This interdependency is shown in Figure 1:

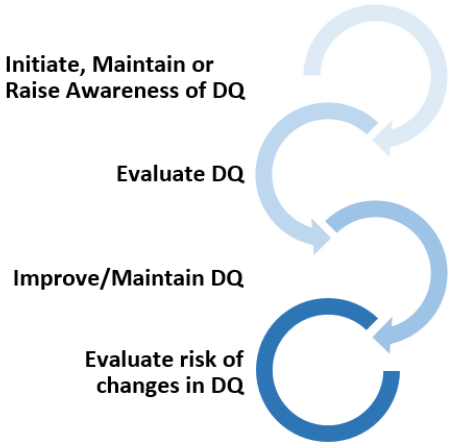

Figure 1: Interdependency between awareness of data quality and risks of data quality.

Awareness

However, addressing awareness is easier said than done, as there may be certain barriers occurring. For example, many employees in organizations are often focused on other high-priority activities and so are unable to prioritize the enhancement of data quality. In other instances, there may be a tendency for denial or a wrong perception: "we have no issues with data quality, so why should we bother".

Identifying and understanding barriers is important in order to facilitate effective interventions to initiate, maintain or raise the awareness of data quality.

We define awareness of data quality as an acknowledgement or realisation by everyone in an organisation, that their current skills and competencies are (becoming) ineffective, resulting in data of inadequate quality [1], [2].

This white paper is intended for those who want to improve data quality in their organization by raising the level of awareness. We first identify possible potential barriers encountered in organizations and then provide several practical enablers/best practices to overcome these barriers . Through this approach, we aim to increase the success rate of any data quality initiative.

Perceived barriers

In this white paper we identify[1] and categorize perceived barriers along three main pillars:

1. Barriers within individuals or teams, e.g. resulting from certain behaviors or attitudes
2. Barriers within Technology, e.g. resulting from the type or maturity of technology, and
3. Barriers within Processes set by Leadership, e.g. work processes defined by Leadership

It is not our intention to indicate an exhaustive list of all possible (combinations of) barriers. Instead, we focus on several of the most common ones we have encountered across different organizations.

[1] Identification and analysis of the most relevant perceived barriers can be achieved through focus groups, interviews, surveys or direct observations.

Barriers within individuals or teams

a. **Denial of any pain points** with data quality; perception: "all is good";
b. **Fixed mindset**: "we can't improve the data quality on the basis of our skills or knowledge, because there is nothing more to learn or improve";
c. **Inflexibility**: "we can improve, but we won't, because of too much workload, too many other responsibilities. This also relates to: WIIFM (What's In It For Me), if I do maintain/improve data quality?);
d. **Lack of trust/skepticism**: Not accepting other person's expertise, authority or views on the subject matter due to that person's lower level of credibility;
e. **Lack of ownership** for end-to-end data quality within, or across teams; ownership is currently only associated with the data quality within one particular team or business unit;
f. **Incomplete handover**: Flexible workforce (contractors and temporary staff) are hired on an incidental or high-priority basis to help fix data quality issues, but then may (suddenly) leave without performing a 'positive handover' back to the team, which results in a halt in the data quality awareness;
g. **Mental fatigue, frustration, or complacency altogether**: Frequent updates in systems or tools result in a mental fatigue, frustration, or complacency altogether: no corrective action or remediation action occurs;
h. **Difficulty to make tacit knowledge explicit.** "Explicit Knowledge" (i.e., documented knowledge) versus "Tacit Knowledge" (i.e., intuitive knowledge & know-how rooted in context, experience and practice, residing in people's minds). Tacit Knowledge is much harder to communicate or transfer to others, especially by less experienced staff. Data quality issues may be of the tacit type, requiring some form of translation to create data quality rules;
i. **Lack of realization** that one's current competencies and skills are not adequate for maintaining or improving the data quality;
j. **Incorrect assumptions or perceptions by employees:** systems will automatically remediate data quality issues;
k. **Paralysis by analysis**, where individuals or entire teams acknowledge data quality issues, but when it comes to remediating these issues, the task seems daunting altogether: no one knows where to start, or even how to start.

Barriers within Technology

a. **No traceable documentation.** Systems get connected or upgraded without any traceable documentation (lack of "Explicit Knowledge", so maintaining or raising the data quality becomes almost impossible).
b. **Incomplete technology roadmap**. Unclear how the technology roadmap will be affected by any changes in (future) regulations, standards, policies, etc ... resulting in either no action, or only a reactive approach.
c. **No Data Quality monitoring system software has been installed:** it is impossible to measure data quality constraints and to remediate the found issues.

Barriers within Processes set by Leadership

a. **Incomplete priority setting.** Leadership prioritize end-to-end data quality on a strategic level, but not on an operational or tactical level.
b. **No explicit knowledge documentation process**. Lack of process for capturing "Explicit Knowledge" within organization.
c. **No tacit knowledge documentation process**. Lack of process for capturing "Tacit Knowledge" within organization.
d. **No learning organization**. Leadership discourages failing, experimentation, developing prototypes, proof-of-concepts, etc. to improve data quality. Essentially, the organization is not a "learning organization".
e. **Motivation discouragement**. Leadership does not celebrate and/or acknowledge any (intermediate) achievements of teams working on data quality.
f. **No data quality mindset.** Leadership is not communicating about the relevance and consequences of data quality work in relation to requirements posed by regulation, audit, stakeholders, clients, etc.
g. **Unknown or unacceptable expectations** raised by the leadership team when asking their staff to take on an extra workload. For example, the following can be(come) dissatisfiers for staff: indefinite time period to take on extra workload, undefined key success factors for data quality project, possible job losses in due time, unknown number of new hires in the pipeline, etc.
h. **Inadequate support of staff**. Leadership is not (adequately) supporting staff in terms of learning & development opportunities, e.g., related to fundamentals in "digital and data literacy", or more advanced topics; attending online courses in one's spare time is also not the solution if a healthy life-work balance is the norm.
i. **Lack of confidence in staff.** Leadership does not abide by the principle of "having confidence in the expertise of employees and giving them autonomy to make decisions".
j. **No approach yet to establish current DQ Maturity level**. So how to know what aspects of DQ to improve and why?

Key questions or guidelines to address barriers

Which key questions or guidelines should we keep in mind to address data quality awareness in order to maintain or enhance data quality? We believe the following three are relevant:

1. **Why maintain or enhance data quality?** Possible answers can be as already mentioned in the introduction. We may lose money, reputation may be damaged, processes may not work as planned, but also because we have to measure and maintain, or enhance the quality to comply with regulatory or governmental directives or policies. Upon receiving reports from external organizations, the reports need to be correct. Everyone who contributes to maintaining or enhancing the data quality should address the 'why question'.
2. **How credible is the person providing advocacy for maintaining or enhancing the quality of the data?** Are persons on the receiving end having trust in the person who is indicating that there is a problem, with a need to change, or are there any concerns with this trust?

3. **What is the priority for maintaining or enhancing the data quality?** What is the rush for improving the data quality? Which priority applies?

Enablers

With these guidelines in mind, we can turn the mentioned barriers in enablers[2]:

- **Leadership**. To have the right level of credibility, let leadership guide the communication (and awareness) of Data Quality;
- **Audit**. The internal audit department can help by making Data Quality more credible. For example, the internal audit department has to audit the data quality remediation process;
- **Quality Policy and Quality Objectives**. Let leadership communicate periodically about Data Quality by defining a policy, showing and following the roadmap, setting clear objectives and celebrating (intermediate) victories;
- **Team**. Assign a special team with roles in the organization for implementing and monitoring Data Quality, such as Data owners or Data stewards;
- **Monitoring**. Implement a proper Data Quality monitoring system with Data Quality rules and show discrepancies to prove Data Quality is important and has to be improved;
- **Prioritizing**. Implement a prioritization methodology to prioritize which data source area, discrepancy type to solve first in relation to data quality. For example, risk based or because of regulatory requirements;
- **Use of coaching:** realizing effective behaviors in the workplace and e.g., identifying people's intrinsic motivations to maintain or improve Data Quality;
- **Use of mentoring**: improving people's (technical) skills for Data Quality in their day-to-day projects;
- **Explicit Knowledge**: use of Knowledge Management systems (to capture knowledge in official documents, Wikis, policies, standards, etc);
- **Workshops**. Make tacit Knowledge available in the form of workshops or "lunch & learns" to derive best practices from more experienced staff;
- **Handover.** Flexible workforce needs to perform a "positive handover" before continuing with the next project, for example in the form of: workshop, demo, walk-through sessions, manual, etc.
- **Job rotation**. Implement job rotation schemes for employees so they can take up roles related to Data Quality across other departments. This can help in sharing knowledge & experience across departments to improve the end-to-end Data Quality.

An overview of several barriers and enablers is shown in Figure 2.

[2] Identification and analysis of the most relevant enablers can be achieved through focus groups, interviews, surveys or direct observations.

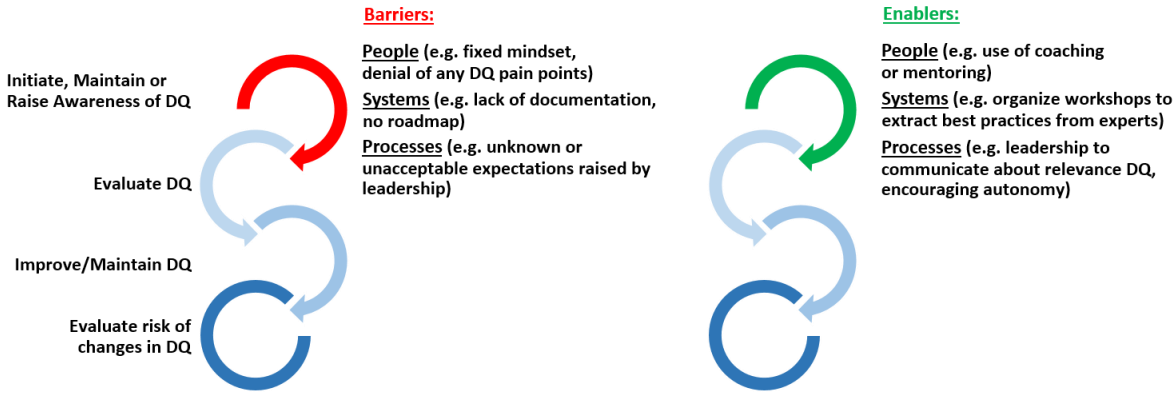

Figure 2: Several barriers and enablers for initiating, maintaining or raising awareness of data quality.

Concluding remarks

Improving Data Quality requires more than installing a Data Quality monitoring system and setting up a team; identifying possible barriers along the way, by means of focus groups, interviews, surveys or direct observations, is equally important. For each identified barrier, ask key-questions related to maintaining or raising the data quality:

- What is the relevance?
- What is the credibility of the person providing advocacy?
- What is the priority?

Associate each of these prioritised barriers with suitable enablers/best practices. Finally, monitor and analyse by means of focus groups, interviews, surveys or direct observations, if overcoming these identified barriers is:

a) helping to enhance Data Quality awareness in your organization, and
b) leading to more successful Data Quality initiatives.

Next steps

The DAMA-NL Data Quality Working Group welcomes any further thoughts on any of the topics raised in this white paper. Comments, suggestions and questions can be sent to info@premium.nl.

References

[1] *"Factsheet Awareness of Data Quality"*, version 3.0, DAMA-NL Data Quality Working Group, 16 January 2022.

[2] White paper, *"A management system for data quality – Towards a More Solid, Coherent and Standardised Approach"*, DAMA-NL Data Quality Working Group, 10 March 2022.