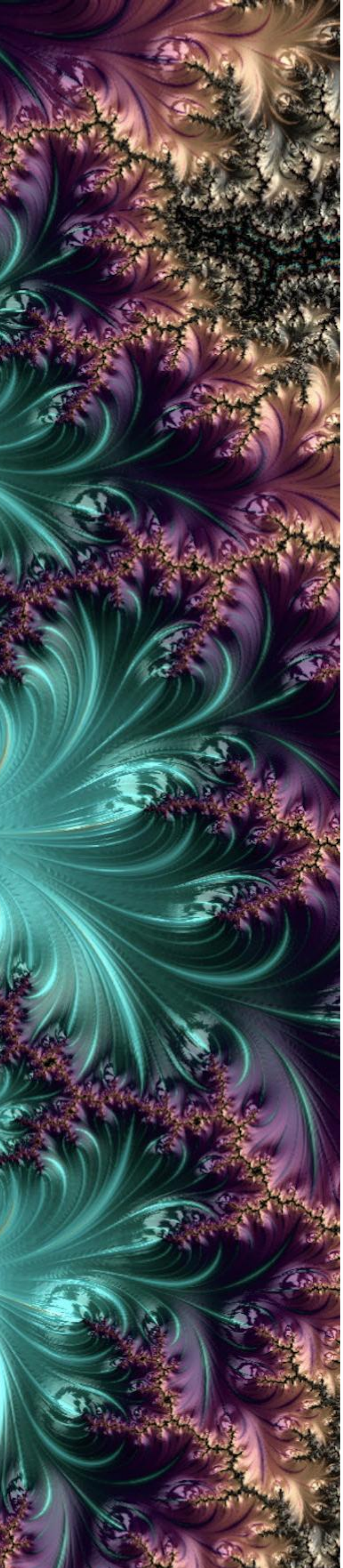


2024 Annual Report on the Dimensions of Data Quality

Year Nine:
Embracing the Future of Data Quality

November, 2024
9 Years of Analysis Sponsored by

DQMatters

Authored by Dan Myers

Executive Summary

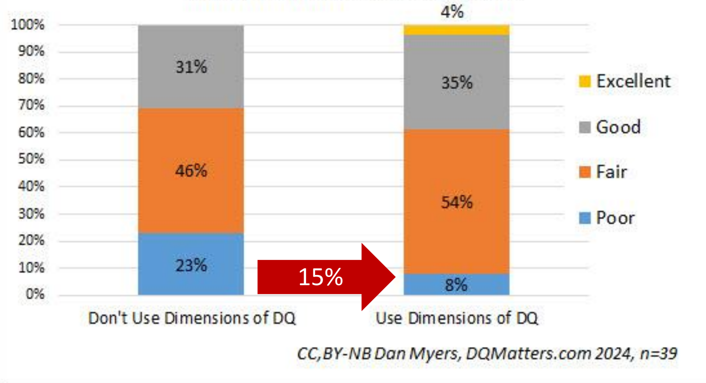


Figure 8c. Overall data quality for organizations that use the Dimensions or don't use them

- Organizations that measure Data Quality (DQ) by using the dimensions of data quality have identified and then reduced **poor** data quality by 15%. Organizations not using these frameworks underestimate the severity of their data quality problems. (red arrow on right) … more here
- 50% of orgs who don't use dimensions of DQ said they couldn't agree on a standard, so they gave up altogether … more here
- 66% of the respondents said their organizations use the dimensions of DQ, but only 21% use them continuously … more here
- Although 67% of organizations already use some data quality dimensions, a significant number (54%) express a strong interest in a standardized set. This suggests a willingness to refine existing approaches or adopt a standard to improve data quality management … more here

SIGNIFICANT RISE IN MEASUREMENT OF THESE DIMENSIONS

- Completeness has seen growth, especially in measuring the completeness of records and the truncation of values.
- Lineage measurement is also on the rise, driven by improved tools that can automatically document data lineage.
- Representation is increasing, particularly in documenting measurement units, presentation language, metadata availability, and the use of appropriate media.

KEY SHIFT IN TRENDS

- From basic data availability (Completeness) towards a deeper understanding of data context and meaning (Lineage, Representation, Metadata).
- From manual documentation to leveraging automated tools for greater efficiency and comprehensiveness.

Underutilization of Certain Dimensions. Despite their importance, some dimensions are not receiving enough attention. Timeliness and Accessibility have seen a decline in measurement. This is unexpected considering the growing need for fast and easy access to data. The report speculates that these drops might be temporary anomalies.

Discuss the report findings in the CDDQ LinkedIn Group

What's your opinion? How did you convince your organization to start using the dimensions of data quality? Let's discuss it here in the LinkedIn Conformed Dimensions group!

Table of Contents

Preface

If you've read this report in the past or subscribed to the Conformed Dimensions of Data Quality newsletter/blog you'll be familiar with the definitions of the dimensions of data quality and specifically the Conformed Dimensions. If that's the case, you can skip to the introduction below. If not, we recommend you read through this overview of the dimensions of data quality and the proposed standard for the dimensions, called the Conformed Dimensions of Data Quality (CDDQ).

Dimensions of data quality are categories used to characterize data and its fitness for use. Many famous data quality authors have written about the dimensions (offering different perspectives and emphasis). A few include: MIT (Lee, Pipino, Funk, Wang), Larry English, Tom Redman, David Loshin and the DAMA organization. We list many of them on our website's research section.

Dimensions Provide Value:

- Act as quick reference, checklist, and guide to quality standards
- Enable people to communicate current and desired state of data to speed up implementation
- Match dimensions against a business need and prioritize which assessments to complete first
- Understand what you will (and will not) get from assessing each dimension.[1]

The author of the Conformed Dimensions compared these author's dimensions, and many more sets of dimensions, and proposed a standardized set in 2013. Then a survey about the use of the dimensions was created in 2015 which was the first time this survey/report was published. More on the history here.

The Conformed Dimensions are hierarchical, meaning that each dimension has "Underlying Concepts" which create sub-groups for definition and measurement. This enables complexity when needed but simplifies high-level reporting and communication. Underlying Concepts are the basis for individual DQ metrics which quantifies a specific aspect for numeric calculation.

The Conformed Dimensions are unique because they emphasize the objective measurability of the dimensions. Each level of measurability is additive. Meaning that dimensions that are inherently measurable are also testable.

Desirability	Measurability	Definition
1. Most desirable	Inherent	This metric can be measured through statistical or logical observation of the data directly.
2. Somewhat desirable	Test-able	This metric can be measured through observation against some objective standard such as reality, another system or business requirement.
3. Least desirable	Subjective	This metric requires human input that is difficult-to-impossible to automate the testing process with computers. Answers typically cannot be predetermed via a mathematical or logical formula.

- Examples of inherent measures are those in the Completeness Dimension including: record population, attribute population, truncation and existence.
- Examples of testable measures are those that require input, like the Validity Dimension including: values in a specified range, conform to business rule, domain of predefined values, data type, or format.

Numerous data quality thought leaders endorse the Conformed Dimensions. See the endorsement page for more on their thoughts and case studies of clients who've taken the DQMatters DQ Jumpstart class that teaches how to use the Conformed Dimensions. The Conformed Dimensions have also been translated into three languages (German, Mandarin Chinese, and Portuguese) as well.

[1] McGilvray, 2008, p. 30-31

Introduction

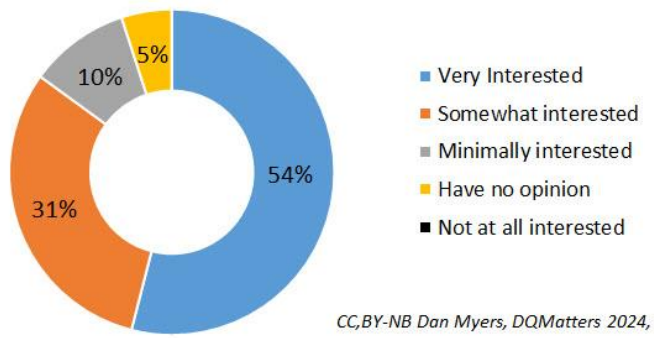
Each time we conduct the [Annual Dimensions of Data Quality Survey](), we learn so much about the current state of DQ and how the industry is changing. Despite a small sample size, the long-term perspective is clear. Organizations are slowly improving their use of the Dimensions of Data Quality, but as we identify in this report they struggle to:

a) Get started (pick dimensions, convince management which dimensions to start with)
b) Use more than the top 5 (Completeness, Consistency, Accuracy, Validity, Integrity)
c) Use more than one metric per dimension[2]
d) Use them continuously (on all data defects, all projects…etc.)

Struggles Facing Organizations

As outlined above, many organizations struggle to get started with the Dimensions of Data Quality because there is no standard. There are too many voices out there who espouse different lists without support. This makes it hard to convince management which to use. Since the first year that we conducted this survey, we've been interested to see if people would accept a proposed standard for the Dimensions of Data Quality, and the results have consistently painted a picture of strong interest.

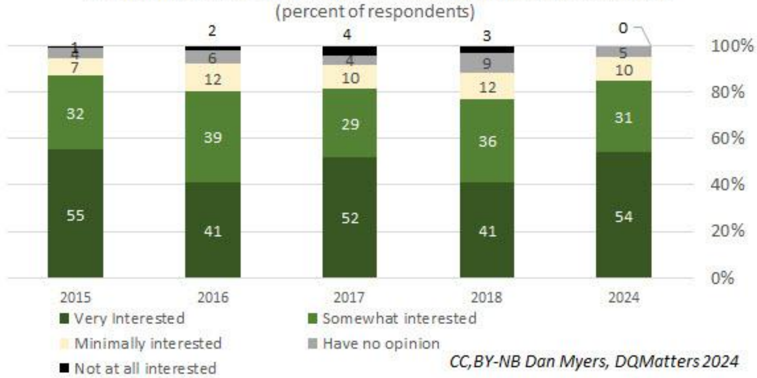
Figure 1a. If an industry standard set of dimensions of data quality was available, how interested would you be in using that at your organization?



This year was the same. In fact, 54% of respondents said they are "Very interested" (above). Although, 67% of organizations already have a set of dimensions of data quality[3], they are still interested in a standard. This implies that they are open to either amending their existing set with additional Dimensions/ Underlying Concepts or even converting to a standard.

As seen on the right, between 40 to 50 percent of the organizations surveyed are Very Interested in a standard. An additional 30-35 percent are Somewhat Interested.

Figure 1d. If an industry standard set of dimensions of data quality was available, how interested would you be in using that at your organization? (percent of respondents)



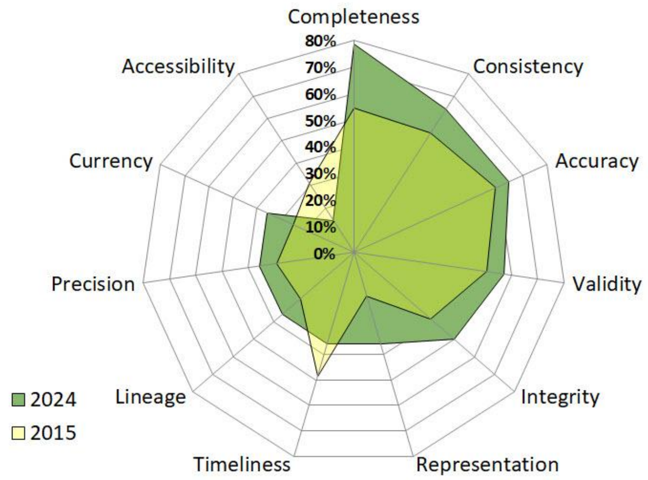
---

[2] For more on this refer to the explanation of Metrics on the levels within the Conformed Dimensions of Data Quality website: http://dqmatters.com/list-metric-examples

[3] See Figure 3d in the *Frequency of Usage of Dimensions of Data Quality* section of this report.

When choosing a set of dimensions, the fear of overwhelming their stakeholders caused leaders to trim their set or choose overly simplistic sets. The result is that organizations don't reap all the benefits desired. As you can see in the following spider diagram, (Figure 4e) the usage of nearly all the dimensions has increased in the last 9 years but it's been inconsistent.

Figure 4e. Improved Usage of the Dimensions of Data Quality from 2015 to 2024



CC,BY-NB Dan Myers, DQMatters 2024

1. Expansion in measurement of Completeness, Lineage and Representation since 2015
2. Reduction in measurement of Accessibility and Timeliness

Although we see a dip in the use of Timeliness and Accessibility, we expect these are short-term anomalies given the increasing need for access to data in a fast-paced environment.

One problem with this illustration is that it abstracts the change in usage of the Underlying Concepts within the dimensions. Although this survey has been conducted since 2015, we didn't collect the data at the Underlying Concept level of detail until 2018. The following section outlines changes at that level of detail.

What is an Underlying Concept?

Example for the Completeness Dimension

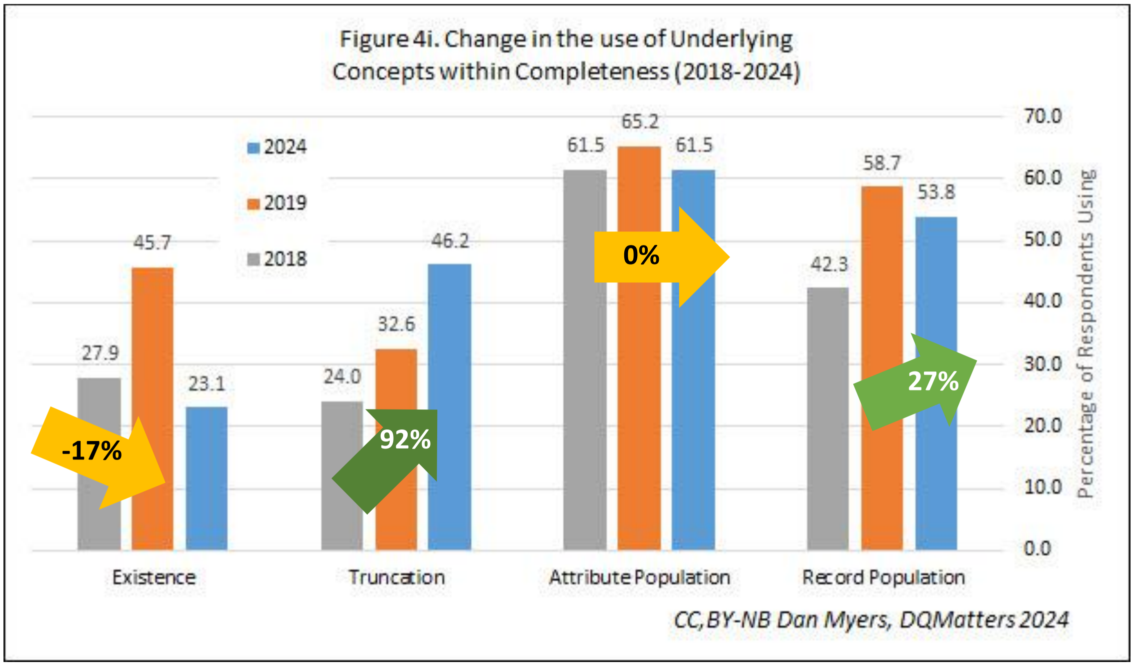
Conformed Dimension: Completeness- the degree of population of data values in a data set.

Underlying Concept:

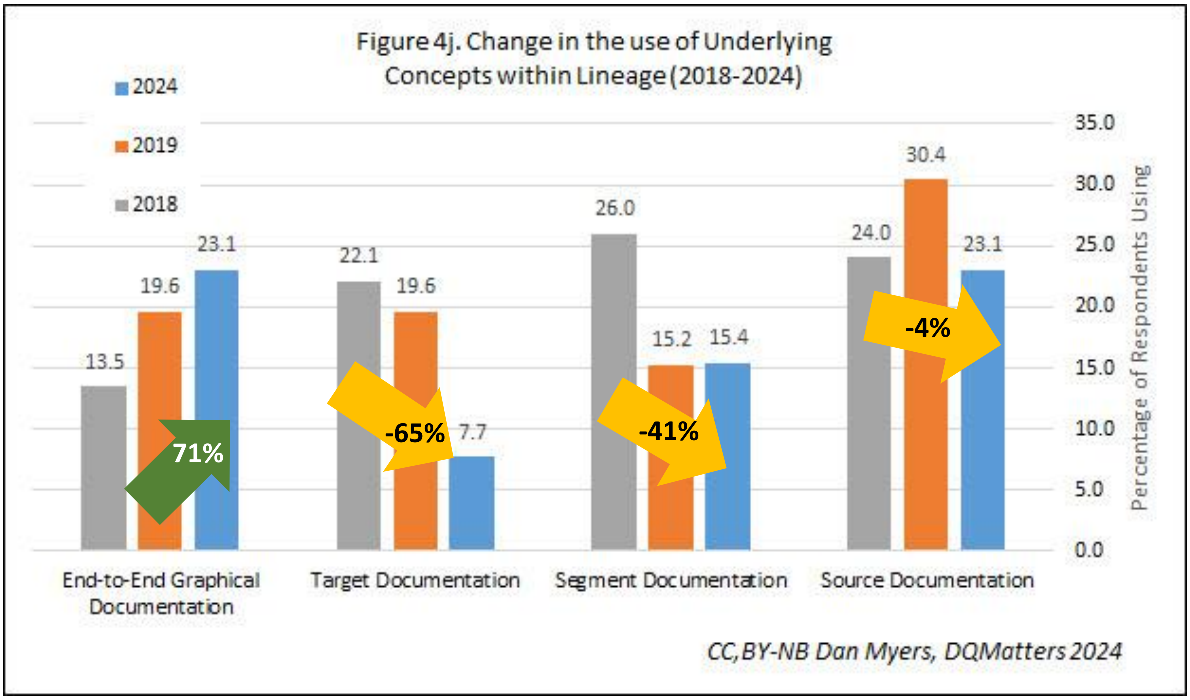
- Record Population- whether a row is present in a data set (table).
- Attribute Population- whether a value is present (not null) for an attribute (column).
- Truncation- whether the value contains all characters of the correct value.
- Existence- identifies whether a real-life fact has been captured as data.

Change in the Use of Specific Underlying Concepts of Data Quality

<u>Increase</u> in Use of Completeness:
We observed growth in the measurement of Completeness, specifically Record Population, which measures whether all rows were moved from source, which grew by 27% since 2018. Truncation showed an even more significant increase of 92%. The reduction in measurement of Existence (-17%), whether data has been captured or not, is puzzling but perhaps organizations are realizing that they don't need as much data. More likely than not, we just don't have enough data to draw the right conclusion.



<u>Mixed Change</u> in Use of Lineage:
As we'd expect, tools that automate much of the documentation of data lineage by parsing ETL code have improved over the last ten years and we see an increase in the level of measurement overall. Most notably was the increase in end-to-end graphical documentation (71%) but unfortunately we saw decrease in the other Underlying Concepts: Target Documentation (-65%) and Segment Documentation (-41%) and Source Documentation (-4%).[4]



The concerns we have is that Source Documentation is valuable when conducting root cause analysis to understand what the original up-stream value was compared to the downstream anomaly. This is typically static metadata about the data in an often times legacy system, which may not have been well documented and there aren't any knowledgeable team members around to populate this even if required.

Unfortunately, even data management functions can become siloed. If DQ and Metadata Management are split between people or IT versus business teams the organization can loose sight of the intensely important relationship. Metadata about the meaning of data, how it's transformed and who stewards it is imperative to data quality research, diagnosis and improvement. The use of a tool to automate End-to-End Graphical Documentation doesn't similarly populate Target, Segment and Source Documentation.

[4] See *Appendix 4- Change in Use of the Dimensions of Data Quality from 2018 to 2014 by Underlying Concept* for the details for every Underlying Concept.

<u>Increase</u> in Use of Representation:

Before we address Representation, it's best we clarify how we define it in the Conformed Dimensions and the survey.

Representation Dimension	Representation measures ease of understanding data, consistency of presentation, appropriate media choice, and availability of documentation (metadata).
Underlying Concept	**Definition of Underlying Concepts**
Includes Measurement Units	Well represented data includes the scale of measurement, such as weight, height, distance…etc.
Presentation Language	Data that is represented well is simple but elegantly formed with good grammar and presented in a standard way.
Metadata Availability	Comprehensive descriptions and other information about the characteristics of the data are provided in plain language.
Media Appropriate	The appropriate media (e.g. Web-based, hardcopy, or audio…etc.) are provided.
Easy to Read and Interpret	Illustrations and charts should be self-explanatory and presented with appropriate labels, providing context.

We saw a very significant growth in documentation about data's Measurement Units (186%). There was also growth in the use of Metadata Availability (33%), Appropriate Media (45%), and Presentation Language (26%). Strangely enough, there was a decline in measurement of the most basic and fundamental concept, Readability (-58%). During follow-up interviews, we found that for at least one respondent they chose not to include the Representation dimension within their enterprise scope to simplify DQ socialization across the company. Based on our consulting experience we found this to consistently tradeoff that middle-level managers feel they have to make.

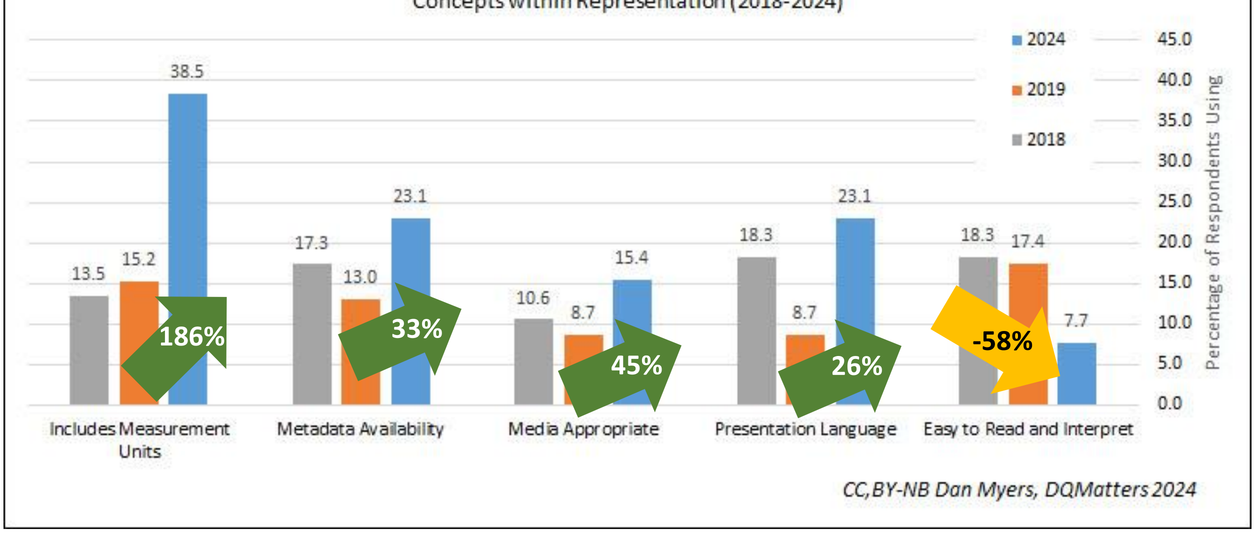


Figure 4k. Change in the use of Underlying Concepts within Representation (2018-2024)

Our recommendation to clients is to us the full set of dimensions of data quality internally, e.g. within the data management team, but focus on specific dimensions during projects and communication to audiences that may not be completely invested yet. The problem is that stakeholders will come back and ask why problems are still occurring and DQ teams will have to explain that issues are outside the original scope and defects may not be tagged with dimension specific information so the organization has no clue how severe the issues are without appropriate labels.

There's also growth in using standardized Presentation Language (26% increase) and choosing Appropriate Media for data representation (45% increase). This signifies an effort to improve both structured and unstructured data clarity and relevance for wider audiences.

Despite these positive trends, there's a concerning 58% decline in measurement of readability, which relates to the ease of understanding data visualizations and presentations. We surmise that this drop can be attributed to the exclusion of the Representation Dimension from some companies' enterprise-wide data quality initiatives but given more and more focus on AI models that struggle to correctly interpret information as naturally as humans, this metadata will be important.

<u>Decrease</u> in Use of Accessibility:

For the first time, since we started tracking the use of Underlying Concepts, in 2018, the use of Accessibility has dropped (right). Because Retention is typically managed (and documented very thoroughly) via the Records Management profession, we're skeptical whether it truly isn't being measured. It seems more likely that it's being tracked by a whole different department and not considered part of data quality. Similarly, Access Control is tracked within Information Security, so it is likely outside most DQ professionals' purview.

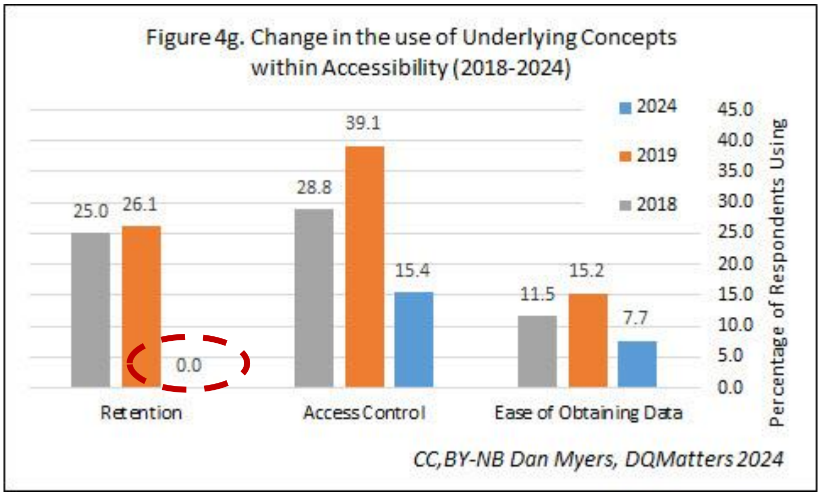


Figure 4g. Change in the use of Underlying Concepts within Accessibility (2018-2024)

In the past, with the introduction of Big Data and the associated hype cycle around Hadoop and data lakes we saw an increased focus on Accessibility, but since 2019 that has fallen across all three areas.

<u>Mixed Change</u> in Use of Timeliness:

As seen earlier, in the spider diagram [Figure 4e](), we've seen the use of Timeliness generally decline over time. Despite the general decline overall, there has been a significant increase in the measurement of [Manual Float]() from 2018 to 2024 grew 256% (8.7% to 30.8% as shown by the red arrow over Figure 4h). As a reminder, Manual Float is "the time from when an observation is made to the point it is recorded in electronic format." This isn't as common in the new digital economy, where we have so many devices, but human-entered data, like in-person audits, some healthcare, and some sales and marketing activities still use manual entry.

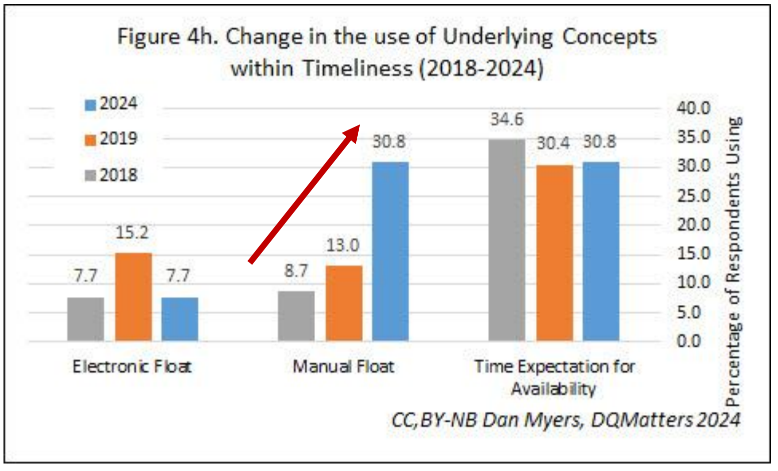
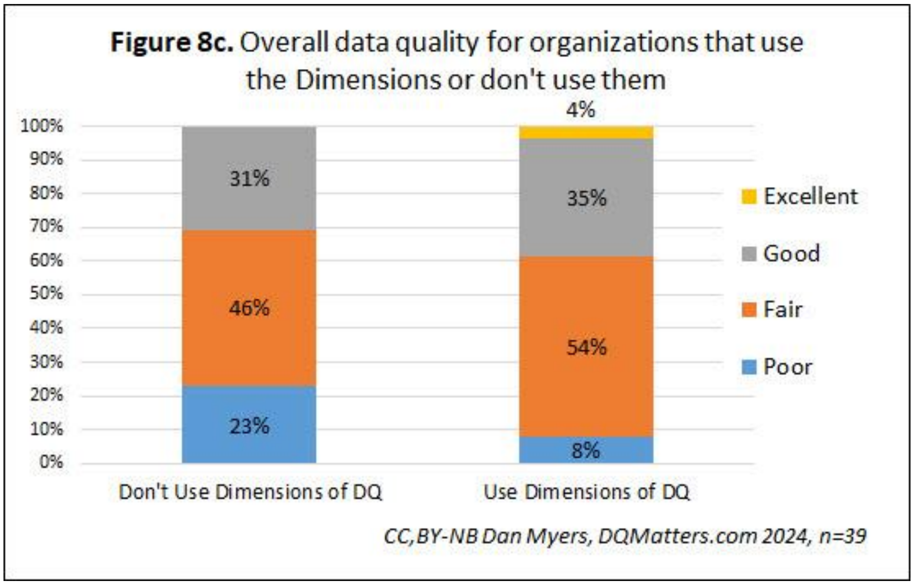


Figure 4h. Change in the use of Underlying Concepts within Timeliness (2018-2024)

Changes in Respondent Reported Organizational Data Quality Levels

Next, we'll talk about not just the decreases in usage of specific Underlying Concepts, but how overall data quality relates to the use of the dimensions of data quality. We should point out two aspects to consider.

What do we Get by Measuring Data Quality?



Logically, measurement should improve decision making and we can see that only organizations that measure data quality (e.g. using the dimensions of data quality) understand their real risk and opportunities for improvement.

Survey respondents were asked about the overall data quality within their organizations. Those that do not use the dimensions of data quality reported 23% of them have "Poor" DQ as opposed to only 8% that do use the dimensions (right). In other words, organizations that measure DQ can identify and then reduce poor data quality by 15%.

This suggests that organizations that don't measure data quality using a structured framework like the dimensions of data quality may significantly underestimate how bad their data quality actually is. This emphasizes that using the dimensions of data quality to quantify data quality issues can reveal a much more insightful picture of an organization's data quality.

Similarly, those self-reporting "Fair" data quality (46%) but don't use the dimensions had 8% lower data quality levels.[5] It should also be noted that only one organization (who does use the dimensions) reported "Excellent" data quality. In other words, we all have a general sense of how poor our data quality is, but unless we quantify it, using the dimensions of data quality, we don't really know how bad it is. This data says it's about three times as bad.[6]

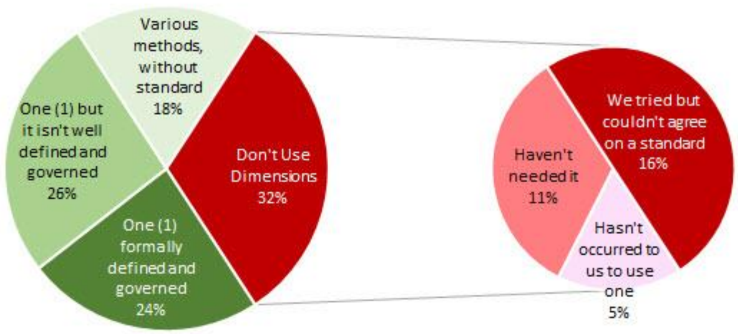
[5] For more detail see Appendix 8, Table 8a.

[6] See Figure 8a and Table 8a in Appendix 8

Why Organizations Aren't Using the Dimensions

Within the third of respondents who don't measure data quality with the dimensions there are various reasons. Half of these organizations (16% of total) couldn't agree on a standard and gave up. It didn't occur to 5% that they need to use the dimensions.

Figure 3ef. Levels of Use of the Dimensions of Data Quality



68% of organizations surveyed use some sort of dimensions of data quality (green shades)

32% of organizations don't use dimensions (red shades)

CC,BY-NB Dan Myers, DQMatters 2024, n=39

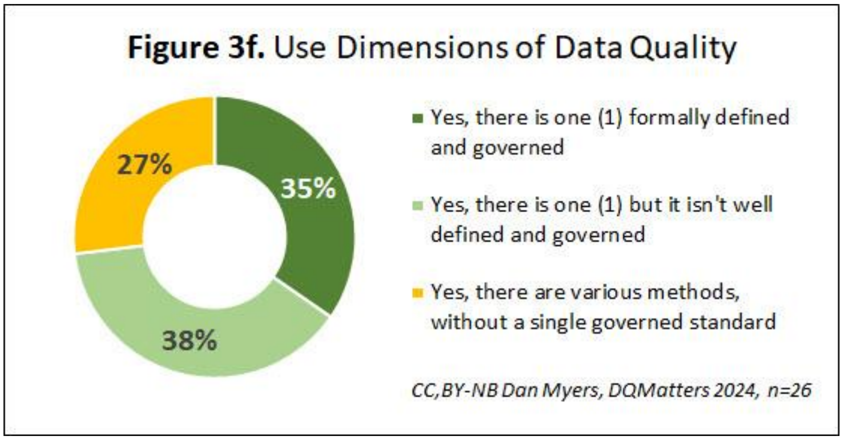
Eleven percent (11%) don't think they need to use the dimensions. One of the respondents falling into the latter category, had consistently poor data quality reported across departments, but is struggling to make the business case with management. The other respondents in this class have self-reported "fair" or "good" data quality so apparently that's why measurement isn't top of mind.

One of the primary reasons for creating the Conformed Dimensions of Data Quality (CDDQ), was to ensure organizations don't give-up and get started with something. One respondent of the 2017 survey, said that "the Conformed Dimensions has prevented fist fights over what is the 'correct' definition of each dimension during implementation" let alone saved time.

Frequency of Usage of Dimensions of Data Quality

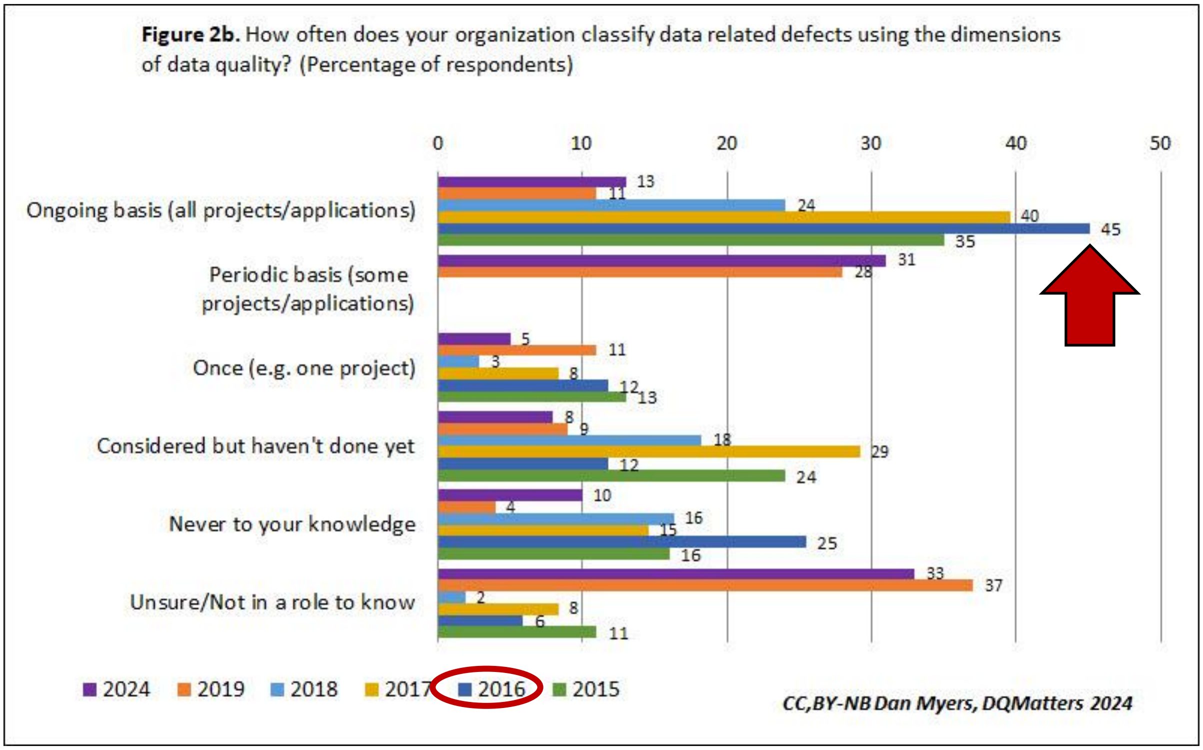
About 68% of respondents do use the dimensions of data quality, which we're glad to hear. Now, taken in context we know that this number is likely overstated because organizations that do not have a data quality program or anyone aware of this topic did not take our survey. Our respondent pool has bias because they were aware of the need for data quality enough to take our survey.

Those that do use the dimensions breakdown into three groups: those with a formally defined and governed set, those without a well-defined set, and those with various methods in use (see breakdown below in Figure 3f). Each group has nearly the same portion.



Although 68% (Figure 3ef, above) of the respondents said their organizations use the Dimensions of DQ, a much smaller number of those **continuously** use them.

Only 21% of all respondents use the dimensions in an ongoing way[7] which shows it's one thing to define a list of dimensions, but true value generation requires a cultural change focused on continuous improvement through adoption. Our data shows that organizations with mature data management programs do use the dimensions more frequently. We draw this conclusion because in 2016 we conducted this survey at the Enterprise Data World (EDW) conference- collaboratively conducted by [Dataversity](#) and the [Data Administration Management Association](#) (DAMA). That year, we saw 2-3 times as many organizations that use the dimensions on an ongoing basis (see Figure 2b below). As many as 45% of the respondents used them continuously according to the 2016 data.



7 See Table 3a in [Appendix 3](#)

The Most Popular Dimensions of Data Quality

As if it's a horse race, it's always fun for the authors of this report to see the results about which dimensions are most used by organizations. Of course, various Conformed Dimensions of Data Quality blog posts have been written, pointing out the importance of fundamental dimensions like Completeness or Validity. But the competition for 3-5th is most interesting. See Table 4e, on the right, with the change since our last survey.

Table 4e. Ranking Changes

2019		2024	
1	Completeness	1	Completeness
2	Validity	2	Consistency
3	Consistency	3	Accuracy
4	Accuracy	4	Validity
5	Integrity	5	Integrity
6	Accessibility	6	Representation
7	Representation	7	Timeliness
8	Timeliness	8	Lineage
9	Lineage	9	Precision
10	Precision	10	Currency
11	Currency	11	Accessibility

To better appreciate the scale of difference the following bar chart (Figure 4a below[8]) has the breakdown for the dimensions.

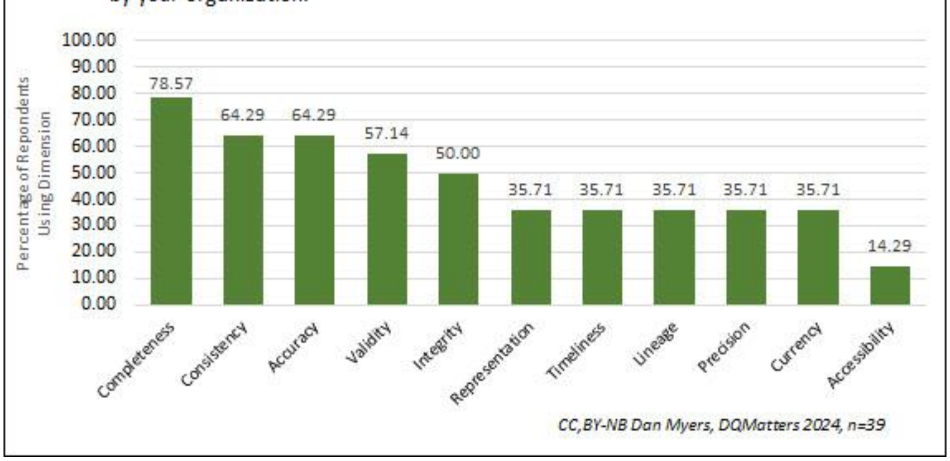


Figure 4a (percentage). Select any of the following dimensions currently used by your organization.

Many organizations have some sort of basic data quality checks in place. For instance, they report these levels of use of the top three most widely used dimensions (e.g. Completeness [78%], or Consistency [64%], Accuracy [64%]). They desperately need to push into more nascent areas like time related dimensions (Timeliness and Currency) which improve data self-service and democratization efforts when executed in conjunction with Accessibility.

[8] Due to the small survey size (39) this year the percentages for the lower end aren't as helpful (e.g. 35.71 for five of the least used dimensions), but please use Figure 4f (in Appendix 4) to identify which of the Underlying Concepts were used within these dimensions to gain additional insight and identify candidates to add to your organizations set of dimensions.

Data Quality Levels by Organizational Subject Area

In 2019, we started to ask organizations about data quality levels by function. This year's results reveal significant deficiencies in certain departments affecting both revenue generation and cost mitigation.

Figure 8.1a Data Quality Levels by Function (percentage)
What level of data quality does your organization have for the following subject areas? (2024)

From Figure 8.1a, above, the highest levels are in:

- Products (Services) and (Physical Goods)
- Technology/IT Infrastructure

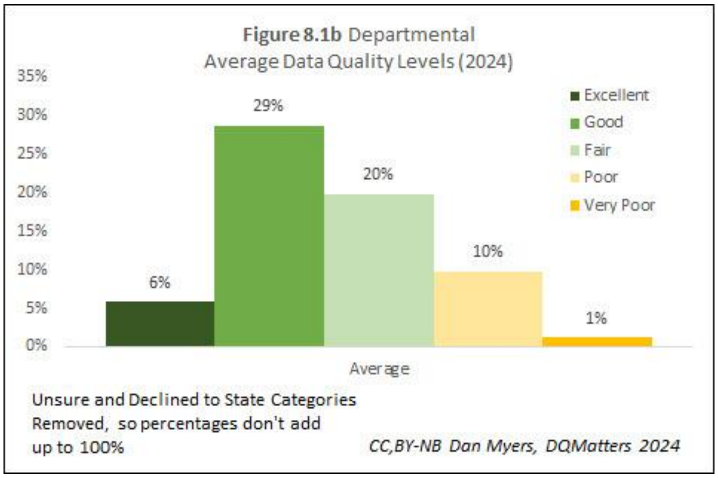
The worst data quality was in:

- Marketing/Sales
- Operations
- Human Relations

The most impactful are the loss of revenue in marketing and sales roles. The Sales Intelligence platform, Cognism (D&B competitor) says, "sales reps spend 20% of their time researching prospects, and bad data causes them to waste another 30%."[9] The business case for Data Quality is much easier to make with direct revenue analogies than cost savings or risk mitigation due to the dollar-for-dollar opportunity costs. Businesses starting their Data Quality journey should consider starting in these related departments: Marketing/Sales, Services, Products, and sometimes Operations that impacts sales or business partners.

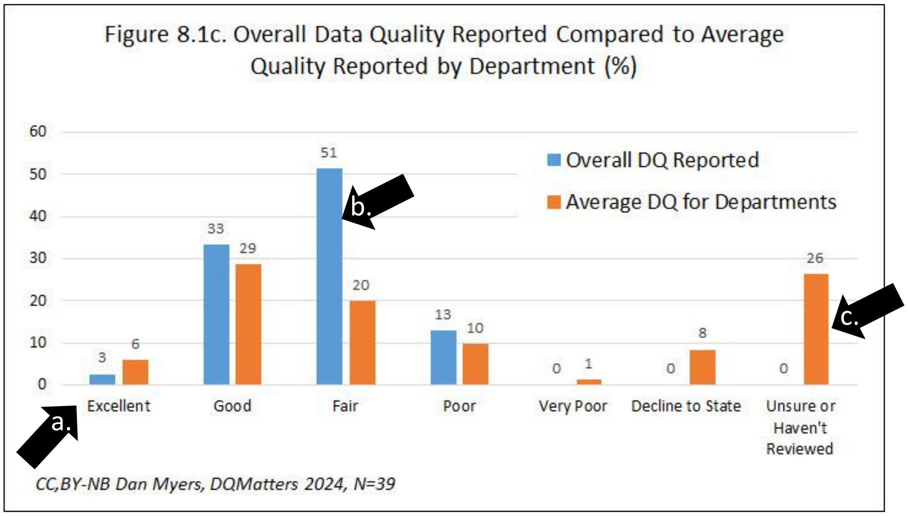
[9] https://www.cognism.com/blog/dun-bradstreet-competitors, November 25, 2024.

In Figure 8.1b we've calculated the average for each quality level by function. It looks like a standard bell curve, doesn't it. Most organizations have "Good" to "Fair" data quality for each function. Interestingly, four functions have high levels that are unsure or haven't reviewed levels of quality: Service, Ops, IT, and HR (see dashed red line on Figure 8.1a above). If you have thoughts on these areas post a note on the [CDDQ Group on LinkedIn](#) regarding your experience relating to why these are less frequently measured.



Among all responses, only one reported "Excellent" organization-wide data quality. But when we saw that multiple organizations were reporting Excellent data quality for individual departments, like Services, Physical goods, and Technology/IT Infrastructure, we started to wonder how these departmental averages compare to the overall reported data quality. Below, in Figure 8.1c, you can see that:

a. People believe that individual departments have double the average level of "Excellent" DQ compared to overall DQ
b. Nearly a third (26%) of respondents don't know about individual departmental DQ levels
c. Half of the people reporting "Fair" DQ at the overall organization level don't actually know what the departmental DQ levels are.



This falls somewhat in line with the fact that about a third of the respondents took the survey in the context of a single department (don't manage or have visibility to the whole organization's DQ) as seen in Figure 9a in [Appendix 9](#). The C-level of leadership needs to emphasize stronger cross-functional measurement of data quality. This will especially become relevant as AI initiatives seek to discover revenue generating (or cost cutting) recommendations that use data from all departments.

One interviewee said that their data quality department was facing steep budget cuts until their C-Suite announced new AI initiatives. Then the cuts were walked back saving jobs and emphasizing more focus on enterprise-wide data lineage and consistent data quality measures.

DQ Tool Usage

Of course, the data quality discussion wouldn't be complete without a discussion about tools. Tools can significantly improve team efficiency and reduce labor intensive tasks. Here are a few quick observations we've made looking at the data:

- The DQ Tool market is highly fragmented (few vendors were identified by more than one respondent)
- More companies have built their own tools than we expected (23% of respondents)

As seen in Figure 20.1, below, almost half of the respondents don't have a tool of any kind (43.6%) and the mix of those that do (56%) includes those that purchased it (33.3%) and those that built their own (23.1%). A very small portion (only one respondent out of the 39 organizations surveyed) both built and bought their tools. Over time, organizations tend to cycle through their tools so it was important that respondents answered only for currently used tools. Often both in-house and purchased tools gain popularity and then become *shelfware* (aren't used) over time.

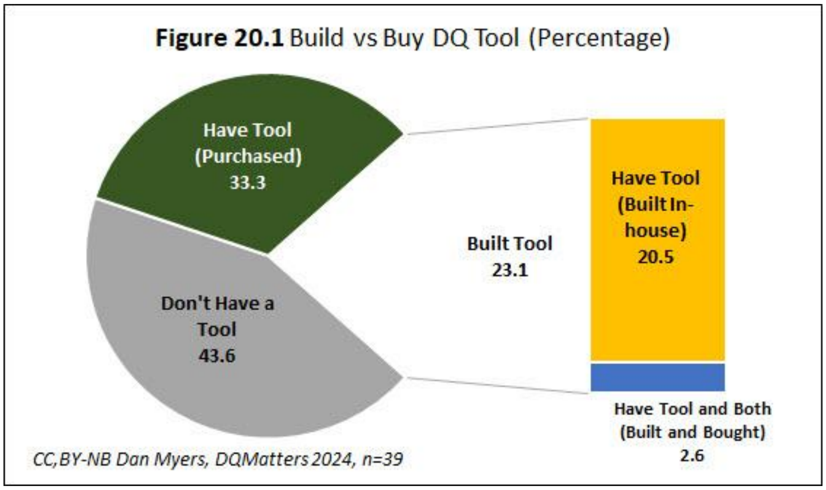


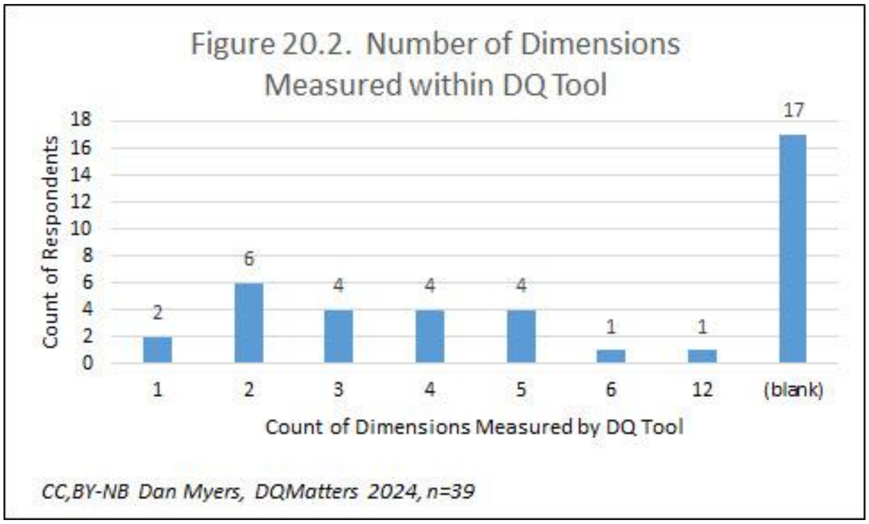
Figure 20.1 Build vs Buy DQ Tool (Percentage)

Vendor Names Listed by Respondents:

- Informatica
- SAP
- Collibra
- Ataccama
- DQ Labs
- IBM Infosphere
- Experian
- SAS Data Flux
- WhereScape Red
- SonarQube (code quality)

We also found that organizations who build their own tools generally measured more dimensions than those that purchased their tools. The average number of dimensions measured by those who have a purchased tool was only 2.7, whereas those that were built in-house were 5.1. In other words, those who built their own measure twice as many dimensions. Unfortunately, not enough of these respondents took the second part of our survey that asks about which Underlying Concepts are measured so we can't provide detail about what Underlying Concepts they were measuring. Please consider registering for the Conformed Dimensions of Data Quality newsletter where we advertise about the survey when it's conducted.

When asked how many dimensions of data quality are measured using these tools, respondents' answers varied widely (see below).



A third of respondents (33%) said they measure Accuracy with their tools. This is confusing at best because Accuracy is an all-encompassing word used by many to mean very different things. Among data quality experts, "Accuracy" generally has two components (definitions below from CDDQ):

- **Agree with Real-world:** Degree that data factually represents its associated real-world object, event, or concept.
- **Match to Agreed Source:** Measure of agreement between data and the source of that data. This is used when the data represent intangible objects or transactions that can't be observed visually.

The former, agree with real-world, is not something you can measure out of the box with any tool, so we assume that when respondents say they measure Accuracy, they measure the agreement with a defined system of record, which is perhaps better defined as the Consistency dimension which the CDDQ defines as "whether or not data is equivalent across systems or location of storage" with four of its own Underlying Concepts (listed below).

- **Equivalence of Redundant or Distributed Data:** The measure of similarity with other sources of data that represent the same concept.
- **Format Consistency:** This measures the conformity of format of the same data in different places.
- **Logical Consistency:** Logical consistency measures whether two attributes of related data are conceptually in agreement, even though they may not record the same characteristic of a fact.
- **Temporal Consistency:** The measure of uniformity of the data compared to historical values.

Briefly looking at that same list, a decent number of these same respondents say they measure Accuracy but don't measure Consistency. This would imply that they are not measuring equivalence across systems, and they have created some internal definition of Accuracy that expands on the "Agree with Real-World" concept.

Conclusion

Embracing the Future of Data Quality

This survey reveals a data quality landscape in transition. Although organizations pay lip service to their adoption of dimensions of data quality, the journey towards data quality excellence is fraught with challenges. The lack of a universally accepted standard for data quality dimensions, coupled with the reticence of complexity, often hinders organizations from embracing a comprehensive approach. As a result, many organizations settle for basic data quality checks, focusing on the most popular dimensions and neglecting other crucial aspects like timeliness, accessibility, and representation. This limited approach prevents organizations from fully realizing the benefits of data quality management.

To navigate this evolving landscape successfully, organizations must overcome these hurdles and embrace a more holistic and structured approach to data quality. ***Adopting a standardized framework, such as the Conformed Dimensions of Data Quality, can streamline implementation, facilitate communication, and ensure a shared understanding across the organization.*** Moving beyond the most popular five dimensions and embracing a wider range of data quality aspects is essential for supporting emerging data needs, including data self-service and democratization efforts. ***Furthermore, organizations must transition from ad-hoc data quality checks to a culture of continuous data quality management, where data quality is embedded in daily operations and decision-making processes.***

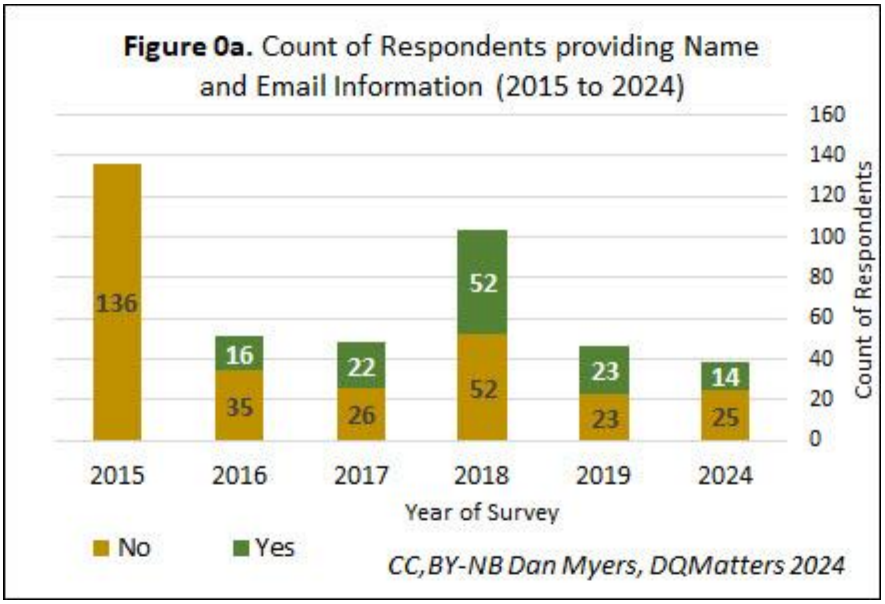
This cultural shift requires not only a commitment to consistent measurement and improvement but also a data-quality-driven mindset throughout the organization. By embracing a standardized framework, incorporating a wider range of dimensions, and fostering a culture of continuous data quality management, organizations can improve operational performance and ensure data fitness for use in an increasingly data-centric world.

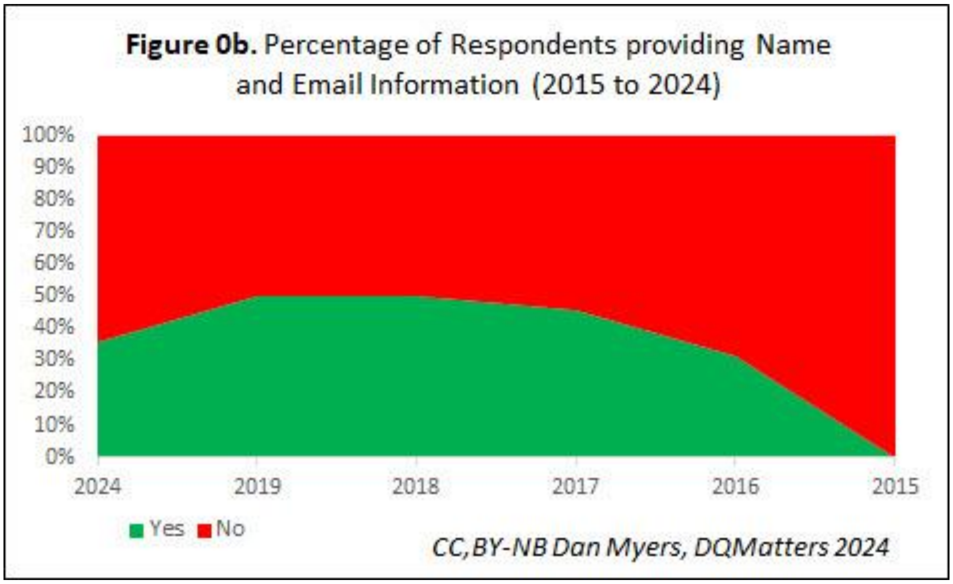
Discuss the report findings here in the CDDQ LinkedIn Group

Appendix

It may seem inconsistent that we don't include all charts in the appendix (e.g. 1a and 1d, where 1b and 1c are missing) but only charts that are cited in the body of the paper are included in this appendix. If you would like a copy of any other figure or table, please contact the authors.

Appendix 0- Respondent Participation





Appendix 1- Interest in a Standard for the Dimensions of Data Quality

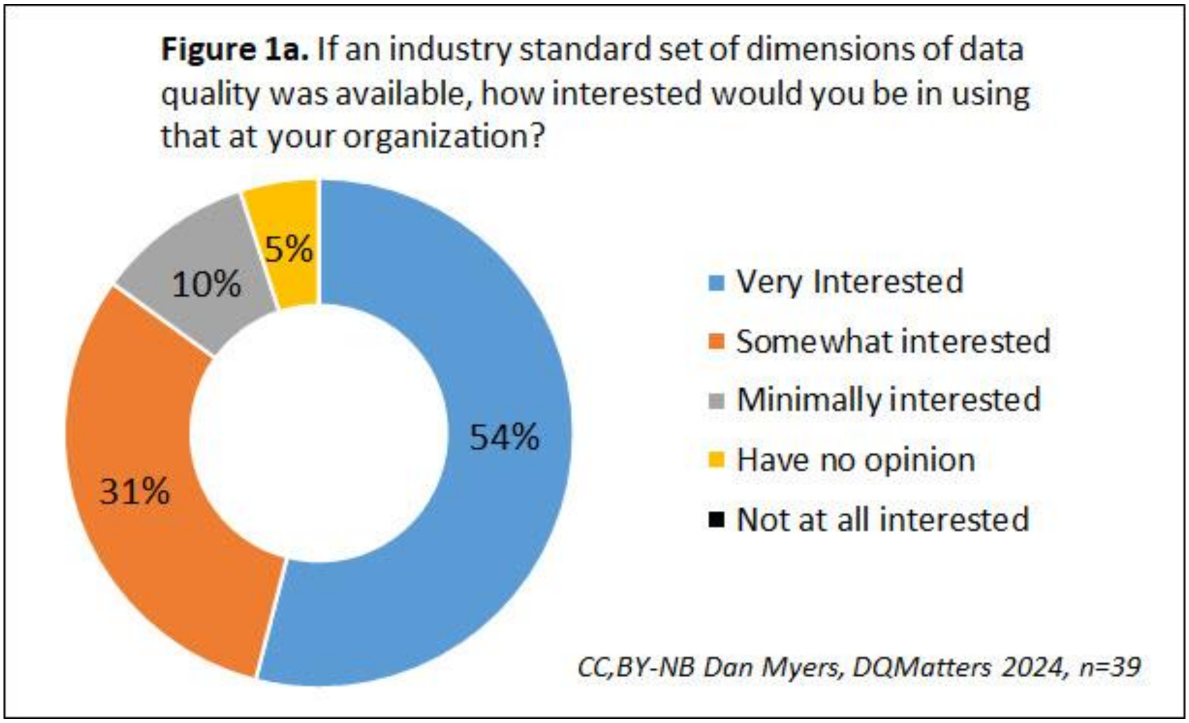


Figure 1a. If an industry standard set of dimensions of data quality was available, how interested would you be in using that at your organization?

Appendix 2- Frequency of Use of Dimensions of DQ

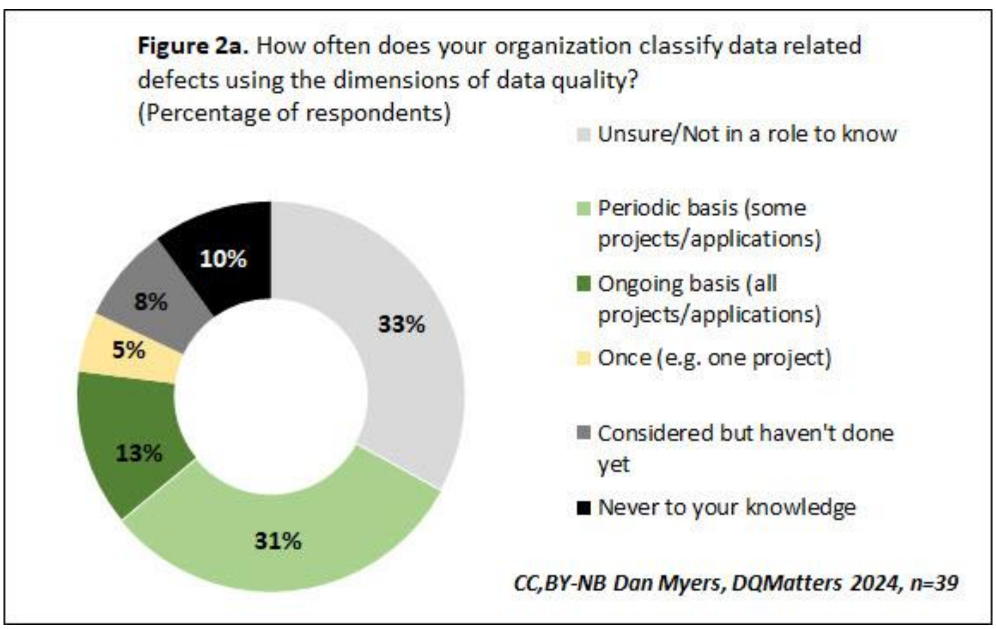
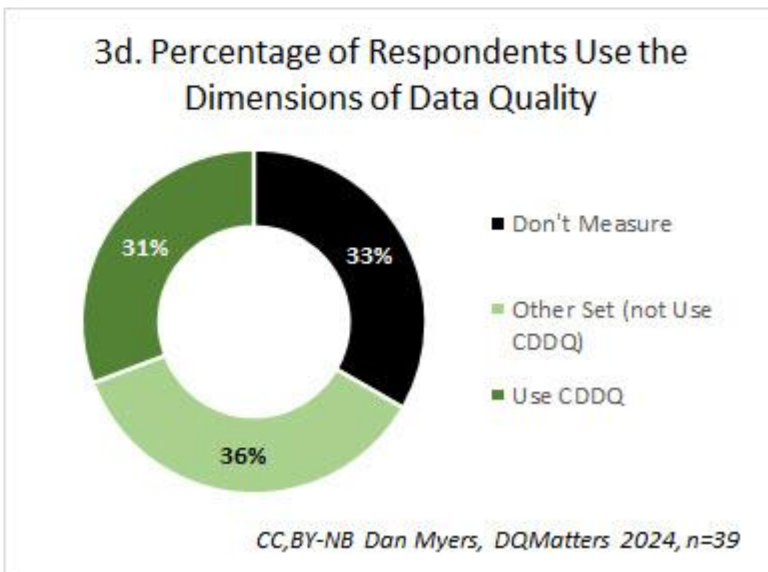


Figure 2a. How often does your organization classify data related defects using the dimensions of data quality? (Percentage of respondents)

Appendix 3- Have method and Using Ongoing

Table 3a. Have method and Using Ongoing

Have a method of categorizing DQ issues	Count of Measuring	% of Measuring	% of All Respondents
Yes, there are various methods, without a single governed standard	7	27%	
Never to your knowledge	2	8%	
Periodic basis (some projects/applications)	5	19%	
Yes, there is one (1) but it isn't well defined and governed	10	38%	
Considered but haven't done yet	2	8%	
Never to your knowledge	1	4%	
Once (one project/one time)	2	8%	
Ongoing basis (all projects/applications)	1	4%	3%
Periodic basis (some projects/applications)	4	15%	
Yes, there is one (1) formally defined and governed	9	35%	
Considered but haven't done yet	1	4%	
Never to your knowledge	1	4%	
Ongoing basis (all projects/applications)	4	15%	10%
Periodic basis (some projects/applications)	3	12%	8%
Grand Total	**26**	**100%**	
Total Survey Respondent Count 2024	39		21%



Appendix 4- Change in Use of the Dimensions of Data Quality

Table 4. Use of the Dimensions of Data Quality (2015-2024)

Average Change from 2015 to 2024: **6.73%**

2019-2024 Change		2024	2019	2018	2017	2016	2015	2015-2024 Change
-4.0%	Completeness	79%	83%	82%	77%	73%	54%	24.16%
-0.9%	Consistency	64%	65%	69%	73%	49%	54%	10.61%
3.4%	Accuracy	64%	61%	61%	81%	73%	59%	5.46%
-23.3%	Validity	57%	80%	77%	58%	51%	51%	6.41%
4.3%	Integrity	50%	46%	51%	46%	35%	38%	11.76%
-1.2%	Representation	36%	37%	34%	29%	22%	17%	18.80%
0.9%	Timeliness	36%	35%	45%	50%	49%	49%	-12.82%
5.3%	Lineage	36%	30%	38%	29%	24%	26%	9.24%
11.8%	Precision	36%	24%	29%	29%	27%	29%	6.30%
11.8%	Currency	36%	24%	17%	33%	33%	25%	10.71%
-29.2%	Accessibility	14%	43%	43%	35%	20%	31%	-16.60%

Figure 4a (count). Select any of the following dimensions currently used by your organization.

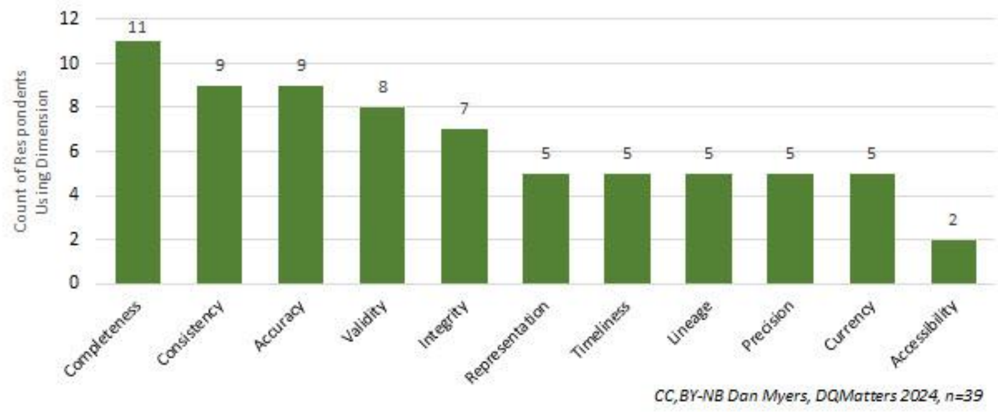


Table 4f. Ranking of Industry Use of the Underlying Concepts of DQ Dimensions (2024). Sorted by 2024

Sort A	Sort B	**Dimensions- Underlying Concept Used**	**2024**	**2019**	**2018**	**Percentage Growth**		**Change**
7	7.3	Accessibility- Retention	0	26	25	-100%		-25
5	5.3	Timeliness- Electronic Float	8	15	8	0%		0
7	7.1	Accessibility- Ease of Obtaining Data	8	15	12	-33%		-4
10	10.1	Representation- Easy to Read and Interpret	8	17	18	-58%		-11
9	9.3	Lineage- Target Documentation	8	20	22	-65%		-14
10	10.3	Representation- Media Appropriate	15	9	11	45%		5
9	9.2	Lineage- Segment Documentation	15	15	26	-41%		-11
7	7.2	Accessibility- Access Control	15	39	29	-47%		-13
9	9.4	Lineage- End-to-End Graphical Documentation	23	20	13	71%		10
8	8.2	Precision- Granularity	23	11	17	33%		6
10	10.4	Representation- Metadata Availability	23	13	17	33%		6
10	10.2	Representation- Presentation Language	23	9	18	26%		5
8	8.1	Precision- Precision of Data Value	23	22	20	14%		3
9	9.1	Lineage- Source Documentation	23	30	24	-4%		-1
2	2.4	Completeness- Existence	23	46	28	-17%		-5
5	5.2	Timeliness- Manual Float	31	13	9	256%		22
3	3.3	Consistency- Logical Consistency	31	28	23	33%		8
5	5.1	Timeliness- Time Expectation for Availability	31	30	35	-11%		-4
4	4.1	Validity- Values in Specified Range	31	54	43	-29%		-13
8	8.3	Precision- Domain Precision	38	4	n/a	n/a		#VALUE!
11	11.1	Currency- Current with World it Models	38	24	13	208%		26
10	10.5	Representation- Includes Measurement Units	38	15	13	186%		25
3	3.4	Consistency- Temporal Consistency	38	20	21	82%		17
6	6.3	Integrity- Cardinality	38	26	25	54%		13
4	4.5	Validity- Values Conform to Format	38	41	28	38%		11
6	6.1	Integrity- Referential Integrity	38	33	29	33%		10
1	1.1	Accuracy- Agree with Real-world	38	35	30	29%		9
4	4.4	Validity- Values Conform to Data Type	38	48	38	0%		0
4	4.3	Validity- Domain of Predefined Values	38	63	43	-11%		-5
2	2.3	Completeness- Truncation	46	33	24	92%		22
3	3.1	Consistency- Equivalence of Redundant or Distributed Data	46	39	32	45%		14
6	6.2	Integrity- Uniqueness	46	39	36	30%		11
1	1.2	Accuracy- Match to Agreed Source	46	41	37	26%		10
3	3.2	Consistency- Format Consistency	54	26	33	65%		21
2	2.1	Completeness- Record Population	54	59	42	27%		12
4	4.2	Validity- Values Conform to Business Rule	62	48	39	56%		22
2	2.2	Completeness- Attribute Population	62	65	62	0%		0

Figure 4f. Ranking of Industry Use of the Underlying Concepts of DQ Dimensions (2024). Sorted by 2024

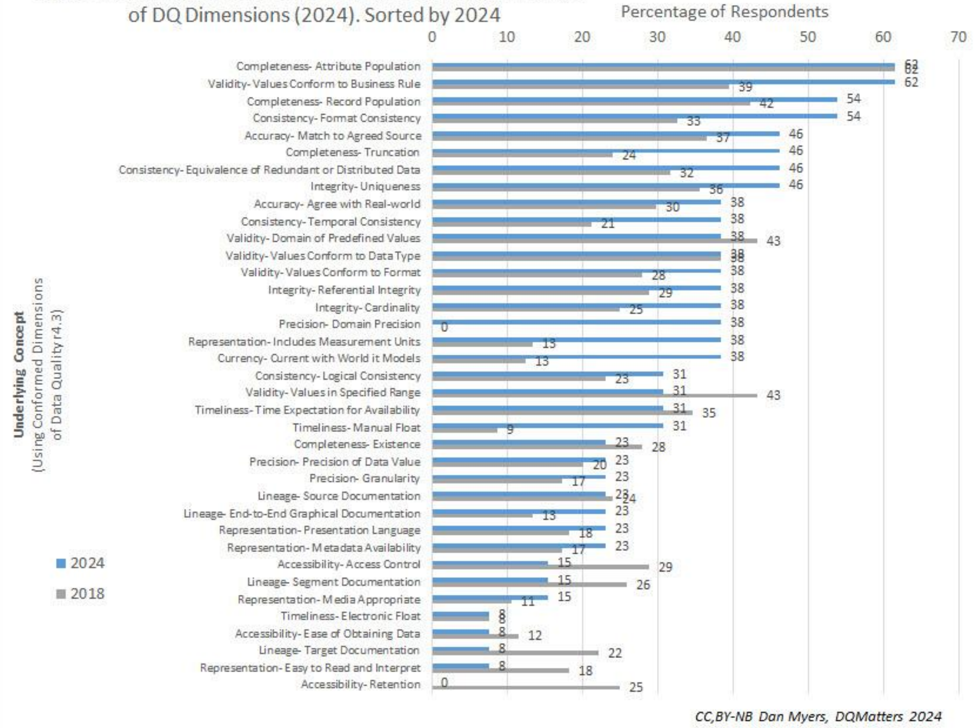
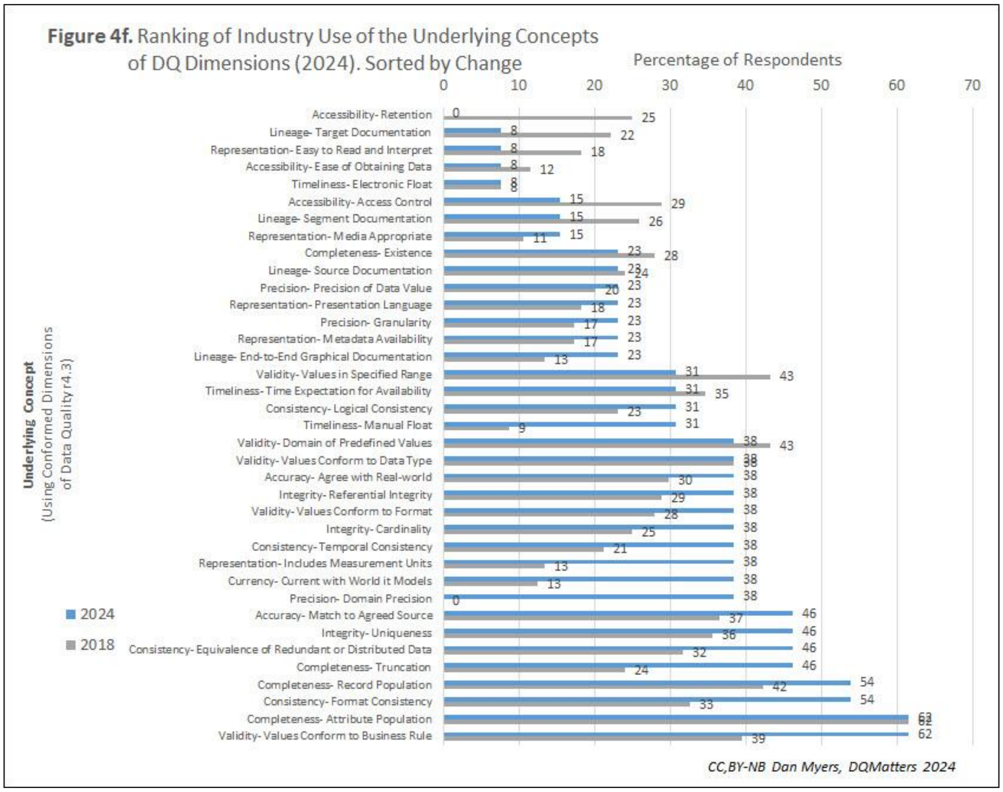


Table 4g. Ranking of Industry Use of the Underlying Concepts of DQ Dimensions (2018-2024). Sorted by Change

Sort A	Sort B	Dimensions- Underlying Concept Used	2024	2019	2018	Percentage Growth	Change
7	7.3	Accessibility- Retention	0	26	25	-100%	-25
9	9.3	Lineage- Target Documentation	8	20	22	-65%	-14
7	7.2	Accessibility- Access Control	15	39	29	-47%	-13
4	4.1	Validity- Values in Specified Range	31	54	43	-29%	-13
9	9.2	Lineage- Segment Documentation	15	15	26	-41%	-11
10	10.1	Representation- Easy to Read and Interpret	8	17	18	-58%	-11
2	2.4	Completeness- Existence	23	46	28	-17%	-5
4	4.3	Validity- Domain of Predefined Values	38	63	43	-11%	-5
7	7.1	Accessibility- Ease of Obtaining Data	8	15	12	-33%	-4
5	5.1	Timeliness- Time Expectation for Availability	31	30	35	-11%	-4
9	9.1	Lineage- Source Documentation	23	30	24	-4%	-1
5	5.3	Timeliness- Electronic Float	8	15	8	0%	0
4	4.4	Validity- Values Conform to Data Type	38	48	38	0%	0
2	2.2	Completeness- Attribute Population	62	65	62	0%	0
8	8.1	Precision- Precision of Data Value	23	22	20	14%	3
10	10.3	Representation- Media Appropriate	15	9	11	45%	5
10	10.2	Representation- Presentation Language	23	9	18	26%	5
8	8.2	Precision- Granularity	23	11	17	33%	6
10	10.4	Representation- Metadata Availability	23	13	17	33%	6
3	3.3	Consistency- Logical Consistency	31	28	23	33%	8
1	1.1	Accuracy- Agree with Real-world	38	35	30	29%	9
9	9.4	Lineage- End-to-End Graphical Documentation	23	20	13	71%	10
1	1.2	Accuracy- Match to Agreed Source	46	41	37	26%	10
6	6.1	Integrity- Referential Integrity	38	33	29	33%	10
6	6.2	Integrity- Uniqueness	46	39	36	30%	11
4	4.5	Validity- Values Conform to Format	38	41	28	38%	11
2	2.1	Completeness- Record Population	54	59	42	27%	12
6	6.3	Integrity- Cardinality	38	26	25	54%	13
3	3.1	Consistency- Equivalence of Redundant or Distributed Data	46	39	32	45%	14
3	3.4	Consistency- Temporal Consistency	38	20	21	82%	17
3	3.2	Consistency- Format Consistency	54	26	33	65%	21
2	2.3	Completeness- Truncation	46	33	24	92%	22
5	5.2	Timeliness- Manual Float	31	13	9	256%	22
4	4.2	Validity- Values Conform to Business Rule	62	48	39	56%	22
10	10.5	Representation- Includes Measurement Units	38	15	13	186%	25
11	11.1	Currency- Current with World it Models	38	24	13	208%	26
8	8.3	Precision- Domain Precision	38	4	n/a	n/a	#VALUE!



Appendix 7- Respondent Role

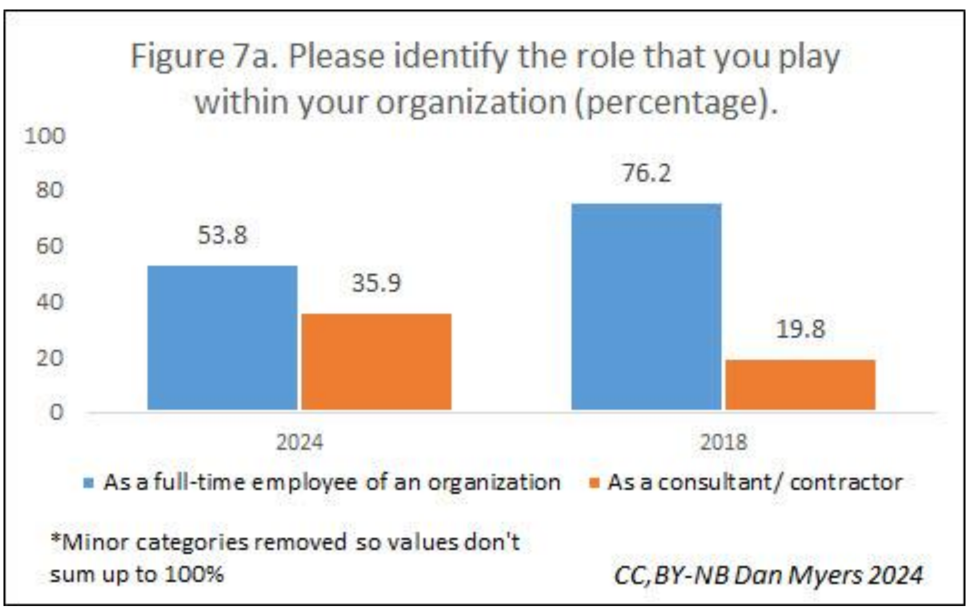
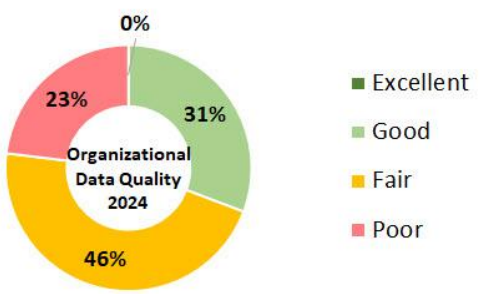


Figure 7a. Please identify the role that you play within your organization (percentage).

Role	2024	2018	2017	2016	2015
As a full-time employee of an organization	53.8	76.2	85.4	75.0	69.1
As a consultant/ contractor	35.9	19.8	10.4	8.3	19.9
Other	0	3	0.0	2.1	0.0
As a teacher	5.1	1	0.0	0.0	2.2
As a vendor	2.6	0	2.1	0.0	2.2
As a retired professional	2.6	0	0.0	0.0	0.7
As a part-time employee of an organization	0	0	2.1	0.0	0.7
As a student	0	0	0.0	0.0	5.1
Decline to State	0	0	0.0	14.6	0.0
	100	100	100	100	99.9

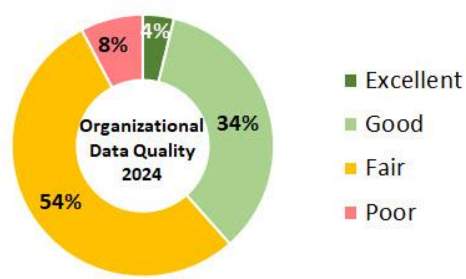
Appendix 8- Organizational DQ Levels

Figure 8a. Organizational Data Quality Level
(Subset of respondents who **Do Not** Use the Dimensions of Data Quality for Data Defects)



CC,BY-NB Dan Myers, DQMatters.com 2024, n=39

Figure 8b. Organizational Data Quality Level
(Subset of respondents who **Do** Use the Dimensions of Data Quality for Data Defects)



CC,BY-NB Dan Myers, DQMatters.com 2024, n=39

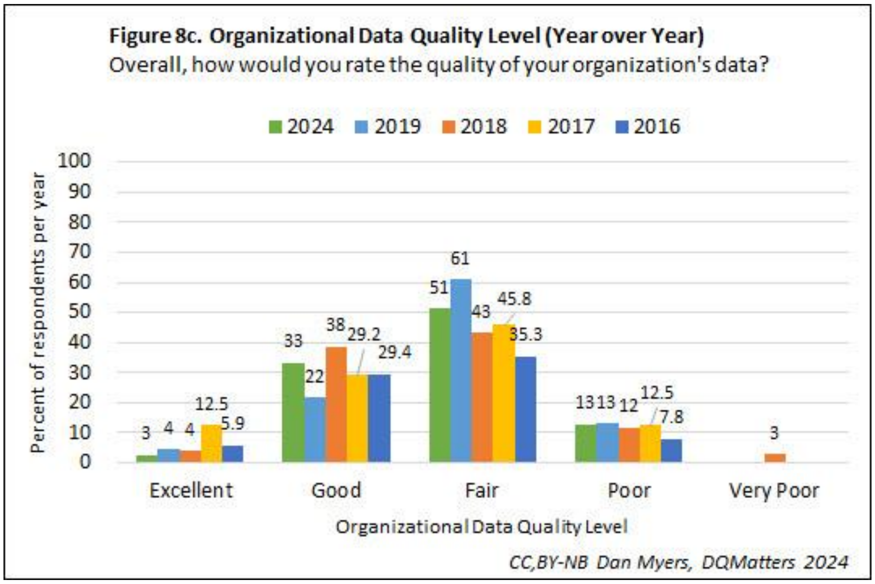
Table 8a. Overall data quality for organizations that use the Dimensions or don't use them

	Poor	Fair	Good	Excellent	Grand Total
Don't Use Dimensions of DQ	3	6	4	0	13
Use Dimensions of DQ	2	14	9	1	26
Grand Total	**5**	**20**	**13**	**1**	**39**

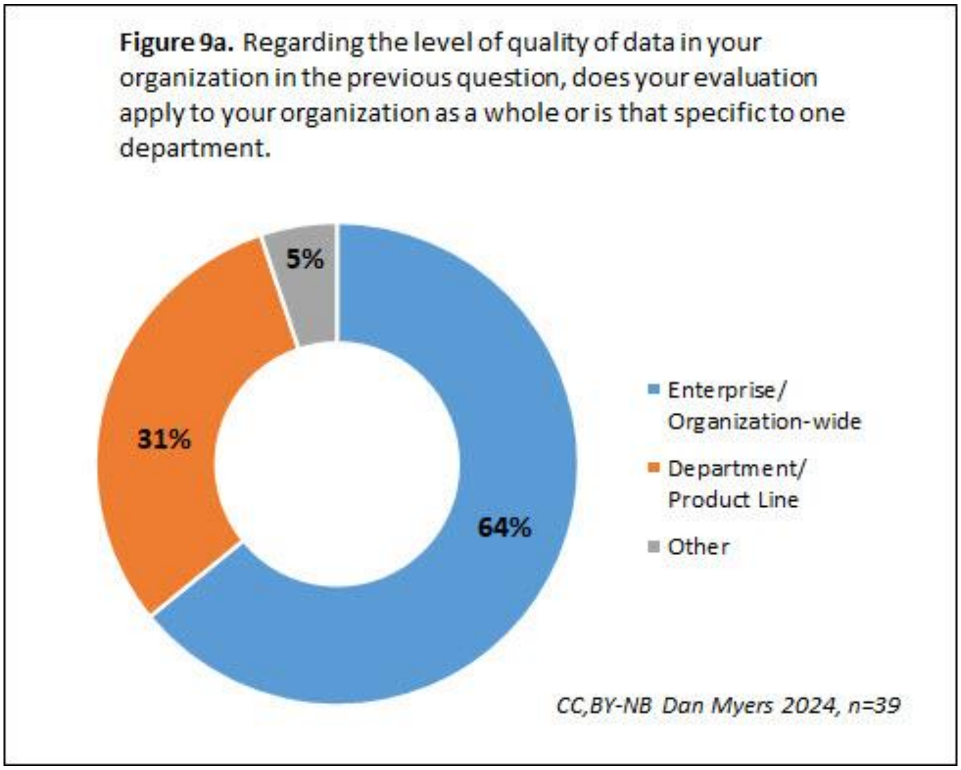
	Poor	Fair	Good	Excellent	Grand Total
Don't Use Dimensions of DQ	8%	15%	10%	0%	13
Use Dimensions of DQ	5%	36%	23%	3%	26
Grand Total	**5**	**20**	**13**	**1**	**39**

How much better by Using Dimensions?

Poor	Fair	Good	Excellent
3%	21%	13%	3%



Appendix 9- Scope of DQ Levels Provided by Respondent



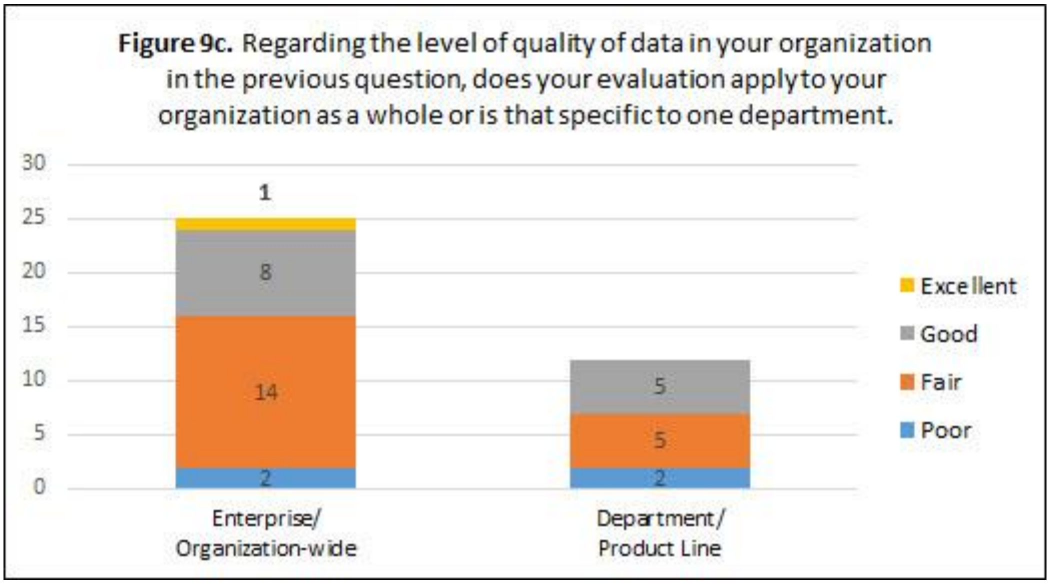


Figure 9c. Regarding the level of quality of data in your organization in the previous question, does your evaluation apply to your organization as a whole or is that specific to one department.
30
25
20
15
10
5
0
1
8
14
2
5
5
2
Excellent
Good
Fair
Poor
Enterprise/ Organization-wide
Department/ Product Line