

Een managementsysteem voor datakwaliteit

Naar een structurele, samenhangende en gestandaardiseerde benadering

Werkgroep Data Quality van DAMA-NL

Peter van Nederpelt, Kamal Ahmed, Andrew Black, Marco Heij, Aris Prins, René Wiertz, Dirk Huetling, Laurens van der Drift

Datakwaliteit, factsheet, ISO 9001, Data Quality Management System

Inleiding

In het data domein is ISO 27001 (*Eisen voor een informatiebeveiligingsmanagementsysteem*) een bekende standaard op het gebied van informatiebeveiliging. Deze standaard wordt veel toegepast en organisaties laten zich hiervoor certificeren. Op deze manier kunnen organisaties aan de buitenwereld tonen, dat zij de informatiebeveiliging op orde hebben.

Voor datakwaliteit bestaan vergelijkbare standaarden, namelijk ISO 8000 (*Data quality*) en ISO 9001 (*Eisen voor een kwaliteitsmanagementsysteem*), maar de eerstgenoemde wordt in een te specifieke context toegepast en de laatstgenoemde wordt zelden toegepast. Belangrijke oorzaak in het geval van ISO 9001 is, dat deze standaard generiek is en niet specifiek op data van toepassing is. ISO 9001 is bedoeld voor kwaliteit van producten en diensten in het algemeen.

In dit white paper wordt daarom een vertaalslag gemaakt van ISO 9001 naar datakwaliteit. De termen voor de elementen die in ISO 9001 worden gebruikt, worden omgezet naar termen die in het data domein gangbaar zijn. Vervolgens wordt kort toegelicht, wat de betekenis van elk van deze elementen is. Doel hiervan is om het makkelijker te maken om ISO 9001 ook toe te passen op het managen van datakwaliteit.

Managementsysteem

In het data domein wordt er in het algemeen gesproken over datamanagement en data governance. ISO hanteert een ander concept, namelijk dat van een managementsysteem. Dit geldt zowel voor ISO 27001 als ISO 9001. Een managementsysteem wordt gezien als een geheel van samenhangende of elkaar beïnvloedende elementen gericht op het vaststellen van doelstellingen van een organisatie en het bereiken van deze doelstellingen.

De ISO-standaarden worden na een wereldwijde consultatie elke vijf jaar aangepast. Certificering voor een ISO-standaard kan plaatsvinden door een externe auditor die hiervoor is geaccrediteerd. Een managementsysteem kan betrekking hebben op verschillende disciplines, zoals kwaliteitsmanagement en informatiebeveiliging, maar ook milieumanagement en financieel management. De ISO-standaarden voor managementsystemen bevatten een gemeenschappelijke kern: de high level structure. Dit, omdat de verschillende managementsystemen ook zaken met elkaar gemeen hebben.

Motieven voor het invoeren van een Data Quality Management System (DQMS)

Redenen om een DQMS te willen invoeren, zijn, dat een organisatie datakwaliteit op orde wil brengen en houden. Dit kan bijvoorbeeld nodig zijn om data voor meerdere doeleinden te gebruiken. In eerste plaats wordt data gebruikt in de reguliere processen, zoals inkoop, productie, verkoop, financiën en HR. Onvoldoende datakwaliteit hierbij leidt tot inefficiënties en irritaties, zowel intern als extern.

In de tweede plaats wordt data gebruikt voor bijvoorbeeld rapportages aan toezichthouders en de maatschappij, voor data gedreven werken, voor het sturen op basis van Key Performance Indicators

(KPI's) en voor het trainen, evalueren en verbeteren van machine learning systemen binnen het domein van kunstmatige intelligentie (AI). Het gebruik van data van onvoldoende kwaliteit kan hierbij leiden tot onjuiste beslissingen, met als gevolg: boetes, levensbedreigende situaties voor individuen of groeperingen, en imagooverlies.

In het financiële domein is het al heel gewoon om aandacht te schenken aan de kwaliteit van de financiële data. Een jaarrekening moet immers in orde zijn en aandeelhouders moeten erop kunnen vertrouwen. Het is niet voor niets dat accountants de jaarrekening jaarlijks controleren van organisaties. Bovendien is het binnen dit domein gebruikelijk om het financiële managementsysteem in de vorm de administratieve organisatie (AO) op orde te hebben.

Alternatieve kaders

In de wereld van datamanagement zijn er kaders die lijken op het DQMS dat in dit paper wordt toegelicht. Allereerst bestaat al geruime tijd de standaard ISO 8000. Deze bestaat uit 13 delen, maar heeft vooral een technisch karakter en richt zich op datakwaliteit van master- en referentiedata. Verder is er het DECAM-systeem. Dit systeem is gericht op datamanagement in zijn geheel en heeft dus een bredere scope. DAMA DMBOK is een body of knowledge en kan niet echt een toetsingskader of framework worden genoemd. Een DQMS is helemaal gericht op datakwaliteit.

Factsheets

De werkgroep Data Quality van DAMA Nederland heeft zich als doel gesteld om alle elementen van het DQMS in overzichtelijke documenten te beschrijven, ook wel factsheets genoemd. De factsheets zijn op de site van DAMA-NL in de datamanagement wiki gepubliceerd (www.dama-nl.org/datakwaliteit). Hierbij is gekozen voor de wiki vorm vanwege de toegankelijkheid en de onderhoudbaarheid van de factsheets.

Een factsheet bestaat uit verschillende rubrieken: een definitie van het element, het doel van het element, relaties met andere elementen, de PDCA-cyclus (Deming) van een element, en de eigenschappen van een element. Ook worden er voorbeelden gegeven en wordt er een verhaal verteld, waaruit blijkt hoe het element toegepast kan worden. De beschrijving van elk element is/wordt uit verschillende bronnen afgeleid zoals DAMA-DMBOK2, ISO 9001, boeken en artikelen die op Internet staan. Zie de lijst van referenties aan het einde van dit paper.

Eisen aan de elementen van het managementsysteem

In dit paper gaan we nog niet zover, dat we de eisen specificeren die aan elke element kan worden gesteld, bijvoorbeeld aan welke eisen moet een datakwaliteitsbeleid voldoen. Dat is ook niet echt nodig, want deze eisen staan eigenlijk al in de ISO 9001 standaard.

Certificering

Als er maar voldoende aandacht wordt besteed aan alle elementen, is het mogelijk om je als organisatie te laten certificeren voor ISO 9001 door een externe, certificerende instantie (externe audit), ook al gaat het om datakwaliteit. Je beperkt de scope van het certificaat dan tot de data verwerkende processen.

Selectie en invoeringsvolgorde

In dit paper wordt geen advies gegeven, welke elementen relevant zijn voor een organisatie of in welke volgorde deze ingevoerd zouden moeten worden. Dit is erg afhankelijk van de omgeving, waarin deze worden toegepast, welke bronnen men tot zijn of haar beschikking heeft en hoever de volwassenheid van de organisatie gevorderd is. Overigens is het bij certificering voor ISO 9001 wel vereist, dat alle elementen van kracht zijn en een invoeringsvolgorde wordt niet aangegeven. Dit mag een organisatie zelf bedenken.

PDCA-cyclus van de elementen

We gaan ervanuit dat er voor elk element een PDCA-cyclus is. Dat wil zeggen dat een element wordt voorbereid (“Plan”), wordt uitgevoerd (“Do”), wordt geëvalueerd (“Check”) en zo nodig wordt verbeterd (“Act”). Alle elementen bevinden zich dus in een verbetercyclus. De cyclus zal daarom niet voor elk element worden herhaald in dit paper.

Doelstellingen en elementen van het DQ-managementsysteem (DQMS)

In tabel 1 staan alle doelstellingen en elementen van het DQMS en worden de termen uit ISO 9001 vertaald naar termen die in het data domein gangbaar zijn.

Tabel 1 Afbeelding van elementen van ISO 9001 op het DQMS

ISO 9001 paragraaf	ISO 9001 Elementen	2 Doelstellingen en 28 Elementen van het DQMS
4.1	Understanding the organization and its context	Context analysis
4.2	Understanding the needs and expectations of interested parties	Stakeholdersanalysis Compliance
4.3	Determining the scope of the quality management system	Scope
5.1	Leadership and commitment	Leadership
5.2	Policy	Data quality policy
5.3	Organizational roles and responsibilities	Roles and responsibilities
6.1	Actions to address risks and opportunities	Risk analysis
6.2	Quality objectives and planning to achieve them	Data quality objectives
7.1	Resources	Resources
7.2	Competence	Staff competence
7.3	Awareness	Awareness of data quality
7.4	Communication	Communication about Data Quality
7.5	Documented information	Metadata
7.5.3	Control of documented information	Document control
8.1	Operational planning and control	Data processes
8.2	Requirements for products and service	<i>Doelstelling:</i> Meeting data quality requirements
		Critical data elements
		Data quality rules
8.3	Design and development of products and services	Data design processes
8.4	Control of externally provided processes, products and services	Data suppliers
8.5.2	Identification and traceability	Data lineage
8.7	Control of non-conforming outputs	Data cleansing Data issues (1)
9.1	Monitoring, measurement, analysis en evaluation	Data profiling Monitoring Data Quality
9.1.2	Customer satisfaction	<i>Doelstelling:</i> Data consumer satisfaction
9.2	Internal audit	Internal audit
9.3	Management review	Management review
10.2	Nonconformity and corrective action	Data issues (2)
10.3	Continual improvement	Continuous improvement

Doelstellingen van het DQMS

Een DQMS kent de twee onderstaande doelstellingen en deze liggen ook in elkaars verlengde: doelstelling 1 draagt ook bij aan doelstelling 2.

1. Datakwaliteitseisen

De eerste doelstelling is, dat wordt voldaan aan de eisen die worden gesteld aan de kwaliteit van de data. Datakwaliteit kent veel dimensies. De werkgroep Data Quality van DAMA NL heeft er 60 geïdentificeerd en van een gestandaardiseerde definitie voorzien. Hieruit moet uiteraard een keuze worden gemaakt. Dit geldt ook voor de data waarop deze dimensies betrekking hebben. Hierbij kan het element 'critical data element' een rol spelen.

2. Datagebruikerstevredenheid

De tweede doelstelling is tevredenheid van de dataconsumenten: degenen die de data gebruiken binnen en buiten de organisatie. De perceptie van de dataconsumenten over datakwaliteit wordt op de één of andere manier gemeten. Uit deze metingen worden conclusies getrokken en zo nodig worden acties gepland en uitgevoerd.

Elementen van het DQMS

In deze paragraaf worden alle elementen van het DQMS kort toegelicht. De elementen worden beschreven in volgorde van het metamodel in figuur 1. Voor een meer uitgebreide, gestructureerde beschrijving van de elementen wordt verwezen naar de factsheets van DAMA-NL in de data management wiki (https://datamanagement.wiki/overview/overview_data_quality_management_system).

1. Compliance

Naleving van wetten, regels en normen is de mate waarin gegevens in overeenstemming zijn met wetten, regels en normen.

2. Datakwaliteitsdoelstellingen

Bepaal KPI's op gebied van datakwaliteit. Dat is zinvol als bepaalde data en kwaliteitsdimensies goed in de gaten moeten worden gehouden. Als bepaalde streefwaarden worden overschreden is het nodig om verbeteracties uit te voeren.

3. Datakwaliteitsbeleid

Geef in het datakwaliteitsbeleid aan welke data en welke kwaliteitsdimensies belangrijk zijn voor de organisatie. Ook kan hier worden genoemd, welke wettelijke eisen van toepassing zijn voor wat betreft datakwaliteit.

4. Interne audit

Ga na of alle elementen van het managementsysteem nog functioneren en of verbetering gewenst is. Dit is ook input voor het managementreview. Een interne audit kan ook door een externe partij worden uitgevoerd. Bij een externe audit daarentegen gaat het altijd om een certificatieproces.

5. Management review

Bespreek als directie periodiek (jaarlijks) of het managementsysteem nog effectief is en of aanpassing nodig is. Hierbij wordt de directie verantwoordelijk gemaakt voor het managementsysteem, zoals voor alle andere bedrijfsfuncties zoals inkoop, productie, marketing, verkoop, financiën, personeel en ICT.

6. Leiderschap

Laat als directie duidelijk merken, dat je datakwaliteit belangrijk vindt en waarom.

7. Scope

Het gebied waarover een activiteit zich uitstrekt.

8. Stakeholder

Leg in de stakeholdersanalyse vast wie belang heeft bij datakwaliteit en wat de verwachtingen hierover zijn. Daar kan dan rekening mee worden gehouden.

9. Dataleverancier

Evalueer periodiek, of data die door externe partijen worden geleverd, voldoen aan de gestelde inputeisen. Dit kan leiden tot overleg met deze partijen, om tot verbetering te komen. Indien nodig en mogelijk, wordt overgestapt op andere leveranciers.

10. Datakwaliteit communicatie

Bepaal met wie en hoe over datakwaliteit wordt gecommuniceerd. Dit kan via overlegstructuren, maar ook andere middelen zijn mogelijk.

11. Rollen en verantwoordelijkheden

Onderscheid op gebied van data en in het bijzonder datakwaliteit verschillende rollen. Daarvan zijn data-eigenaar en datasteward de meest genoemde. Duidelijk moet zijn wie deze rollen vervult en wat hun verantwoordelijkheden zijn.

12. Competenties medewerker

Zorg ervoor, dat medewerkers die met data werken, competent zijn om hun taak te vervullen. Het volgen van opleidingen en cursussen kunnen daarbij nodig zijn.

13. Datakwaliteitsbewustzijn

Maak medewerkers bewust van het belang van datakwaliteit. Ontwikkeling van datageletterdheid kan hierbij van nut zijn.

14. Gedocumenteerde informatie

Zorg ervoor dat documentatie goed wordt beheerd, zodat de inhoud, afkomst en de laatste versie duidelijk is (titel, auteur, versie).

15. Risicoanalyse

Inventariseer mogelijke situaties en gebeurtenissen die kunnen leiden tot onvoldoende datakwaliteit. Hierbij wordt ook nagegaan of al genomen maatregelen afdoende zijn dan wel aanvullende maatregelen nodig zijn.

16. Context analyse

Geef in de context analyse de sterke en zwakke punten en de kansen en bedreigingen m.b.t. datakwaliteit weer (SWOT-analyse). Op grond hiervan kunnen al verbeteracties worden vastgesteld.

17. Continue verbetering

Richt een proces in voor het managen van verbeteracties op gebied van datakwaliteit. Hierbij is ook ruimte voor spontane verbeter suggesties van medewerkers.

18. Data model

Beschrijf welke dataverzamelingen er aanwezig zijn en hoe deze met elkaar zijn verbonden. Hierbij kunnen ook datamodellen worden toegepast, die de datastructuren weergeven.

19. Dataproces

Beschrijf het dataverwerkingsproces al dan niet in combinatie met het werkproces. Dit bevordert de overdraagbaarheid van kennis van de processen en maakt verbeteracties beter mogelijk.

20. Resource

Resources om rekening mee te houden bij het opstellen van een datakwaliteit business case om data te schonen en de kwaliteit te verbeteren zijn medewerkers, applicaties, data platformen en budget.

21. Metadata

Beschrijf onder andere de betekenis van de velden in de bestanden. De inhoud van de velden moet daarbij passen. Veldnamen zijn niet altijd zelfverklarend.

22. Datakwaliteitsregels

Bepaal datakwaliteitsregels die nodig zijn om onjuiste input of bewerkingen te voorkomen, maar ook om mogelijke fouten achteraf te signaleren. Deze regels zijn ook nodig bij data cleansing.

23. Kritieke data elementen

Bepaal de velden die van essentieel belang zijn voor de organisatie. Verbetering van de datakwaliteit kan het beste in eerste instantie hierop gericht zijn.

24. Data lineage

Maak duidelijk via welke weg output tot stand is gekomen. Dit is van nut bij het opsporen van afwijkingen en bij voorgenomen wijzigingen in het proces.

25. Data cleansing

Verbeter in dit proces de datakwaliteit eenmalig. Conversies en migraties zijn hier vaak aanleiding voor. Naast eenmalige verbetering is het uiteraard ook nodig om ervoor te zorgen, dat afwijkingen in de toekomst zo veel mogelijk worden voorkomen.

26. Datakwaliteit monitoring

Bewaak periodiek dan wel continue de datakwaliteit. Als er iets fout gaat, kan er worden ingegrepen. Monitor ook of het DQMS effectief is m.a.w. leiden alle elementen samen ertoe dat de doelstellingen van het DQMS worden behaald.

27. Datakwaliteit issue

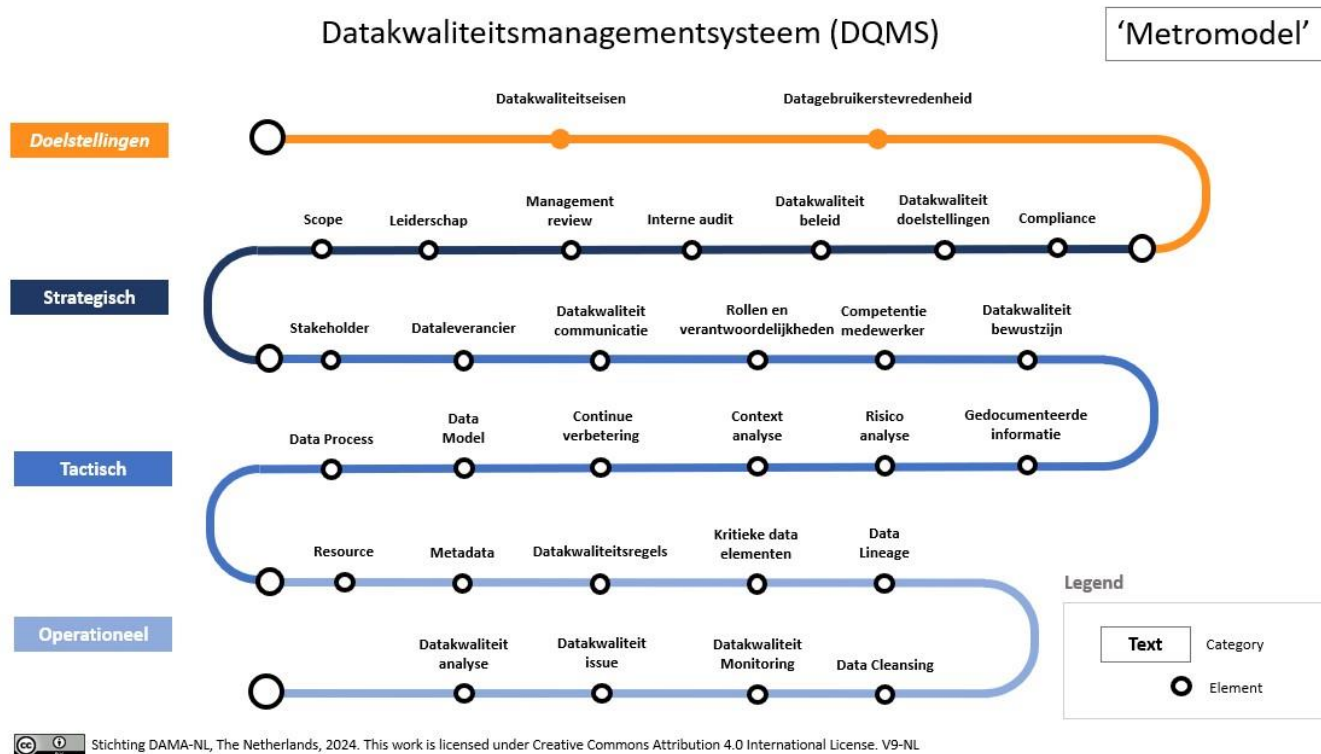
Handel afwijkingen in de data adequaat af ('correctie'). Ook moet worden bekeken of deze afwijkingen in de toekomst voorkomen kunnen worden ('correctieve actie'). De organisatie moet weten waar ze terecht kunnen voor een issue (afwijking).

28. Datakwaliteitsanalyse

Spoor afwijking op via data-analyse. Hierbij wordt soms ook AI toegepast.

Metromodel

In figuur 1 zijn de doelstellingen en elementen van het DQMS nog een keer samengevat in de vorm van een metromodel.



Figuur 1 Doelstellingen en elementen van het DQMS

Conclusie

Het is van belang dat een organisatie de datakwaliteit op orde heeft. Het toepassen van een managementsysteem op dit gebied kan daarbij zeker helpen. ISO 9001 biedt alle handvaten om dit op een gestandaardiseerde manier te doen.

Ook kan een organisatie zich laten certificeren door een certificerende instelling (externe auditor), zodat ook aan de buitenwereld kan worden getoond, dat de datakwaliteit goed wordt gemanaged.

We kunnen concluderen, dat het goed mogelijk is om het concept van een managementsysteem toe te passen op datakwaliteit, ondanks dat er een ander vakjargon wordt gebruikt.

Tot slot

De werkgroep Data Quality van DAMA-NL stelt verder commentaar op dit document zeer op prijs. Reacties kunnen worden gestuurd naar WGDQ@planetbox.nl. De werkgroep is vooral benieuwd, welke vragen het DQMS oproept, welke overwegingen de invoering van een dergelijk kader in de weg kunnen staan, en welke potentie de DQMS heeft voor uiteenlopende organisaties.

Referenties

- DAMA (2017). DAMA-DMBOK. *Data Management Body of Knowledge*. 2nd Edition. [Technics PublicationsLlc](https://www.dama-nl.org/publications). August 2017.
- DAMA Dictionary of Data Management. 2nd Edition 2011. Technics Publications, LLC, New Jersey.
- Data quality. (2021, December 4). Stichting Dama. <https://www.dama-nl.org/datakwaliteit>
- DECAM. [About DCAM - EDM Council](https://www.dama-nl.org/decam)
- Factsheets van DAMA-NL. <https://datamanagement.wiki/>
- ISO 27001:2017. *Informatietechnologie – Beveiligingstechnieken – Managementsystemen voor informatiebeveiliging – Eisen*
- ISO 8000. 13 parts.
- ISO 9000:2015. *Quality Management Systems – Requirements*.

ISO 9001:2015. *Quality Management Systems – Fundamentals and vocabulary.*

Versie 2.4 d.d. 11 december 2024