

Een managementsysteem voor datakwaliteit

Naar een meer structurele, samenhangende en gestandaardiseerde benadering

Werkgroep Data Quality van DAMA-NL

Peter van Nederpelt, Kamal Ahmed, Andrew Black, Marco Heij, Aris Prins, René Wiertz, Merijn Douwes, Dirk Hueting, 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 (DQ MS)

Redenen om een DQ MS 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 imagoverlies.

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 DQ MS 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 DQ MS is helemaal gericht op datakwaliteit.

Factsheets

De werkgroep Data Quality van DAMA Nederland heeft zich als doel gesteld om alle elementen van het DQ MS in overzichtelijke documenten te beschrijven, ook wel factsheets genoemd. Een aantal hiervan zijn al gereed en op de site van DAMA-NL 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 (DQ MS)

In tabel 1 staan alle doelstellingen en elementen van het DQ MS 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 DQ MS

ISO 9001 paragraaf	ISO 9001 Elementen	2 Doelstellingen en 25 Elementen van het DQMS
4.1	Understanding the organization and its context	Issue analysis
4.2	Understanding the needs and expectations of interested parties	Stakeholdersanalyse
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
7.2	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 DQ MS

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

1 Data quality requirements

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 Data consumer satisfaction

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 DQ MS

In deze paragraaf worden alle elementen van het DQ MS kort toegelicht. De elementen worden beschreven in volgorde van de laatste kolom van tabel 1. Voor een meer uitgebreide, gestructureerde beschrijving van de elementen wordt verwezen naar de factsheets van DAMA-NL.

1 Issue analysis

In een issue analyse wordt nagegaan, wat de belangrijkste interne en externe problemen zijn met betrekking tot datakwaliteit. Dit is van belang om prioriteiten te stellen bij verbeteracties en je te concentreren op de hoofdproblemen.

2 Stakeholdersanalyse

In een stakeholdersanalyse wordt nagegaan, wie betrokken zijn bij het dataverwerkingsproces en welke belangen, verwachtingen en percepties zij hebben bij datakwaliteit. Belanghebbenden ('stakeholders') kunnen worden aangetroffen aan de kant van de input (suppliers), de throughput en de output (interne en externe dataconsumenten). In het managen van datakwaliteit wordt rekening gehouden met de invloed die stakeholders hebben op datakwaliteit en op de impact die datakwaliteit heeft op hun functioneren.

3 Leadership

Een DQ MS komt niet van de grond zonder betrokkenheid van het management. Het vereist naast een bottom-up aanpak ook een top-down aanpak. Het management is trouwens ook direct betrokken bij het managementsysteem via het management review.

4 Data quality policy

In het datakwaliteitsbeleid staat welke intenties het management heeft met betrekking tot datakwaliteit. Welke dimensies van datakwaliteit zijn bijvoorbeeld van belang en om welke data gaat het? Ook kan hierin worden genoemd welke wet- en regelgeving van toepassing is op datakwaliteit. Datakwaliteitsbeleid geeft richting aan acties op gebied van datakwaliteit.

5 Roles and responsibilities

Roles en responsibilities wil zeggen, dat duidelijk is wie, welke rollen en verantwoordelijkheden hebben met betrekking tot datakwaliteit. Rollen die daarbij vaak worden genoemd zijn de Chief Data Officers, de data owner en de data steward, maar ook andere medewerkers spelen zeker een rol bij datakwaliteit.

6 Risk analysis

Bij een risk analysis wordt nagegaan welke gebeurtenissen of oorzaken de doelstellingen van het systeem (voldoende datakwaliteit, tevredenheid dataconsumenten en dataproducten) kunnen bedreigen. Ook wordt beschreven, welke maatregelen al zijn genomen om dit risico te verminderen. Tot slot wordt beoordeeld of de maatregelen afdoende zijn, of dat er aanvullende maatregelen nodig zijn.

7 Data quality objectives

Bij quality objectives worden op een meetbare manier geformuleerd aan welke eisen de data moet voldoen. Dit wordt vervolgens ook gemeten en aan een norm getoetst. Als niet aan de norm wordt voldaan, zijn verbeteracties gewenst. Het gaat hier om het dashboard met betrekking tot datakwaliteit.

8 Staff competence

Medewerkers die betrokken zijn bij dataverwerking, dienen door het management en HR ondersteund te worden bij het opdoen van kennis en vaardigheden om hun taak goed uit te kunnen voeren. Het is duidelijk, dat als medewerkers onvoldoende competent zijn, dit gevolgen zal hebben voor de datakwaliteit.

9 Awareness of Data Quality

Awareness of Data Quality is een erkenning of besef van eenieder in de organisatie, dat alleen met de juiste vaardigheden en competenties, data van voldoende kwaliteit kan worden gerealiseerd. Zij zijn zich ervan bewust dat er een kwaliteitsbeleid is, dat er kwaliteitsdoelstellingen zijn, en wat hun bijdrage is aan een effectief DQ MS.

10 Communication about Data Quality

Communicatie over datakwaliteit heeft tot doel om stakeholders bewust te maken van het belang van datakwaliteit en hun rol daarin. Via communicatie wordt ook duidelijk gemaakt welke verbeteringen al zijn uitgevoerd en wat de geplande verbeteracties zijn.

11 Metadata

Metadata beschrijft de betekenis van data. Zonder metadata kan de kwaliteit van data niet worden vastgesteld. Ook is data zonder metadata betekenisloos voor dataconsumenten. Metadata is juist, volledig en duidelijk. Het wordt ook onderhouden.

12 Document control

Van alle documentatie rond datakwaliteit die worden onderhouden, is de versie bekend en is duidelijk welke aanpassingen hebben plaatsgevonden. Hiermee wordt voorkomen, dat niet duidelijk is, welke versie van toepassing is.

13 Data processes

Dataverwerkingsprocessen zorgen ervoor, dat input wordt verwerkt tot output met de juiste datakwaliteit. Output kan ook bestaan uit een database die geraadpleegd wordt door medewerkers of systemen.

14 Critical Data Elements

Critical Data Elements zijn attributen die van wezenlijk belang zijn voor de organisatie. Dit kan ook in gradaties worden uitgedrukt. Bij het management van datakwaliteit zijn dit de attributen of velden waar het eerst aan moet worden gedacht. Als het handiger is om bepaalde dataverzamelingen als kritisch te beschouwen, dan is daar natuurlijk geen enkel bezwaar tegen (critical data sets).

15 Data quality rules

Datakwaliteitsregels zijn regels waaraan data moeten voldoen. Als deze worden geïmplementeerd in de systemen, kunnen hiermee fouten worden voorkomen. Ook datakwaliteitsregels moeten worden gemanaged. Zij kunnen incompleet zijn/worden of niet kloppen. Er kan daardoor een datakwaliteitslek ontstaan.

16 Design processes

Ontwerpprocessen leiden ertoe, dat de juiste data wordt geregistreerd. Daarvoor wordt de informatiebehoefte van dataconsumenten in kaart gebracht. Bovendien wordt de data in de juiste structuur gebracht (datamodel).

17 Data suppliers

Data suppliers kunnen bestaan uit organisaties die bestanden aanleveren, maar ook gebruikers van binnen en buiten de organisatie die data registreren, kunnen als data suppliers worden beschouwd. Beoordeeld wordt, of de suppliers input van voldoende kwaliteit leveren. Is dit niet het geval, dan zijn maatregelen nodig.

18 Data lineage

Data lineage is metadata dat beschrijft wat de herkomst van data is. Via deze metadata kan worden getraceerd wat de bron is van bepaalde output en hoe de data is verwerkt. Soms is dit van belang, om je als organisatie te kunnen verantwoorden. Ook kan deze metadata van belang zijn bij aanpassingen en herontwerpen. Deze mogen niet leiden tot te grote terugval van datakwaliteit.

19 Data profiling

Data profiling gaat vooraf aan data cleansing. Bij data profiling wordt de datakwaliteit via statistieken in beeld gebracht. Er worden uitspraken gedaan in cijfers over gehele dataverzamelingen. Op basis hiervan kan worden gekozen, welke data in aanmerking komt voor verbetering van kwaliteit. Data profiling kan ook worden ingezet bij monitoring van data quality.

20 Data Cleansing

Bij data cleansing worden data issues actief opgespoord en vervolgens verholpen. Het DQ MS kent een werkwijze om data cleansing uit te voeren, al dan niet met behulp van een tool.

21 Monitoring data quality

Een monitoring proces zorgt ervoor, dat de kwaliteit van de data voortdurend wordt bewaakt. Dit zal vooral op basis van data quality rules plaatsvinden. Als de ontwikkeling in een tijdlijn zichtbaar wordt gemaakt, kunnen trends worden onderkend en kan tijdig worden ingegrepen.

22 Internal audit

Bij een interne audit wordt onderzocht namens het management, of alle elementen van het DQ MS aantoonbaar aanwezig zijn en of deze ook effectief zijn. De resultaten van een interne audit worden gerapporteerd aan het management. Deze bepaalt op basis van de uitkomsten, of verbeteringen in het systeem gewenst zijn.

23 Management review

Het management beoordeelt periodiek of het DQ MS voldoende effectief is en of het verbetering behoeft. Zij doet dit op basis van meerdere rapportages en documentatie. Hiermee wordt het management verantwoordelijk gemaakt voor het systeem. Deze verantwoordelijkheid kan niet worden gedelegeerd.

24 Data issues

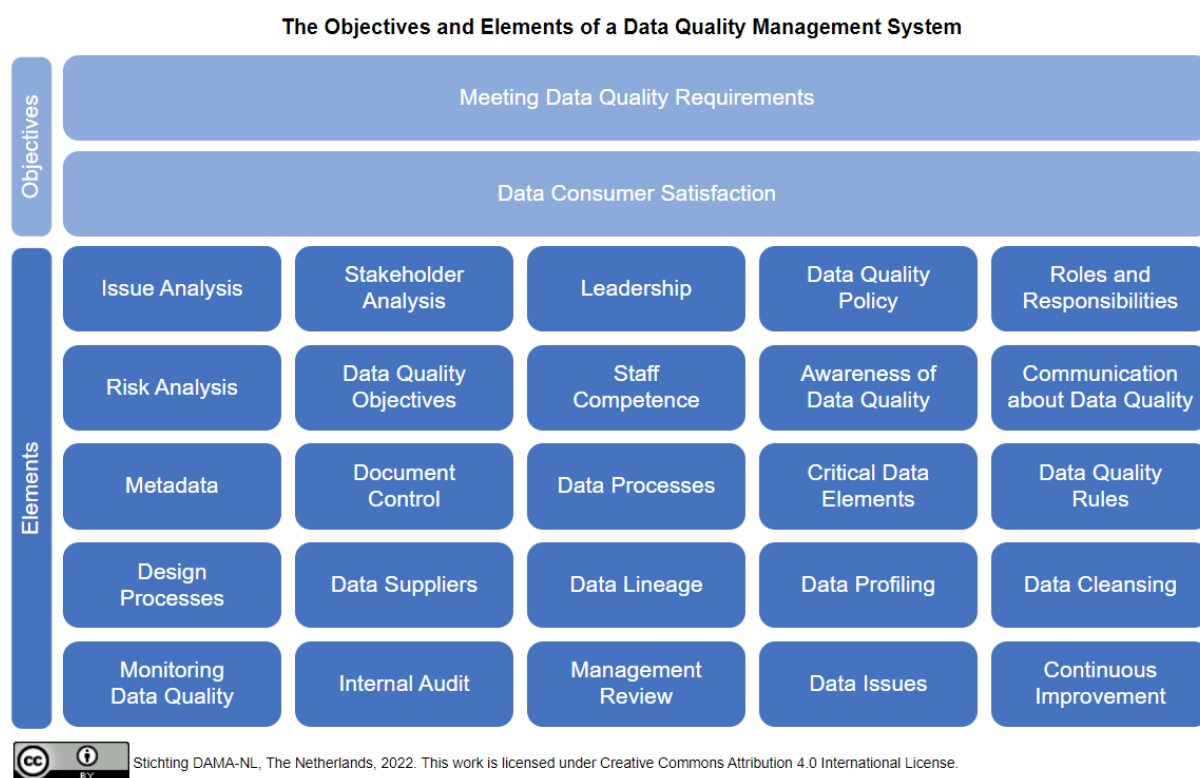
Het afhandelen van data issues is een belangrijk en concreet element binnen een DQ MS. Als een medewerker één of meer data issues tegenkomt, kan hij deze melden. Vervolgens worden deze beoordeeld en afgehandeld ('correction'). Ook de wijze van afhandeling wordt geregistreerd en terug gemeld aan de melder. Periodiek wordt beoordeeld, of er een patroon is te herkennen in de data issues en of er maatregelen kunnen worden genomen om de issues te voorkomen ('corrective actions').

25 Continuous improvement

Continuous improvement moet er niet alleen voor zorgen, dat alle elementen van het managementsysteem periodiek worden geëvalueerd en verbeterd. Er moet ook ruimte zijn voor spontane ideeën voor verbetering. Dat wil zeggen, dat er een werkwijze is die dit ondersteunt.

Diagram

In figuur 1 zijn de doelstellingen en elementen van het DQ MS nog een keer samengevat in de vorm van een diagram (in het Engels).



Figuur 1 Doelstellingen en elementen van het DQ MS

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 DQ MS oproept, welke overwegingen de invoering van een dergelijk kader in de weg kunnen staan, en welke potentie de DQ MS heeft voor uiteenlopende organisaties.

Referenties

- DAMA (2017). DAMA-DMBOK. *Data Management Body of Knowledge*. 2nd Edition. [Technics PublicationsLlc](#). 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](#)
- 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 2.2 d.d. 17 maart 2022