

Datakwaliteit in beeld

Hoe kun je datakwaliteit bekijken en verbeteren voor onderwijsteams in het MBO

Inhoudsopgave

INTRODUCTIE.....	1
HOOFDSTUK 1. - WAT IS DATAKWALITEIT?	2
Aspecten van datakwaliteit	2
Van belang op verschillende momenten in het proces	3
HOOFDSTUK 2. - STAPPENPLAN OM DE DATAKWALITEIT TE METEN EN VERBETEREN.....	4
Stap 1: Voorbereiding en inventarisatie – wat is het vertrekpunt.....	4
Stap 2: Situatie in beeld - aan de slag met de indicatoren	4
Stap 3: Diagnose & onderliggende oorzaak.....	4
Stap 4: Interventie.....	5
Stap 5: Borging & opname in de organisatie	6
TIPS EN OVERWEGINGEN.....	7

Introductie

Vanuit de datacoalitie datagedreven onderzoek wordt gekeken welke kansen data biedt aan het MBO-onderwijs. Om die kansen te kunnen benutten moet de data wel 'op orde zijn'. Data van **onbekende** of **onvoldoende** kwaliteit verzwakt onderbouwing, creëert geen vertrouwen en kan uiteindelijk leiden tot verkeerde keuzes. Dit document bevat een instrument dat helpt om de kwaliteit van data in kaart te brengen (diagnose), de aanleiding in kaart te brengen (oorzaak) en oplossingsrichtingen te adviseren (interventie). Het document bestaat een introductie (hoofdstuk 1) en het instrument (hoofdstuk 2).

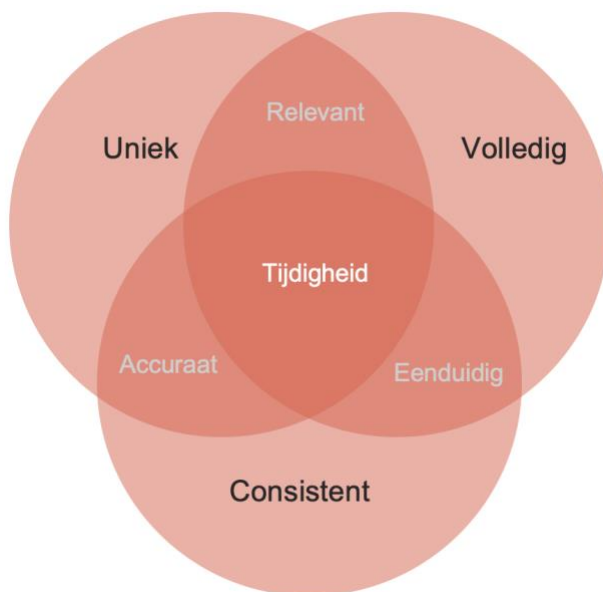
Hoofdstuk 1. - Wat is datakwaliteit?

In een onderwijsinstelling is veel data beschikbaar. En de beschikbaarheid van data neemt, als gevolg van sterk toenemende digitalisering, alleen maar toe. In de afgelopen jaren maken we steeds vaker gebruik van deze data om dashboards te vullen met KPI's voor ondersteunend personeel en voor onderwijsteams zodat zij hun studenten nog beter kunnen volgen in hun ontwikkeling. Het steeds frequenter inzetten van data bij monitoring en besluitvorming stelt ook eisen aan de data en de datakwaliteit.

Een beperkte datakwaliteit kent verschillende oorzaken. Soms worden gegevens niet vastgelegd met als doel daar management- of sturingsinformatie van te maken, maar om bijvoorbeeld studenten in te schrijven of de juiste aanstelling van medewerkers te registreren. Het oorspronkelijke doel kan dan andere eisen stellen aan de data(kwaliteit). Het komt voor dat gebruikers data verschillend invoeren, invoer niet wordt gecontroleerd of dat definities niet voldoende helder zijn.

Aspecten van datakwaliteit

Datakwaliteit wordt in de literatuur bekeken langs verschillende aspecten ¹



1. Uniek

Data moet uniek zijn. Dat wil zeggen dat gegevens maar 1 keer voorkomen in de systemen. Een student komt maar 1 keer voor. Als hij zich twee keer aanmeldt, komt de betreffende student toch maar 1 keer voor in het studenten registratiesysteem.

2. Relevant

Zijn de gegevens relevant voor de gebruiker en het beoogde doel?

¹ <https://smartbridge.com/data-done-right-6-dimensions-of-data-quality/>

3. Volledig

Data moet volledig zijn. Bijvoorbeeld een voornaam en een achternaam van een student. Misschien is de roepnaam optioneel. Dan is de data zonder dat gegeven toch volledig.

4. Eenduidig

Is de data maar op 1 manier te interpreteren? Is de definitie helder en voor iedereen hetzelfde?

5. Consistent

Data moet over systemen heen en binnen systemen consistent zijn. Als een student uitgeschreven is, en er wordt nog steeds presentie geregistreerd, dan is er iets mis.

6. Accuraat

Geven de gegevens die je gebruikt de werkelijkheid goed weer? Maak je gebruik van de gemeente of de woonplaats? Wordt vsv berekend op basis van leeftijd tot 23 of tot en met 23?

7. Tijdigheid

De data moet tijdig verwerkt worden. Als er een 1 oktober telling opgesteld wordt, dan moeten de studenten gegevens voor die tijd in de systemen verwerkt zijn.

In aanvulling op eenduidig en consistent wordt validiteit soms als los onderdeel genoemd.

8. Validiteit

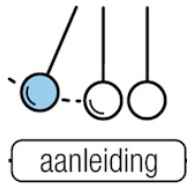
Validiteit heeft betrekking op de vorm van de ingevoerde data. Staan er – of / tussen de geboortedatum.

Van belang op verschillende momenten in het proces

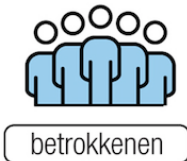
Als een teamdashboard geen gegevens bevat voor 20% van de studenten is voor iedereen de consequentie helder. Dat is vaak het moment dat er wordt gekeken naar de datakwaliteit en wat er verbeterd kan worden.

Gebrekkige datakwaliteit leidt dan tot herstel (foutoplossing) maar idealiter ook tot verbeteringen in het proces. Fouterstel is namelijk een continu proces, waarbij procesverbetering kwaliteit bij de bron probeert te verbeteren. Maar op welke manier kun je als MBO-instelling hiermee zelf aan de slag?

Hoofdstuk 2. - Stappenplan om de datakwaliteit te meten en verbeteren



Stap 1: Voorbereiding en inventarisatie – wat is het vertrekpunt



De aanleiding

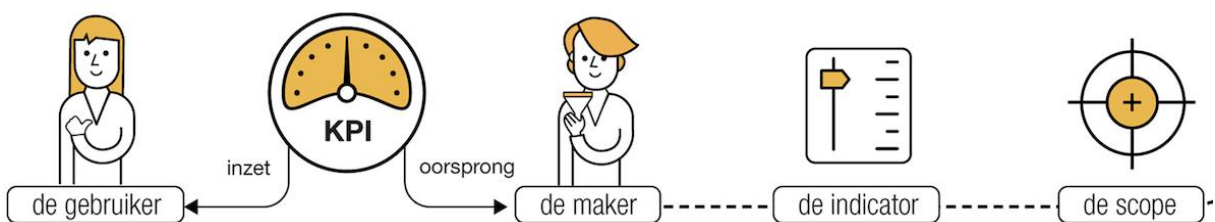
- Wat is de aanleiding/reden om dit stappenplan te doorlopen?
- Is er een bestaande visie en beleid met betrekking tot datakwaliteit?
- Hoe wordt datakwaliteit nu in kaart gebracht? Zijn daarvoor metingen of raamwerken?
- Hoe signaleren we dataproblemen, welke processen hebben we daarvoor?

Wie zijn er betrokken?

- Welke medewerkers zijn er betrokken en welke rollen/verantwoordelijkheden zie ik?
- Is er voldoende mandaat voor de betrokken medewerkers?
- Beoogde verbeteringswensen vaststellen
 - Kwaliteit op orde door fouten te herstellen
 - Probleem oplossen en processen verbeteren

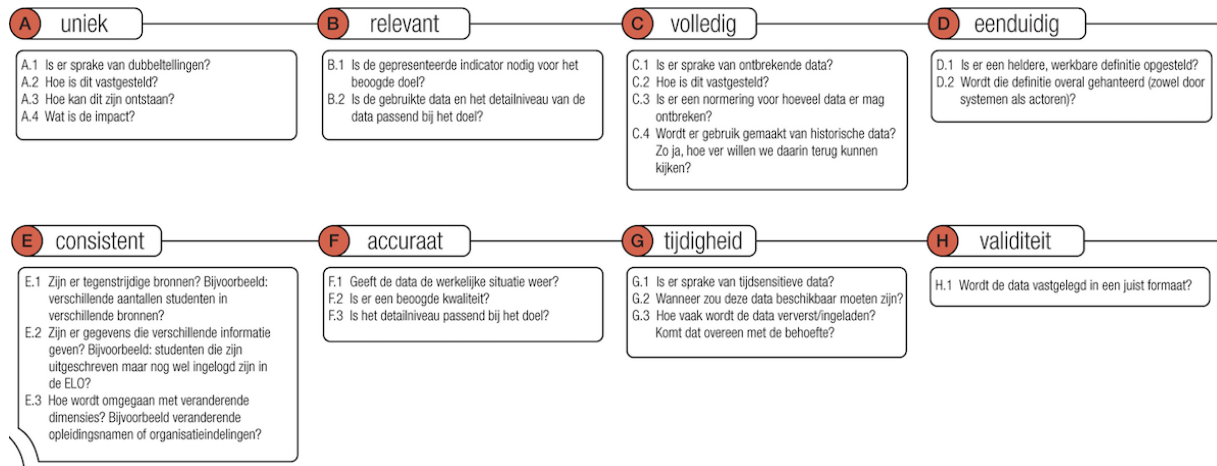
Stap 2: Situatie in beeld - aan de slag met de indicatoren

Na de voorbereidende fase kan per indicator worden bekeken wat er gedaan kan worden rondom de datakwaliteit. Dit doen we door middel van onderstaande 4 stappen in het instrument:



Stap 3: Diagnose & onderliggende oorzaak

Na het bepalen van het onderzoeksonderwerp en het identificeren van de betrokken gebruikers en makers, kan worden bekeken wat de datakwaliteit is op de 8 benoemde aspecten. Dit doen we aan de hand van een serie vragen.



Bovenstaande in ogenschouw nemend zijn er drie opvolgende vragen om dieper in te gaan op de oorzaak de mogelijke verbeteringen die we zien:

1. Op welke van de genoemde aspecten signaleren we een probleem met de datakwaliteit?
2. Hoe is dit gebrek aan kwaliteit ontstaan?
3. Als de kwaliteit leidt tot problemen: wat kunnen we doen om deze kwaliteit te verbeteren?

Stap 4: Interventie

Na het constateren en uitwerken van het probleem kan verder worden gekeken naar de oorzaak en de verbetermogelijkheden. Dit wordt gedaan aan de hand van 4 stappen met een aantal onderliggende vragen.

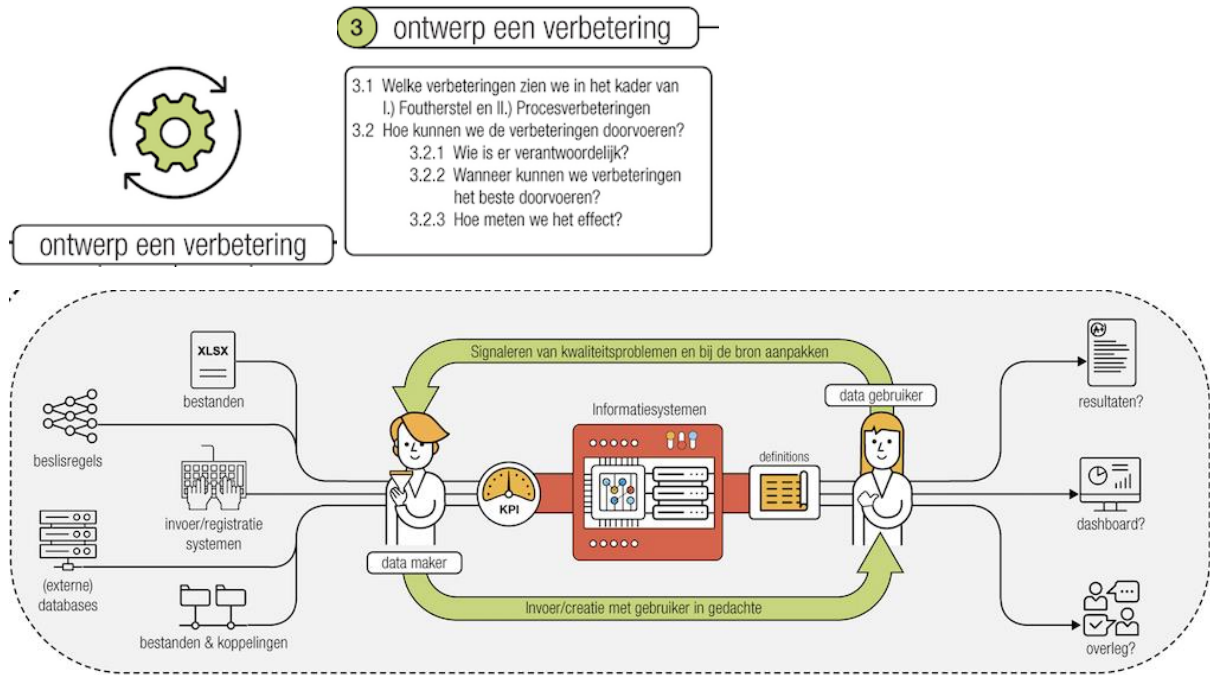
Stap 1. Onderzoek de oorzaak



Stap 2. Prioriteiten vaststellen



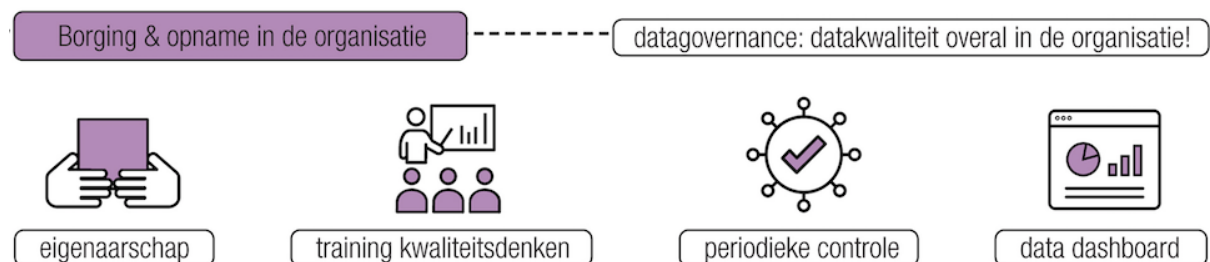
Stap 3. Ontwerp een verbetering



Stap 4. Monitor en evalueer



Stap 5: Borging & opname in de organisatie



Met datagedreven werken komt automatisch de behoefte om de data-huishouding op orde te brengen. Dat vraagt niet alleen om technische en procesoplossingen, maar om bredere aanpak: datagovernance. Data Governance is de specificatie van beslissingsrechten en een verantwoordingsraamwerk om te zorgen voor passend gedrag bij de waardering, creatie, opslag, gebruik, archivering en verwijdering van informatie. Het omvat de processen, rollen en beleidsregels,

standaarden en metrics die zorgen voor een effectief en efficiënt gebruik van informatie om een organisatie in staat te stellen haar doelen te bereiken (Gartner).

Naast het organiseren van processen, beleggen van taken en meten van kwaliteit is het van belang dat de impact van beperkte datakwaliteit overall in de organisatie helder is. In dit kader valt te denken aan:

- Zorg voor eigenaarschap bij de proceseigenaren (en niet alleen bij IT of informatievoorziening)
- Training in kwaliteitsdenken op datagebied
- Maak het makkelijk om issues te melden en zorg voor periodieke controle van je belangrijkste data
- Data dashboard voor de controle van de gegevens

Tips en overwegingen

- Fouten herstellen is vaak moeizaam, kostbaar en tijdsintensief. Er worden ingewikkelde tools & scripts voor bedacht en uitgevoerd. Maar erger nog: als je er eenmaal mee begint dan kom je er niet meer vanaf. Richt op een goede vastlegging/creatie van data vanaf de bron. Fout herstellen op historische data is dan nog steeds een optie, maar het is dan een eenmalige exercitie.
- De verbetering van de kwaliteit van data werkt het beste als het hand in hand gaat met procesverbetering. Als processen goed gedocumenteerd zijn is het makkelijker om knelpunten en overdrachtsmomenten te achterhalen. En er is een dan een duidelijke werkwijze om naar te verwijzen. Afspraken kunnen gemaakt worden op basis van een duidelijk beschreven proces.
Dit heeft de voorkeur boven het repareren van de foutieve vastlegging voor een kleine stap in de werkwijze. Deze manier kost echter meer tijd, maar levert op de lange termijn meer op.
- In het verlengde van het voorgaande punt ligt het creëren van draagvlak voor de aanpassingen. Als collega's (mede-)eigenaar zijn van de data en de producten die ermee gemaakt worden, is de kans groter dat de aanpassingen in de werkwijze ook op de lange termijn beklijven.
- Het opstellen van nieuwe KPI's is eenvoudiger als er goed gedocumenteerde processen zijn om deze op te baseren.
- Zet datakwaliteit op de agenda van MT's en directeurenoverleg.
- Richt een gebruikersgroep of klankbordgroep op
- Benoem verantwoordelijken per KPI en voor het dashboard. Inhoudelijk verantwoordelijken en technisch verantwoordelijken. Vaak komt de data in een dashboard uit diverse organisatieonderdelen.