

De groei van de inzet van GIS-gebaseerde oplossingen om organisaties te ondersteunen het hoofd te bieden aan de toenemende informatievragen, stelt hogere eisen aan de kwaliteit van het GIS en de daarin gepresenteerde data. Dit laatste artikel van een drieluik over datakwaliteit richt zich op het concretiseren van het abstracte begrip datakwaliteit. Wanneer is kwaliteit voldoende, en hoe realiseren we dit?

## Belang data(kwaliteit)

In het eerste artikel (december 2013) is ingegaan op het belang van juiste data ter ondersteuning van de bedrijfsprocessen. Daarbij is tevens geschetst hoe GIS als presentatie en integratieplatform hierin een belangrijke rol kan spelen, onder andere als katalysator van datakwaliteitverbetertrajecten.

Het tweede artikel (februari 2014) schetste de traditionele onderwaardering van data ten opzichte van de aandacht voor het inrichten van de organisatie en informatievoorziening. Tevens zoomde het artikel in op belangrijke aspecten die de datakwaliteit definiëren (volledig, correct, actueel, juiste detailniveau,...) Als inleiding op dit artikel werd tevens de vraag opgeworpen wanneer de datakwaliteit goed genoeg is. Moet we altijd 100% nastreven, of kunnen we volstaan met een "bijna" 100% kwaliteit.

## Kwaliteitsdoel

Het definiëren van het kwaliteitsdoel vraagt inzicht in de processen, de behoeften van de organisatie en een gezond verstand. Een eerste afweging betreft de omvang van de gebruikersgroep. Van datasets gebruikt door één of enkele medewerkers, weet men normaliter de betrouwbaarheid. In het benutten ervan interpreteert men de gegevens met die kennis. Voor datasets die men (organisatie)breed inzet, ligt dit anders. Gebruikers zullen ervan uitgaan dat de data klopt. De eisen gesteld aan de datakwaliteit daarvan liggen dan ook beduidend hoger. Een andere afweging is het belang van de data binnen de bedrijfsprocessen. Betreft het data die bepalend is voor het welslagen van processen, of de veiligheid ervan, of is het data die de medewerker een globaal inzicht in de situatie moeten geven. Zo zal het kwaliteitsaspect nauwkeurigheid bij de locatie van ondergrondse leidingen een andere dimensie kennen dan het ontwerpen van bijvoorbeeld de exacte positie van een kabeldoorvoerkanaal binnen een kunstwerk. Wetende dat 100% datakwaliteit niet altijd reëel (betaalbaar) is, dient men ook over de afwijkingen in de data na te denken. Wat is waardevoller voor de gebruiker / het bedrijfsproces? Indien geen 100% zekerheid is laten we de velden dan leeg? Of tonen we wat we weten evt. met indicatie van de onzekerheid?



Binnen Net4s vormt het definiëren van het kwaliteitsdoel de basis van elk datakwaliteitstraject. Via een nul-meting brengen we de status-quo in kaart. Op basis daarvan ondersteunen we de organisatie, gebruikmakend van gestandaardiseerde procedures, in het definiëren van het te realiseren kwaliteitsdoel. In die definitie differentiëren we naar de verschillende kwaliteitsaspect. Bijvoorbeeld hoeveel procent van de datavelden dient gevuld te zijn, hoeveel procent mag incorrect zijn, hoeveel procent moet aan de gedefinieerde nauwkeurigheid voldoen, hoeveel procent moet (hoe) actueel zijn etcetera.

Binnen Net4s vormt het definiëren van het kwaliteitsdoel de basis van elk datakwaliteitstraject. Via een nul-meting brengen we de status-quo in kaart. Op basis daarvan ondersteunen we de organisatie, gebruikmakend van gestandaardiseerde procedures, in het definiëren van het te realiseren kwaliteitsdoel. In die definitie differentiëren we naar de verschillende kwaliteitsaspect. Bijvoorbeeld hoeveel procent van de datavelden dient gevuld te zijn, hoeveel procent mag incorrect zijn, hoeveel procent moet aan de gedefinieerde nauwkeurigheid voldoen, hoeveel procent moet (hoe) actueel zijn etcetera.

## Kwaliteitsdoel realiseren

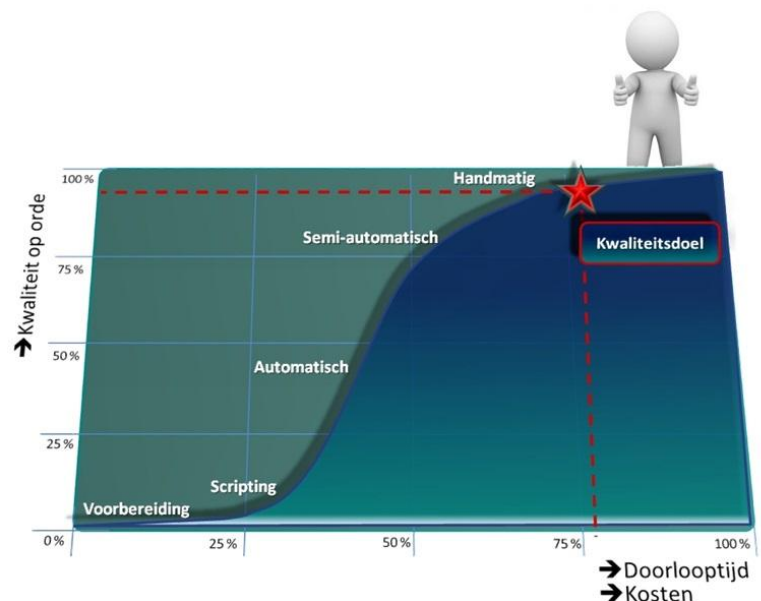
Is het kwaliteitsdoel eenmaal gedefinieerd begint het “echte” werk. Het realiseren van dat doel. Mede bepaald door de uitgangssituatie en de gedefinieerde ambitie zijn hier grofweg twee scenario’s denkbaar: Volledig handmatig het datakwaliteitsniveau realiseren, of (deels) gebruik maken van geautomatiseerde procedures. Gezien de hoeveelheid data die er vaak mee gemoeid gaat, levert het tweede scenario veelal efficiënter en sneller het gewenste resultaat.

### Voorbereiding

Start daarvan is de *voorbereiding* fase. Daarin bepaald men op basis van het inzicht in de status-quo welke gegevensbestanden relevant zijn, eventueel aangevuld met extern beschikbare gegevens. Veel datakwaliteitverbetertrajecten ontstaan als verdergaande informatiebehoefte het integraal benutten van verschillende informatiesystemen vraagt. Het confronteren van bestanden met complementaire gegevens over dezelfde objecten, is vaak letterlijk confronterend voor de organisatie en betrokkenen. Het resulteert in inzicht in discrepanties en hiaten in de informatievoorziening. Denk bijvoorbeeld aan een situatie waarin een technisch bestand objecten met hun kenmerken bevat, een GIS omgeving de locatie en samenhang van deze objecten en een beheerpakket de onderhouds- en logistieke gegevens. In een recente cases zagen we in zo’n situatie dat een match van bestanden tot een lage overeenstemming (orde grootte 60%) leidde. Gegevens als objectidentificatie ontbraken, locatie was niet bekend, met bijbehorende vraag bestaat het object (niet) meer of mist het in de GIS omgeving etc. In de voorbereidingfase definieert men op basis van proceskennis, een analytisch inzicht en het inzicht in de bestanden en discrepanties de logica om de datakwaliteit te verbeteren. Daarbij zijn eventueel tevens externe bestanden als beschikbaar fotomateriaal (luchtfoto’s, cyclorama’s ed.) in te zetten.

### Automatisch traject

De logica vertaalt zich in *scripts* die na een test en pilotfase de bulk van de datakwaliteit verbetering *automatisch* kunnen doen verlopen. Ervaring leert dat daarmee zo’n 60-70% van het juiste kwaliteitsniveau is te realiseren. Uiteraard is het denkbaar naast de bulk ook de complexe en uitzonderingssituatie via scripts te verwerken. Daarbij is het evenwel zaak alert te blijven op de verhouding tussen inspanning in het automatiseringsproces en hetzelfde handmatig uitvoeren.



### Uitzonderingen Semi-automatisch

Een *semi-automatische* benadering is hier zinvol. Door de discrepanties automatisch inzichtelijk te maken kan men gericht en daarmee efficiënt de correcties handmatig door voeren. Ten slotte blijft er altijd een deel van de verbeteringen zo complex dat deze alleen handmatig en met inhoudelijke kennis van de organisatie, haar assets en processen is te realiseren.

Zoals de figuur illustreert zijn de laatste loodjes het zwaarst, zowel in benodigde tijd als in kosten. Een onderstreping van het belang van een goede definitie van het kwaliteitsdoel!

## Binnen – buiten

De geschetste benadering resulteert via een “binnen-tenzij” benadering in een efficiënt traject. Maar de praktijk wijst uit dat we vrijwel altijd ook naar de fysieke situatie buiten moeten kijken. Dat vormt onderdeel van de semi-automatische fase. Medewerkers van de organisatie, of bij eenduidige vraagstelling evt. derden, schouwen de situatie in het veld. Net4s benut voor de objectenregistratie een toegespitste app die in het veld de te onderzoeken discrepanties en gevraagde actie inzichtelijk maakt, bijvoorbeeld object wel/niet aangetroffen, juiste locatie aangeven of het nummer noteren. GIS is daarbij ook ingezet om overzichtelijk te presenteren waar discrepanties uit het (semi) automatische proces naar voren kwamen, en ook om (live) de voortgang van het schouwproces te monitoren. Zie de voorbeelden van de app en monitor omgeving. In een recente casus is het op basis van de geschetste benadering en de Net4s standaarden gelukt het kwaliteitsniveau van zo'n 60% gematchte data naar 97% te brengen. Dit binnen een doorlooptijd van enkele weken.

## Proces kwaliteit

Een laatst onderwerp betreft de kwaliteit van het dataverbetertraject zelf. Hoe kunnen we garanderen dat na de verbeteringen de datakwaliteit correct is? Uiteraard kan men een 100% controle uitvoeren, maar dat is gezien de hoeveelheid gegevens veelal niet haalbaar. De procesindustrie levert hier goede benaderingen aan. Één daarvan is de Acceptance Quality Limit (AQL). AQL structureert en kwantificeert het beslissingsproces voor het goed- of afkeuren van een verwerking. Het legt op basis van statistische methoden objectief vast hoeveel en welke fouten men mag terugvinden in een steekproef. Op basis van representatieve steekproeven is zo te bepalen of het gewenste kwaliteitsdoel is bereikt. Daarbij is het resultaat te differentiëren naar de genoemde kwaliteitsaspecten als compleetheid, correctheid, actualiteit etc. De afwijkingen weegt men veelal naar bedrijfskritisch, belangrijk, ondergeschikt om een compleet beeld te genereren. Indien een steekproef het minimale acceptabele niveau niet bereikt betekent dat, dat de hele “verwerkingsbatch” afgekeurd is en opnieuw moet plaatsvinden. In een kwaliteitstraject helpt deze benadering in de aanvangsfase ook om het proces steeds verder te verbeteren. Eerste steekproeven (in pilotfasen) zullen groot zijn om een goed inzicht in het proces te krijgen. Naarmate dat inzicht ontstaat en het proces is geoptimaliseerd zijn de steekproeven meer toegespitst.

Het benodigde datakwaliteitsniveau is zo via een gestructureerde benadering te realiseren en ondersteunt organisaties daarmee in het volledig benutten van de potentie van een GIS omgeving als belangrijk onderdeel van de totale informatievoorziening.

 [j.roodzand@net4s.nl](mailto:j.roodzand@net4s.nl)

 [www.net4s.nl](http://www.net4s.nl)

 [j.roodzand@ruimteschepper.nl](mailto:j.roodzand@ruimteschepper.nl)

 [www.ruimteschepper.nl](http://www.ruimteschepper.nl)

 @JanRoodzand

 06 200 133 65

Jan Roodzand heeft een ruime ervaring als ondernemer, adviseur en programmamanager in het implementeren van GIS-omgevingen.

Kenmerkend daarbij is de integratie ervan met andere informatiesystemen en de aandacht voor de datakwaliteit.

Hij is mede initiatiefnemer en partner in Net4s en Ruimteschepper

