

Data Kwaliteit & Testen

Technieken en Tools voor het beoordelen en verbeteren van datakwaliteit



- Korte introductie
- Data Kwaliteit en Testen
- Data Analyse een Model
- Aansluiting Data Kwaliteit en Testen
- Tools en Creativiteit

- ✍ Managing Consultant bij Capgemini
- ✍ 18 jaar ervaring: van operator tot thought leader
- ✍ Oprichter Conversion Squad, gericht op:
 - ✍ data kwaliteit
 - ✍ data migraties
- ✍ Thought leader vakgroep Migration Services
- ✍ Trekker COP Data Migration

- ✍ Meten van datakwaliteit
- ✍ Doel hiervan:
 - ✍ Verzamelen informatie over data kwaliteit (“meten is weten”)
 - ✍ Verspreiden verkregen informatie voor diverse doeleinden
 - ✍ Conversie
 - ✍ Testen
 - ✍ Systeemontwikkeling / inrichting
 - ✍ Maar ook: fraude preventie, operationeel, kostenbesparing
 - ✍ Problemen in kaart brengen (“vinger op de zere plek”)
 - ✍ Oplossen (“oplossen voordat het een probleem wordt”)

Misconception #1: You Can Fix Data

Fixing implies that there was something wrong with the original data, and you can fix it once and be done with it. In reality, the problem may have been not with the data itself, but rather in the way it was used. When you manage data you manage data quality. It's an ongoing process. Data cleansing is not the answer to data quality issues. Yes, data cleansing does address some important data quality problems and offers a solid business value ROI, but it is only *one* element of the data quality puzzle. Too often the business purchases a data cleansing tool and thinks the problem is solved. In other cases, because the cost of data cleansing tools is high, a business may decide that it is too expensive for them to deal with the problem.

Misconception #2: Data Quality is an IT Problem

Data quality is a *company* problem that costs a business in many ways. Although IT can help address the problem of data quality, the business has to own the data and the business processes that create or use it. The business has to define the metrics for data quality - its completeness, consistency, relevancy and timeliness. The business has to determine the threshold between data quality and ROI. IT can enable the processes and manage data through technology, but the business has to define it. For an enterprise-wide data quality effort to be initiated and successful on an ongoing basis, it needs to be truly a joint business and IT effort.

Column published in DMReview.com
June 3, 2004 By [Rick Sherman](#)

Data Kwaliteit En Testen

Focus bij testen:

- Werkende processen
- Werkende functionaliteit
- Business wensen / eisen
- Operationeel systeem
- Toetsen kwaliteit en gemaakte afspraken

"An error detected during the project testing phase can cost up to one hundred times more to correct than the same error found during the design phase."
- The Quality Assurance Institute

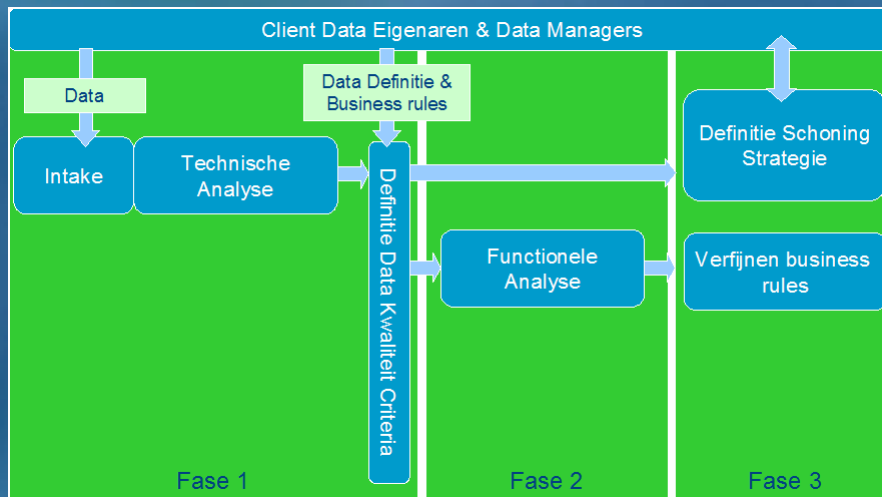
Basis voor werking van applicaties = 'data'

- Sturing
- Informatie voor organisatie
- Werking van de organisatie

"Data drives the application and organization. Errors in data will prevent an organization from growing to a new level"

Data uit 'oude systemen' zegt veel over het 'nieuwe systeem'

Data Kwaliteit Meten Hoe Doe Je Dat
"Model"



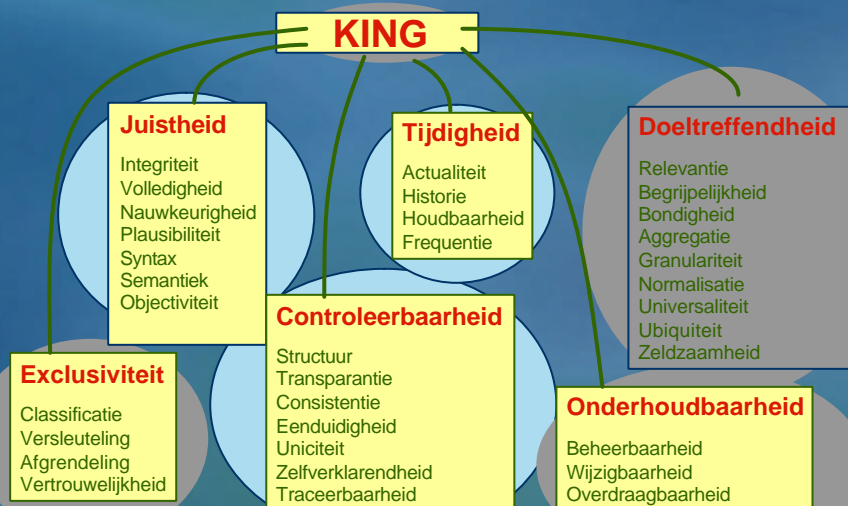
- ✎ Extra informatie ter onderbouwing van testen
 - ✎ Min / Max / Domein / Interne Lengte / Gebruik
 - ✎ Plausibiliteit vraagstukken ((jaar)domeinen)
 - ✎ Relaties tussen gegevens (referentiele integriteit)
- ✎ Betere testscripts / completer testset
- ✎ Uitzonderingen in gegevens

Snel inzicht in bestaande situatie



- ✎ Extra informatie ter onderbouwing van testen
 - ✎ Complexe vraagstukken vanuit business
 - ✎ Wat kunnen ze nu met de applicatie?
 - ✎ Waar zitten de hiaten?
- ✎ Betere testscripts / completer testset
- ✎ Meer gericht op samenhang / business
- ✎ Uitzonderingen en speciale situaties

Zoeken naar een spel in een hooiberg



Logisch

- Multiple sites en applicaties
- Multiple definities voor hetzelfde gegeven
- Eigenaarschap

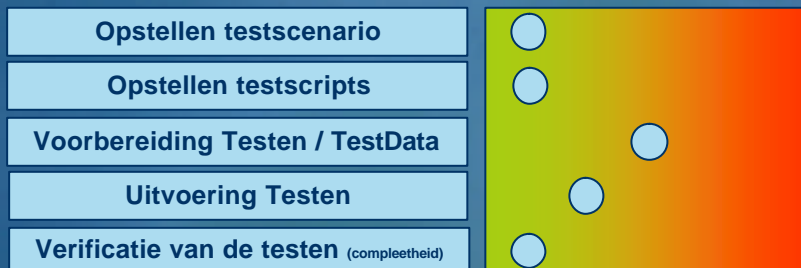
Physical

- Data wordt gekopieerd tussen systemen (distributie)
- Data inconsistenties (tijd)



“Corporate datamodelling”, één definitie binnen één organisatie

Nut van Data Kwaliteit voor Testers



- ✎ Testdata selecteren op basis van oude systeem
 - ✎ Consistente set data op basis van datamodel toetsen
- ✎ Opstellen van testscripts op basis van brondata
 - ✎ Minimum - / maximum waarden / interne lengte
- ✎ Wijzigingen in omgeving naar aanleiding van testen
 - ✎ Welke waarden komen nu voor t.o.v. eerdere situatie
- ✎ Vroegtijdig analyseren is mogelijk (bron bestaat al!)
- ✎ AS-IS test situaties geven herkenning en rust

Tooling en Creativiteit





Fouten zeggen wat over het gebruik

Naam	Parent	Child	Size	Created	Modified	Access
0			0	17-12-2005 14:0		Read only
300			0	18-12-2005 13:0		Check
0			0	17-12-2005 14:0		Read only
0			0	17-12-2005 14:0		Read only
0			0	17-12-2005 14:0		Read only



Maar het hoeft niet altijd fout te zijn

Naam	Parent	Child	Size	Created	Modified	Access
0			0	18-12-2005 13:0		Read only
0			0	17-12-2005 14:0		Read only
0			0	17-12-2005 14:0		Read only
0			0	17-12-2005 14:0		Read only





 Tools voor Analyse

-  Pervasive Data Profiler
-  Trillium Software Discovery
-  Compuware File-AID
-  Diverse ETL tools zoals SAS, Crystal Reports

 Tools voor Schoning en Kwaliteit behoud

-  Human Inference
-  Trillium

 Gebruik MS Excel / MS Access Let wel op:

-  Consequente vraagstelling
-  Herhaalbaarheid
-  Toegankelijkheid tot resultaten
-  Stel eerst de goede vragen voordat je start

