

Effecten op operationeel niveau

Op gebruikerstevredenheid en imago	Verlaagd, klanten en gebruikers hebben weinig begrip voor fouten.
Op de kosten	Hoge kosten: 10% tot 50% van de omzet (T.C. Redman)
Op het moreel van de werknemers	Sterk verlaagd, foute data geven veel frustratie en inefficiëntie
Op de, financiële, rapportages (ook tactische en strategische impact)	Grote problemen met invoeren IAS, IFRS, Basel II, Sarbanes Oxley (SOX)

Effecten op operationeel niveau

Op gebruikerstevredenheid en imago	Verlaagd, klanten en gebruikers hebben weinig
Op de kosten	Hoge kosten: 10% tot 50% van de omzet (T.C. Redman)
Op het moreel van de werknemers	Sterk verlaagd, foute data geven veel frustratie en inefficiëntie
Op de, financiële, rapportages (ook tactische en strategische impact)	Grote problemen met invoeren IAS, IFRS, Basel II, Sarbanes Oxley (SOX)

Gebruikerstevredenheid
"Uw computer heeft mij een kind toebedeeld. Maar ik heb helemaal geen kind. En al helemaal niet van uw computer"
 bron: citaten uit brieven aan verzekeraars

Effecten op tactisch niveau

Op de besluitvaardigheid	Verlaagd. Management neemt slechte beslissingen, laat of helemaal niet
Op data warehousing en business intelligence	Vertraagde implementatie door slechte data in bronsystemen
Op systeemontwikkeling, onderhoud en migratie	Slechte datakwaliteit is een groot struikelblok
Op het vertrouwen tussen organisaties, partners	Sterk verlaagd
Op de productiviteit van werknemers	Verlaagd: werken onder hun niveau aan triviale fouten

Effecten op strategisch niveau

Op de strategiebepaling	Kost meer tijd, is moeilijker
Op de implementatie van strategische wijzigingen	Bemoelijkd door verwarring op tactisch en operationeel niveau
Op het vermogen om meer waarde te creëren uit bedrijfsinformatie	Bemoelijkd o.a. het beleggen van informatie-eigenaarschap
Op de synergie en samenwerking in de organisatie	Gehinderd. Afdelingen communiceren slecht door niet eenduidige data.
Op management focus	Verminderd, wordt afgeleid van de wezenlijke issues.

Effecten op strategisch niveau

Op de strategiebepaling	Kost meer tijd, is moeilijker
Op de implementatie van strategische wijzigingen op operationeel niveau	Bemoelijkd door verwarring op tactisch en operationeel niveau
Op het vermogen om meer waarde te creëren uit bedrijfsinformatie	Bemoelijkd o.a. het beleggen van informatie-eigenaarschap
Op de synergie en samenwerking in de organisatie	Gehinderd. Afdelingen communiceren slecht door niet eenduidige data.
Op management focus	Verminderd, wordt afgeleid van de wezenlijke issues.

Strategische achterstand
"Ik heb zoveel formulieren moeten invullen, dat ik veel liever had gehad dat mijn geliefde man helemaal niet was gestorven"
 bron: citaten uit brieven aan verzekeraars

Kortom

1. Meten en verbeteren van datakwaliteit is belangrijk, en wordt komende jaren nog belangrijker.
2. Deze discipline zal een volwassen plaats gaan krijgen in de ICT.
3. Dat vergt goede Mensen, goede Methodes en goede Middelen

Waarom deze avond

- Wij, als testers, bewaken belangrijke projecten
 - Datawarehouses, BI, CRM, ERP implementaties, migratie, ...
- Ook wij (vooral wij?!) bewaken datakwaliteit
- Maar we doen dat niet professioneel
 - ad-hoc, zonder algemeen geaccepteerd methodisch kader
 - datakwaliteit is voor informatie-analisten en database experts?
 - we laten testen toch ook niet (alleen) aan systeemanalisten, ontwerpers en programmeurs over
- We hebben wellicht behoefte aan:
 - een overall visie en kader
 - een referentiemodel voor datakwaliteit, met goede definities
 - methoden, technieken en tools

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

13

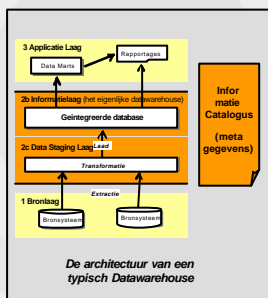
Inhoud

- Introductie: het belang van datakwaliteit
- Aandacht voor datakwaliteit in het testproces
 - Wat doen we al en wat weten we al?
- Datakwaliteit: literatuur, goeroes en organisaties
- Modellen voor het definiëren van datakwaliteit
 - het KING model voor Kwaliteit van Informatie en Gegevens
 - Het PSP/IQ model van het M.I.T.
- Methoden voor het beoordelen van datakwaliteit
 - IQA, TDQM
 - Data Profiling
- Conclusies

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

14

Een typisch datawarehouse



- Drie lagenmodel:
 - Onderkant:** Bronsystemen
 - Binnenkant:**
 - Data Staging (bewerking)
 - Geïntegreerde database
 - Bovenkant:** Applicaties en gebruikers
- Het product:
 - Gegevens
 - Metagegevens
- Informatiecatalogus met metadata essentieel

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

15

Voorbeeld acceptatiecriteria DWH

Acceptatiecriteria Datawarehouse: de hoofdcategorieën	
A	Autorisaties en security
B	Beheerbaarheid van de DWH componenten (technische en functionele voorzieningen)
C	Controles en overige waarborgen voor datakwaliteit aan bronzijde
E	Extractieproces: volledigheid, consistentie, juistheid en afgesproken formaat van de aanlevering
G	Gegevens in laag 3: volledigheid, juistheid en consistentie, aansluiting bij fysiek datamodel
H	Historie
I	Integratie en consistentie met reeds in DWH aanwezige gegevens
J	Justigheid (inhoudelijke) en plausibiliteit van de gegevens, aansluiting bij aanpalende systemen
M	Metagegevens en Informatie Catalogus: volledigheid, juistheid, bruikbaarheid
O	Beheerorganisatie rondom DWH en de aanleveringen
P	Performance, betrouwbaarheid, robuustheid
S	Technische InfraStructuur en interfaces
T	Transformatie- en laadprocessen

Alles draait om data en metadata

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

16

Testboeken over datawarehousing

- 2 technieken in Tmap Test Topics, hoofdstuk datawarehousing
 - Controletoeleningen (hash totals, vierkantstelling, audit trail)
 - Steekproeven
 - Verder diverse tips voor (creëren van) testdata
- Verder niet veel bijzonders



EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

17

Wat zegt de testliteratuur verder?

- 2 technieken in o.a. "Kwaliteit door Acceptatietesten":
 - Levenscyclustest: aan de hand van de CRUD matrix
 - Legitimiteitstest: controle op relaties en constraints)
- Wat artikelen in o.a. STQE magazine (stickyminds.com)
- KING



Een magere oogst !

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

18

Inhoud

- Introductie: het belang van datakwaliteit
- Aandacht voor datakwaliteit in het testproces
 - ☞ Wat doen we al en wat weten we al?
- Datakwaliteit: literatuur, goeroes en organisaties
- Modellen voor het definiëren van datakwaliteit
 - ☞ het KING model voor Kwaliteit van Informatie en Gegevens
 - ☞ Het PSP/IQ model van het M.I.T.
- Methoden voor het beoordelen van datakwaliteit
 - ☞ IQA, TDQM
 - ☞ Data Profiling
- Conclusies

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

19

De literatuur en de goeroes



EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

20

Organisaties en conferenties



- DWQ project: ESPRIT programma van de Europese unie.
- Nederland: DDMA richt werkgroep datakwaliteit op www.ddma.nl

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

21

Inhoud

- Introductie: het belang van datakwaliteit
- Aandacht voor datakwaliteit in het testproces
 - ☞ Wat doen we al en wat weten we al?
- Datakwaliteit: literatuur, goeroes en organisaties
- Modellen voor het definiëren van datakwaliteit
 - ☞ het KING model voor Kwaliteit van Informatie en Gegevens
 - ☞ Het PSP/IQ model van het M.I.T.
- Methoden voor het beoordelen van datakwaliteit
 - ☞ IQA, TDQM
 - ☞ Data Profiling
- Conclusies

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

22

Wat is datakwaliteit?

- Allereerst: Data = informatie Data ? informatie
- "Data are of high quality if they are fit for their intended use in operations, decision making and planning" (Joseph Juran, kwaliteitsgoeroe)
 - ☞ Fit for use: gebruikersperspectief, situationeel bepaald
 - ☞ Operationeel, tactisch, strategisch
- Dit heeft vele dimensies, net als softwarekwaliteit
- Er zijn modellen beschikbaar. Enkele bruikbare worden hierna behandeld

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

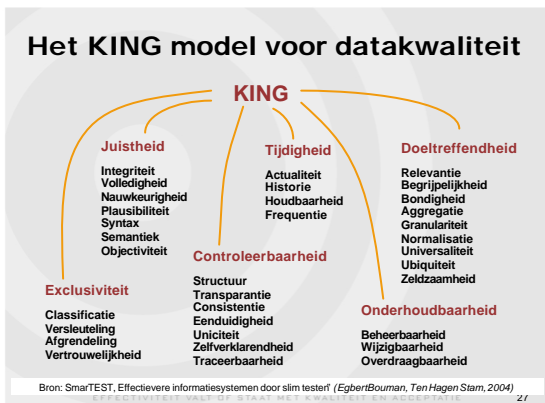
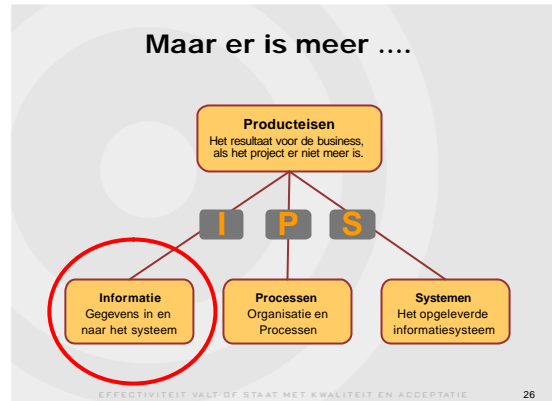
23

Twee modellen voor datakwaliteit

- Het **KING model** van Maintain
 - ☞ Praktisch, compleet, nederlandstalig, onderdeel SmarTEST
- Het **PSP/IQ model** van het M.I.T.
 - ☞ Gedegen, wetenschappelijk onderbouwd

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

24



Met definities

KENMERK	Definitie	Synoniemen, tegengestelden
1 Juistheid	De mate waarin de informatie qua presentatie en inhoud correct is vastgelegd	Betrouwbaarheid, Waarheid
Integriteit	De mate waarin relaties tussen de data-elementen technisch en syntactisch juist en volledig zijn. Zie ook "structuur".	
Volledigheid	De mate waarin alle gewenste en relevante gegevens beschikbaar zijn in de informatieverzameling	
Nauwkeurigheid	De mate van detail waarmee de informatiekenmerken (attributen) worden beschreven	Accuraatheid. Zie ook granulariteit.
Plausibiliteit	De mate waarin de informatie geloofwaardig is en overeenkomt met de werkelijkheid buiten het informatiesysteem	Waarheid, Juistheid

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 28

Met definities

KENMERK	Definitie	Synoniemen, tegengestelden
1 Juistheid	De mate waarin de informatie qua presentatie en inhoud correct is vastgelegd	Betrouwbaarheid, Waarheid
Integriteit		
Volledigheid		
Nauwkeurigheid		
Plausibiliteit		

Plausibiliteit

"Ik reed eerst met mijn auto tegen de vangrail, sloeg toen over de kop en knalde tenslotte tegen een boom. Toen verloor ik de macht over het stuur"

bron: citaten uit brieven aan verzekeraars met de werkelijkheid buiten het informatiesysteem

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 29

Quality is KING

- KING = **K**waliteit van **I**nformatie en **G**egevens
- Dekt alle aspecten van data / informatiekwaliiteit
- Met uitgebreide definitielijst, nederlandstalig
- Gebruikers-, beheerders- en leveranciersperspectief
- Gepubliceerd in boek: "SmarTEST, Effectievere informatiesystemen door slim testen" (Egbert Bouman, Ten Hagen Stam, 2004)

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 30

Een alternatief: het PSP/IQ model

- In 1999-2001 heeft het M.I.T. alle op dat moment beschikbare opsommingen van datakwaliteitsattributen op een rijtje gezet
- Van zowel "Academics" als "Practitioners"
- Dit alles kritisch bekeken en vergeleken
- Vervolgens is hieruit het PSP/IQ model gemaakt
- Dit model kiest het *gebruikersperspectief*

Input uit theorie en praktijk

		Table: The Academics' View of Information Quality			
		Accuracy IQ	Completeness IQ	Representational IQ	Accessibility IQ
Wang and Strong [36]	Accuracy Believability Rapidity Objectivity	Value-related Relevance Completeness Timeliness Appropriate Access	Understandability Interoperability Consistent Representation Consistent Representation	Accuracy Ease of Operation Security	
Zasad [40]	Accuracy Fidelity	Quantity Reliability Timeliness	Appropriateness Readability Security		

		Table: The Practitioners' View of Information Quality			
		Accuracy IQ	Completeness IQ	Representational IQ	Accessibility IQ
ORCA [38]	Accuracy Completeness Consistency Validity	Timeliness	Completeness		
MITRE [39]	Accuracy in [192]	Accuracy in [192]	Accuracy in [192]	Accuracy in [192]	Accuracy in [192]
IRI [39]	Accuracy	Timeliness			Reliability (of delivery)
Yates [33]	Accuracy Consistency Reliability	Completeness Timeliness			Security Privacy

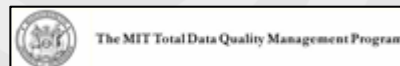
Het PSP/IQ model

	Confirms to Specifications	Meets or Exceeds Customer Expectations
Product	Sound Information	Useful Information
Quality	IQ Dimensions: <ul style="list-style-type: none"> Free-of-Error Consistent Representation Completeness Consistent Representation 	IQ Dimensions: <ul style="list-style-type: none"> Appropriate Access Relevance Understandability Interoperability Objectivity
Service	Dependable Information	Usable Information
Quality	IQ Dimensions: <ul style="list-style-type: none"> Timeliness Security 	IQ Dimensions: <ul style="list-style-type: none"> Believability Accessibility Ease of Operation Rapidity

Bron: "AIMQ: A methodology for Information Quality Assessment, Lee et al., 2001"

Wetenschappelijk onderbouwd

- Uit de vakgroep van Richard Wang e.a., Massachusetts Institute of Technology
- Uit dezelfde stal komen:
 - TDQM: een "Total Data Quality Management" raamwerk
 - AIMQ: een methode voor assessment van datakwaliteit
 - IQA: een instrument (vragenlijst) hierbij
- Site: <http://web.mit.edu/tdqm> (aanrader!)



Waarschuwing

Voorkom analysis paralysis

Gebruik modellen als KING en PSP/IQ als checklist, neem niet alle mogelijke kwaliteitsaspecten mee, maar selecteer de belangrijkste

[Redman], field tip 17.2:

"Data Quality involves meeting the most important needs of the most important customers. Be prepared to make the tough choices".



Inhoud

- Introductie: het belang van datakwaliteit
- Aandacht voor datakwaliteit in het testproces
 - Wat doen we al en wat weten we al?
- Datakwaliteit: literatuur, goeroes en organisaties
- Modellen voor het definiëren van datakwaliteit
 - het KING model voor Kwaliteit van Informatie en Gegevens
 - Het PSP/IQ model van het M.I.T.
- Methoden voor het beoordelen van datakwaliteit
 - IQA, TDQM
 - Data Profiling
- Conclusies

Het IQA instrument

- IQA: Information Quality Assessment
 - ⚡ Ontwikkeld door het M.I.T.
 - ⚡ Statistisch getoetste en geoptimaliseerde vragenlijst
- Model-gebaseerd
 - ⚡ Beschikbaar voor PSP/IQ
 - ⚡ KING kan ook
- Met vragen of stellingen, zoals:
 - ⚡ "Deze informatie wordt consistent gepresenteerd"
 - ⚡ "Deze informatie is relevant voor mijn werk"
- Een stakeholder-perceptie onderzoek



Total Data Quality Management

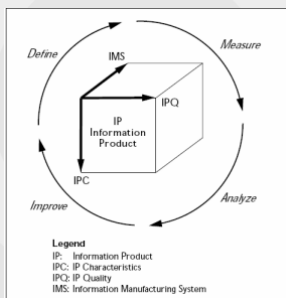
- Een product van het M.I.T.
- Parallel getrokken met de klassieke industrie
- Volgt het Total Quality Management stramien (Deming en Juran)



	Fysieke productie	Informatie productie
Input	Ruwe materialen	Ruwe data
Proces	Productie lijn	Informatie systeem
Output	Fysieke producten	Informatie producten

Leestip:
"A product Perspective on Total Data Quality Management"
(<http://web.mit.edu/dqm>)

De TDQM cyclus



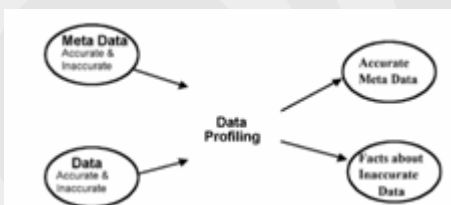
- Gebaseerd op de verbetercyclus van Deming (Plan, Do Check, Act)
- Als denkkader zeer bruikbaar
- Voor testers is de rechterhelft interessant

Een commerciële toepassing



Een "implementatie" van de TDQM cyclus van de firma Firstlogic.

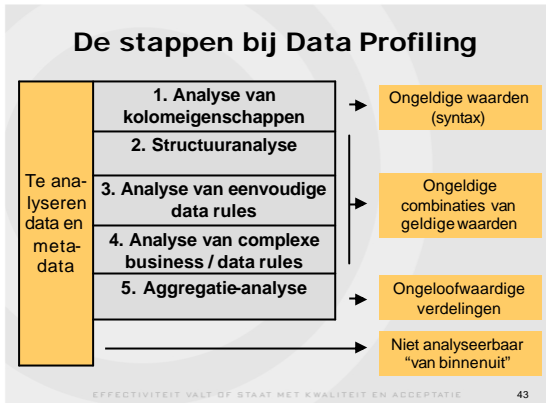
Data Profiling: hoe werkt het?



- Een "van binnen uit" aanpak
- De data en metadata zelf zijn je enige input
- Dat maakt de methode heel rechttoe rechtaan


Data Profiling: Discovery en Test

- Discovery
 - ⚡ Onttrafelen van de gegevensstructuur
 - ⚡ Verbeteren en aanvullen van metadata
- Test
 - ⚡ Voldoet de data aan de verwachting?
 - ⚡ met technieken als assertion testing, visuele inspectie, metadata verificatie, etc.



Samenvatting Data Profiling

- Systematische aanpak
 - Vanuit de data zelf: inside out
- Inventarisatie, Meting en Analyse
 - Heeft wel iets van Exploratief Testen
 - Heuristics,
- Jack Olson schreef er een boek over.
 - Vol met concrete, toepasbare technieken
 - Teveel om hier op te noemen
- Mijns inziens erg geschikt voor testers



EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 44

Tot slot: de nieuwe millenniumcrisis?



Kwaliteit van bedrijfsdata onvoldoende?

Grote kans op ruzie met deze mannen!

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 45

Inhoud

- Introductie: het belang van datakwaliteit
- Aandacht voor datakwaliteit in het testproces
 - Wat doen we al en wat weten we al?
- Datakwaliteit: literatuur, goeroes en organisaties
- Modellen voor het definiëren van datakwaliteit
 - het KING model voor Kwaliteit van Informatie en Gegevens
 - Het PSP/IQ model van het M.I.T.
- Methoden voor het beoordelen van datakwaliteit
 - IQA, TDQM
 - Data Profiling
- Conclusies


EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 46

Conclusies

- Grip op datakwaliteit is de ICT uitdaging anno 2005
- Als testers van informatiesystemen laten we het tot op heden behoorlijk afweten
- De handvatten zijn beschikbaar, zoals: KING, TDQM en Data Profiling

**De hamvraag:
(Hoe) pakken we dit op?**

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE 47



Boekreviews



EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

English



Improving Data Warehouse and Business Information Quality

Larry P. English, 1999



Commentaar: Een klassieker, zeker onder datawarehousebouwers. Veel praktische tips en cijfers, maar ook snorkende taal en theoretische managementmodellen. Voor de liefhebber van visionairs.

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

49

Redman



Data Quality: The Field Guide

Thomas C. Redman, 2001



Commentaar: Helder boek, nadruk op totaalaanpak en voorkomen in plaats van herstellen. Veel aandacht voor organisatie/cultuuraspecten. Klant/leverancier model. Minder praktisch bruikbare zaken voor testers dan Olson.

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

50

Olson



Data Quality: The Accuracy Dimension

Jack E. Olson, 2003

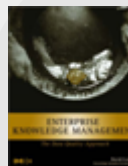


Commentaar: Top voor testers! Een uiterst helder geschreven boek, vooral interessant om de goede beschrijving van de Data Profiling methode. Visie, maar ook veel techniek(en). Veel breder dan de titel doet vermoeden. Beetje DBMS kennis wenselijk.

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

51

Loshin



Enterprise Knowledge Management, the Data Quality Approach

David Loshin, 2001.



Commentaar: Een aantal goede hoofdstukken, en toespitsing op kennismanagement is interessant. Kwaliteitsaspecten met uitvoerige definities. Nadeeltje: onhandige pil, veel tekst, weinig figuren en tabellen, daardoor slecht toegankelijk.

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

52

Wang



Enterprise Knowledge Management, the Data Quality Approach

Richard Y. Wang, et al., 2001.



Commentaar: Het eerste hoofdstuk over TDQM, PSP/IQ model e.a. is goed leesbaar en interessant, maar dat is ook als los artikel op het internet te vinden. De rest is uiterst gedegen, maar zeer academisch. Voor wie de verdieping zoekt. Erg duur (US\$ 117,-).

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

53

Nederlandstalige Boeken



Datawarehouseing, een inleiding

Karien Verhagen - 2005 - 194 blz

Een inleiding inderdaad, leerboek.



Datawarehouse: bron van kennis, voor marketing en sales

Frank Buytendijk, Cok de Groot - 2000

De sales insteek, voor gebruikers.

EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

54

Extra slides

De impact van slechte data

- Datakwaliteit kritische succesfactor voor:
 - ✗ ERP, DWH, CRM
 - ✗ Corporate Performance Management
 - ✗ Enzovoort (waar eigenlijk niet?)
- Zichtbaar
 - ✗ Iedereen kent de voorbeelden
 - ✗ Vervuilde klantenbestanden
- Onzichtbaar
 - ✗ Veel doorwerking, gevolgschade, toekomstige schade
 - ✗ Een sluipend element: "Bad data are like viruses" (T. Redman)
- Op alle niveaus
 - ✗ Strategisch, Tactisch, Operationeel

Belang anno 2005 extra groot

Internationale wet- en regelgeving voor financiële rapportage en risicomanagement:

IAS, IFRS: uniforme financiële rapportage

BASEL II: risicomanagement voor banken

Sarbanes-Oxley: strenge rapportageregels voor bedrijven met Amerikaanse beursnotering

Dit zijn momenteel belangrijke drivers voor verbetering van datakwaliteit.

Testen van datawarehouses

- Kwaliteit van data in de bronsystemen (achterkant)
 - ✗ [Verhagen]: "de gegevensaudit"
- Testen van ETL proces en data (binnenkant)
 - ✗ ETL: Extractie, Transformatie en Laad
 - ✗ De geïntegreerde database: het eigenlijke DWH
- Testen van datamarts, aggregaties, queries, rapporten (voorkant)
- Testen en toetsen van de IC met metadata
- FURPS testen: functionaliteit, gebruikersvriendelijkheid, betrouwbaarheid, performance, veiligheid en onderhoudbaarheid

Begin bij de bron: de gegevensaudit

- Onjuiste definities
- Onjuiste, ontbrekende en vervuilde veldwaarden
 - ✗ Als de leverancier er geen voordeel van heeft gaat het mis
- Homoniemen en synoniemen
 - ✗ Zelfde namen voor iets anders en omgekeerd
- Afwijkende waarden in verwant systeem
 - ✗ Geslacht = M/V of 0/1 (=M/V) of MN/O of
- Interpretatiefouten
 - ✗ Strandbezoekers mei 50, juni 60, juli 70. Dus oktober 100.
- Combinatiefouten
 - ✗ Het controllersdilemma

Leestip: "Datawarehouse, een inleiding" Hoofdstuk 4

Resultaat IQA assessment

Het verschil in perceptie wordt zichtbaar

De vier stappen van TDQM

- **Define**
 - ☒ De karakteristieken: het logisch gegevensmodel (ERD)
 - ☒ De requirements: formuleer de kwaliteitseisen kwalitatief
- **Measure**
 - ☒ Formuleer de kwaliteitseisen kwantitatief (metrics)
 - ☒ Voer metingen en assessments uit
- **Analyze**
 - ☒ Statistische analyse
 - ☒ Vind de "root causes"
- **Improve**
 - ☒ Implementeer verbetermaatregelen
 - ☒ Evalueer en herhaal



EFFECTIVITEIT VALT OF STAAT MET KWALITEIT EN ACCEPTATIE

61