

DATA INTEGRITEIT BINNEN KLM ENGINEERING & MAINTENANCE

AANBEVELINGEN TER BEHOUD VAN DATA INTEGRITEIT OP HET MENSELIJK VLAK



E.J.P. MARKERINK
STUDENTNUMMER: 1015508
BACHELORTHESIS
VAKCODE: 194100010

UNIVERSITEIT TWENTE
1E BEGELEIDER: DR. IR. A.A.M. SPIL
2E BEGELEIDER: DRS. A.M.G.M. HOOGEBOOM
DATUM: 30 SEPTEMBER 2013

Voorwoord

Voor u ligt het rapport geschreven in opdracht van de afdeling Change Development, Component Services, KLM Engineering & Maintenance, ten behoeve van het afronden van mijn bachelor opleiding Bedrijfskunde aan de Universiteit Twente.

Deze opdracht bestond uit het doen van onderzoek naar de data integriteit binnen Component Services, KLM Engineering & Maintenance. Vragen die hierbij centraal stonden waren: Wat zijn de hoofdoorzaken van variërende data kwaliteit welke resulteert in een verslechtering van de data integriteit? Is er een verband tussen variërende data, technische kennis en gedrag binnen de organisatie KLM E&M? Wat is de invloed van variërende data kwaliteit op de operationele en mogelijk financiële performance? Ligt het grootste probleem aan de mensen of aan de systemen welke binnen de organisatie? Wat zijn uiteindelijk mogelijke oplossingen om de data integriteit te verbeteren en eventuele operationele en financiële performance te verbeteren?

Om deze vragen te beantwoorden is op voorhand literatuuronderzoek uitgevoerd om te kijken naar bestaande theorieën & onderzoeken welke de theoretische kant van het probleem beschrijven. Hierbij is een selectie gemaakt van twee hoofdartikelen waarin de hoofdoorzaken van data integriteit worden omschreven op zowel het menselijk als systeem technische vlak. Uit dit hoofdartikel een framework opgesteld welke de leidraad vormt binnen dit onderzoek. Dit framework bestaat uit vier kwadranten te weten: Mens, Systeem, Migratie & Opschoning en Overig. Omdat er op dit moment weinig literatuur te vinden is over het menselijke kwadrant en de organisatie geïnteresseerd is in deze specifieke kant, is in dit rapport voornamelijk gericht op dit kwadrant.

Tevens is er praktijk gericht onderzoek gedaan binnen de organisatie om zodoende de theorie tegen de praktijk te kunnen weerspiegelen. Het praktijk onderzoek is gedaan aan de hand van een etnografische aanpak, waarbij ik zelf intern vijf maanden binnen de organisatie KLM E&M heb meegedraaid. Hierbij is gekeken naar het totale procesverloop en is er met een grote hoeveelheid mensen binnen diverse afdelingen gesproken. Deze gesprekken, de uitkomsten en de discussie welke door andere afdelingen is gedaan omtrent deze uitkomsten zijn de basis waarop de praktijk informatie in dit rapport is gebaseerd. De discussie is met opzet toegepast om zodoende persoonlijke kleuring van de uitkomsten zoveel mogelijk uit te sluiten. Hiermee is getracht de afhankelijkheid van de vergaarde informatie te kunnen waarborgen. (Bryman, 2012).

In dit rapport vindt u de probleemstelling, theoretisch kader, methodologie, praktijkonderzoek, integratie en analyse van theorie en praktijk en de uiteindelijk aanbevelingen.

Naast het belang van KLM E&M is er ook nog een belang ten aanzien van de Universiteit Twente. Dit onderzoek is aller eerst een afronding van mijn Bachelor opleiding Bedrijfskunde waarbij aangetoond wordt dat ik in staat ben om binnen een opdracht, het onderwerp uit te splitsen, hier gedegen onderzoek op toe te passen en uiteindelijke conclusies en aanbevelingen te doen. Daarnaast is dit onderzoek van wetenschappelijk belang aangezien er over dit specifieke vraagstuk nog weinig wetenschappelijke onderzoeken zijn verricht. De meeste wetenschappelijke artikelen omtrent data integriteit, zoals (Ghaeb & Chebil, 2009), (Nouha, Leneutre, & Roudier, 2011), (Ghaeb, Smadi, & Chebil, 2011) en (Sivathanu, Wright, & Zadok, 2005) zijn gericht op de aanpak van data integriteit op het systeem niveau. Beter gezegd zijn deze artikelen gericht op het opslaan van data en hoe daarbij data integriteit kan worden behouden aan de hand van verschillende technieken. Naast deze artikelen is er ook veel wetenschappelijk onderzoek gedaan naar het belang van data integriteit op Big Data en hoe hier mee om te gaan, zoals de onderzoeken van (Klein, Tran-Gia, & Hartmann, 2013) en (Tien, 2013). Vrijwel al deze artikelen zijn op de systeem technische kant gericht van het data integriteitsvraagstuk. Onderzoeken welke een menselijke factor benadrukken zijn voornamelijk gericht op slechts één element uit het opgestelde framework, maar niet alle elementen. Dit rapport is dan ook specifiek gericht op de menselijke elementen welke in beeld komen ten aanzien van data integriteit. Doordat in

dit rapport een directe toepassing plaatsvindt van artikelen welke minder wetenschappelijk onderbouwd zijn, maar voornamelijk op praktijkervaring van de auteurs berust, kan er met een wetenschappelijke insteek en onderbouwing van aanvullende wetenschappelijke literatuur wel onderzoek gedaan worden naar dit vraagstuk. Dit rapport draagt dan ook zeker bij aan het vraagstuk data integriteit door de focus meer te leggen op niet alleen het systeem maar ook de mens.

Dankwoord

Bij het uitvoeren van dit onderzoek ben ik vanuit de Universiteit Twente begeleid door dhr. Dr. Ir. A.A.M. Spil en mevr. drs. A.M.G.M. Hoogeboom, waarvoor mijn dank voor hun hulp, begrip, doorzettende en meedenkende vermogen en het zeer snelle willen en kunnen schakelen.

Vanuit KLM Engineering & Maintenance ben ik begeleid door dhr. R. Jagbandhan en dhr. J. Hamelink, waarvoor mijn dank voor hun vertrouwen, openheid, sturing, mogelijkheden tot persoonlijke ontwikkeling binnen het bedrijf en directe beantwoording van al mijn vragen.

Tevens wil ik alle mensen bij KLM Engineering & Maintenance bedanken welke mij hebben geholpen bij het schrijven en meedenken van dit rapport en hun bijdrage aan een geweldige afstudeer periode.

Graag bedank ik ook mijn zus voor haar hulp en sturing tijdens mijn afstudeerperiode.

Als laatste wil ik mijn ouders bedanken voor het überhaupt mogelijk maken van mijn studie, hun jarenlange onvermoeide inzet voor mij en motiverende factor tijdens mijn bachelor opleiding.

Egbert Markerink

Management Samenvatting

Dit onderzoek is uitgevoerd naar aanleiding van de vraag wat de mogelijke oorzaken, gevolgen en oplossingen zijn van data integriteit binnen KLM Engineering & Maintenance. Nu de koers binnen de organisatie wordt verlegt richting het sturen op data, is het van belang dat de integriteit hiervan op niveau wordt gebracht en gewaarborgd. Op dit moment weet men binnen de organisatie dat hier problemen mee zijn, echter is het probleem nooit systematisch onderzocht. Voor dit betreffende vraagstuk is daarom de volgende doelstelling opgesteld voor het onderzoek:

Het aandragen van oplossingen en aanbevelingen ter verbetering van de huidige data integriteit binnen KLM Engineering & Maintenance, door middel van het onderzoeken van menselijke oorzaken welke bijdragen aan het huidige data integriteitsprobleem.

Naast deze doelstelling, welke tevens als hoofdvraag beschouwt kan worden, zijn drie deelvragen opgesteld om tot deze doelstelling te komen. Deze deelvragen zijn stapsgewijs behandeld om uiteindelijk de betreffende oplossingen en aanbevelingen zoals geformuleerd in de doelstelling te kunnen aandragen. Deze deelvragen zijn als volgt geformuleerd:

1. *Wat zijn de hoofdoorzaken van verslechterde data integriteit volgens de huidige literatuur?*
2. *Welke oorzaken gevonden bij deelvraag 1 zijn van toepassing op KLM E&M in de praktijk en welke mogelijke verbanden en conclusies kunnen hieruit worden getrokken?*
3. *Welke wetenschappelijke oplossingen kunnen er worden aangedragen voor de verbetering van de oorzaken welke bij de tweede deelvraag zijn uitgelicht?*

Deze drie deelvragen zijn elk apart behandeld waarbij als volgt te werk is gegaan. De eerste deelvraag wordt beantwoord door middel van onderzoek op het gebied van literatuur. Tussen een groot aantal artikelen is uiteindelijk gekozen om twee artikelen te hanteren als leidraad. Deze artikelen beschrijven de hoofdoorzaken van data integriteit binnen een organisatie, waarbij zowel op het menselijk vlak als het systeem technische vlak wordt gekeken. Deze twee hoofdartikelen zijn met elkaar geïntegreerd voor wat betreft de overeenkomsten en verschillen. Aan de hand hiervan is er een vier kwadranten framework opgesteld waarbij de volgende factoren van belang zijn: Mens, Systeem, Migratie & Opschoning en Overig. Dit framework wordt als leidraad gehanteerd binnen dit rapport. Tevens zijn wordt dit framework op de diverse elementen aangevuld met andere literatuurstukken om de uitspraken welke gedaan worden te onderbouwen.

In de tweede deelvraag zijn de oorzaken, welke beschreven zijn bij de eerste deelvraag, tegen de praktijk afgezet. Dit praktijk onderzoek is gedaan aan de hand van een intern onderzoek binnen KLM E&M. Hierbij is een etnografische aanpak gehanteerd vanwege de omvang van de opdracht en de omgeving waarin deze zich positioneert. Er is vijf maanden lang intern onderzoek verricht waarbij het proces in kaart is gebracht, er gesprekken met verschillende afdelingen en een groot aantal van hun medewerkers zijn gevoerd, en de gegeven antwoorden van verschillende afdelingen bij elkaar ter discussie zijn gesteld. Vervolgens is de opgedane informatie verwerkt en na enige periode geverifieerd bij diverse bronnen en afdelingen om te achterhalen of de uitkomsten van het eerdere onderzoek gegrond waren. Door deze aanpak kunnen uitkomsten van onderzoeksresultaten beter gewaarborgd worden zonder dat deze teveel persoonlijk gekleurd zijn. Aan de hand van de verkregen informatie tijdens deze vijf maanden is de beantwoording van de tweede deelvraag gedaan. Deze beantwoording illustreert hoe de praktijk zich verhoudt ten opzichte van theorie.

Als laatste is er aan de hand van de eerste en tweede deelvraag gekeken naar doelgerichte aanbevelingen welke kunnen leiden tot algemene verbeteringen ten aanzien van data integriteit op het menselijke kwadrant. Hierbij is tevens naar voren gekomen dat mentaliteit, communicatie en kennis de koppeling zijn tussen de menselijke elementen welke beschreven worden in het opgestelde framework.

Tot slot is er gekeken naar vervolgonderzoek waarbij aanbevelingen zijn gedaan op het dieper onderzoeken van data integriteit. Hierbij wordt gericht op het verrichten van onderzoek ten aanzien van het opgestelde framework, en de andere drie kwadranten.

Inhoudsopgave

Voorwoord	1
Dankwoord.....	2
Management Samenvatting.....	3
Hoofdstuk 1: Inleiding	7
1.1 De organisatie KLM Engineering & Maintenance	7
1.2 Introductie van de opdracht	7
1.3 Probleemstelling.....	7
1.4 Doelstelling.....	8
Hoofdstuk 2: Theoretisch kader	9
2.1 Het begrip data integriteit.....	12
Hoofdstuk 3: Methodologie.....	13
3.1 Onderzoek strategie	13
3.2 Onderzoek ontwerp.....	13
3.3 Uitvoering praktijk onderzoek	13
3.4 Enquête ontwerp.....	14
3.5 Aanvulling	14
Hoofdstuk 4: Deelvraag 1	15
4.1 Mens.....	15
4.2 Systeem	16
4.3 Migratie & Opschoning.....	17
4.4 Overig	18
4.5 Conclusie	18
Hoofdstuk 5: Deelvraag 2.....	19
Hoofdstuk 6: Deelvraag 3	19
6.1 Manual data entry	19
6.2 New data uses & information obfuscation.....	20
6.3 Loss of Expertise	20

6.4	Communication	20
6.5	System.....	20
Hoofstuk 7	Conclusies	20
Hoofstuk 8	Discussie.....	21
Hoofstuk 9	Limitatie.....	22
Referenties.....		23
Appendix A	figuren	25

Hoofdstuk 1: Inleiding

In dit eerste hoofdstuk zal een korte introductie plaatsvinden van de organisatie KLM Engineering & Maintenance waarna vervolgens de opdracht zal worden geïntroduceerd waarop dit onderzoek is gebaseerd. Vervolgens zal uit deze opdrachtomschrijving de probleemstelling worden beschreven met voortvloeiend daaruit de doelstelling van dit rapport.

1.1 De organisatie KLM Engineering & Maintenance

KLM Engineering & Maintenance is de onderhoudsdivisie van KLM. Samen met Air France Industries (AFI) is KLM E&M een toonaangevende speler in de wereldwijde vliegtuigonderhoudsindustrie.

KLM E&M en AFI onderhouden naast de vloot van KLM en Air France, de vliegtuigen, motoren en componenten van ongeveer 150 klanten wereldwijd. Bij KLM E&M werken ruim 4000 medewerkers, bij AFI-E&M ongeveer 15.000.

Component Services (CS) is een bedrijfsonderdeel van KLM E&M. Bij CS dient gedacht te worden aan een zeer breed scala van componenten. Dit gaat van flaps, wielen, slides (glijbanen) tot cockpit instrumenten, stoelen in de cabine en zelfs koffieapparaten. Deze hoeveelheid aan verschillende componenten brengt een zeer variërend proces met zich mee wat dagelijks verschilt in manier van zowel werken als processturing. Daarnaast zorgen de verschillende contracten met klanten ervoor dat er nog meer diversiteit tijdens het proces naar voren komt. Het is dan ook een uitdaging om deze, soms zeer gecompliceerde operatie, zo te sturen dat alle componenten binnen hun gegeven tijd worden afgerond en terug naar de klant kunnen worden gestuurd.

1.2 Introductie van de opdracht

Vanuit de afdeling Change Development, KLM Engineering & Maintenance (KLM E&M), is de vraag gesteld hoe data integriteit binnen de organisatie bewaakt kan worden. In de gesprekken welke gevoerd zijn met diverse mensen van de afdeling Change Development is naar voren gekomen dat, met het oog op de toekomst en het project 'Securing Our Future' (SOF), het belang van integere data steeds groter wordt. Dit project SOF houdt in dat de KLM in de toekomst door middel van een breed pakket aan maatregelen haar toekomst tracht veilig te stellen. SOF omvat de volgende kerntaken: het bereiken van financiële doelen en het zorgen voor een duurzaam blijvend herstel van de concurrentiepositie & financiële kracht. Om deze doelen te kunnen bereiken dient er, ook binnen KLM E&M, in de toekomst meer sturing te worden verricht op data waardoor op dit moment onderzoek naar dit probleem van belang is. Inherent aan het project SOF is het behouden en uitbreiden van het klantportfolio.

1.3 Probleemstelling

Omdat dit vraagstuk zeer omvangrijk is, is getracht door middel van interne gesprekken met diverse mensen binnen KLM E&M deze omvang te verkleinen tot een bepaalde basis waar onderzoek naar verricht kan worden. Tijdens deze gesprekken zijn alle facetten van het proces aan bod gekomen waarbij gekeken is naar de verschillende afdelingen, werkzaamheden, manier van werken en systemen waarmee gewerkt dient te worden. Omdat er veel informatie valt te verkrijgen over de systemen en weinig over de menselijke interventie, is er gekozen voor een onderzoek wat zich richt op deze menselijke factor. Systemen zullen deels tijdens dit onderzoek worden meegenomen maar niet het hoofdonderzoek omvatten. Hieronder zijn een aantal systemen uitgewerkt welke tijdens dit rapport van belang zijn ten aanzien van de werkzaamheden:

Crococ Een systeem waarin alle componenten worden bijgehouden voor wat betreft part-serienummer, uren & cycles, manufacturing date en ruilartikelcode. Een ruilartikelcode is een code waaronder diverse serienummers van componenten zijn

geschaard welke met elkaar uitwisselbaar zijn. Hierdoor kan er snel gekeken worden of componenten met elkaar uitwisselbaar zijn en voor vervanging niet direct hetzelfde serienummer benodigd is. De componenten welke verwerkt worden in dit systeem zijn vrijwel alle mogelijke componenten welke in een vliegtuig te vinden zijn. Dit gaat van wielen, stoelen, vlieginstrumenten en koffiezetapparaten tot motoren waarvan elk onderdeel in de motor in dit systeem te vinden is. Dit systeem is ontwikkeld door en voor KLM waarbij vrijwel alle informatie uit dit systeem gehaald wordt. Crocos dient daarom als een bronsysteem beschouwd te worden.

- Sap** Wereldwijd bekend ERP systeem waarin eveneens informatie van componenten verwerkt zit. Tevens wordt onder andere de financiële afhandeling in dit systeem gedaan. Steeds meer processen tracht men binnen KLM E&M in dit systeem te verwerken, zoals het project Trace2Sap aantoont (uitfasering van Trace en implementatie in Sap).
- Trace** Informatievoorziening over de route welke een component aflegt binnen de verschillende afdelingen. Tevens wordt hierin bijgehouden wat de status van een component is voor wat betreft Servicable, Non-Servicable, locatie en bewerkingsregels welke toegepast dienen te worden.
- Aerexchange** Wereldwijd online platform wat door luchtvaartmaatschappijen onderling wordt gebruikt voor het uitwisselen van informatie over beschikbare componenten, verstuurde componenten en reparatie benodigdheden.

De manier waarop dit interne onderzoek is gedaan zal verder worden toegelicht in paragraaf 1.5. Omdat mogelijk incorrecte data kan leiden tot operationele missturing, verkeerde werkzaamheden welke op het component worden uitgevoerd, etc.) is het noodzaak met oog op de toekomst om data integer te houden.

1.4 Doelstelling

Naar aanleiding van het bovenstaande, dient er onderzoek verricht te worden naar het bewaken en continue verbeteren van de data integriteit. Hierbij staan de volgende zaken centraal: de hoofdoorzaken volgens de literatuur welke bijdragen aan data integriteitsproblemen met focus op de menselijke interventie, hoe deze hoofdoorzaken mogelijk terug te vinden zijn in de praktijk en de uiteindelijk conclusies en aanbevelingen welke gedaan kunnen worden aan de hand van het onderzoek met de daar bijbehorende analyses. De doelstelling binnen dit rapport is daarom als volgt geformuleerd:

Het aandragen van aanbevelingen ten behoudt en continue verbetering van de data integriteit binnen organisaties, door middel van het onderzoeken van menselijke oorzaken welke ten grondslag liggen aan data integriteitsproblemen.

Om deze doelstelling te behalen zijn er drie deelvragen opgesteld welke bijdragen in de onderbouwing en analyse voor de beantwoording hiervan. De eerste deelvraag richt zich op de literatuur en de hoofdoorzaken welke hierin bekend zijn op zowel het menselijk als technische vlak:

1. Wat zijn de hoofdoorzaken van data integriteitsproblemen volgens de huidige literatuur?

De tweede deelvraag omvat het praktijk onderzoek wat gedaan is binnen KLM E&M. In deze deelvraag worden de resultaten van de eerste deelvraag tegenover dit praktijk onderzoek gezet. Hoe verhouden de resultaten van de eerste deelvraag zich ten opzichte van het praktijk onderzoek en welke verbanden

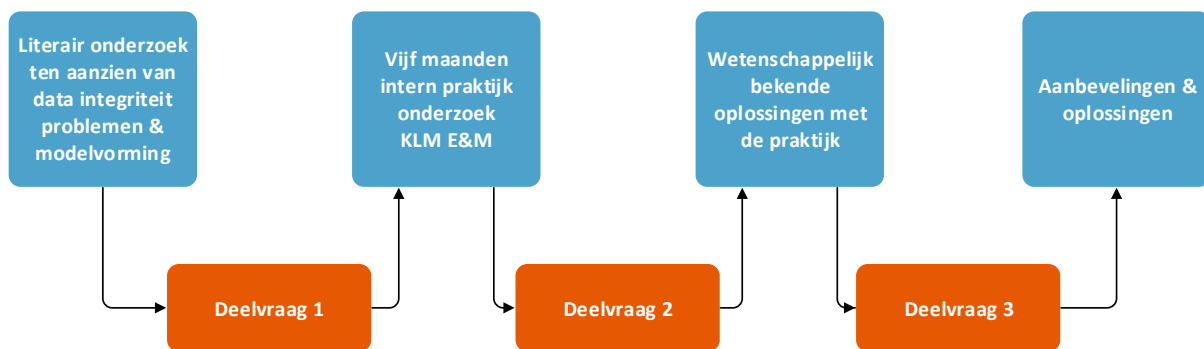
en opvallende uitkomsten kunnen hieruit worden geconcludeerd? Om dit samen te vatten is de tweede deelvraag als volgt geformuleerd:

2. *Welke oorzaken gevonden bij deelvraag 1 zijn van toepassing op KLM E&M in de praktijk en welke mogelijke verbanden en conclusies kunnen hieruit worden getrokken?*

De conclusies en mogelijke verbanden welke bij de tweede deelvraag gevonden worden zullen in de derde deelvraag worden meegenomen. In deze deelvraag zal een combinatie worden gemaakt tussen wetenschappelijke verbeteringen en de mogelijke implementatie van deze verbeteringen in de praktijk. De derde deelvraag welke gebruikt zal worden is daarom als volgt:

3. *Welke wetenschappelijke oplossingen kunnen er worden aangedragen voor het verbeteren van de oorzaken welke bij de tweede deelvraag zijn uitgelicht?*

In deze derde deelvraag zal gebruik worden gemaakt van wetenschappelijke onderbouwing welke niet alleen op menselijke interventie op systemen is gericht, maar ook op de manier hoe mensen beïnvloed kunnen worden om veranderingen door te voeren, zoals beschreven door Cameron & Green. Na de beantwoording van deze laatste deelvraag kunnen specifieke procesgerichte aanbevelingen en oplossingen worden aangedragen welke dienen te leiden tot het bewakren en continue verbeteren van data integriteit. De opbouw van dit rapport is in figuur 1.1 terug te vinden waarbij wordt verduidelijkt hoe dit rapport is opgesteld.



Figuur 1.1 Opbouw van het verslag ten aanzien van onderzoek en deelvragen.

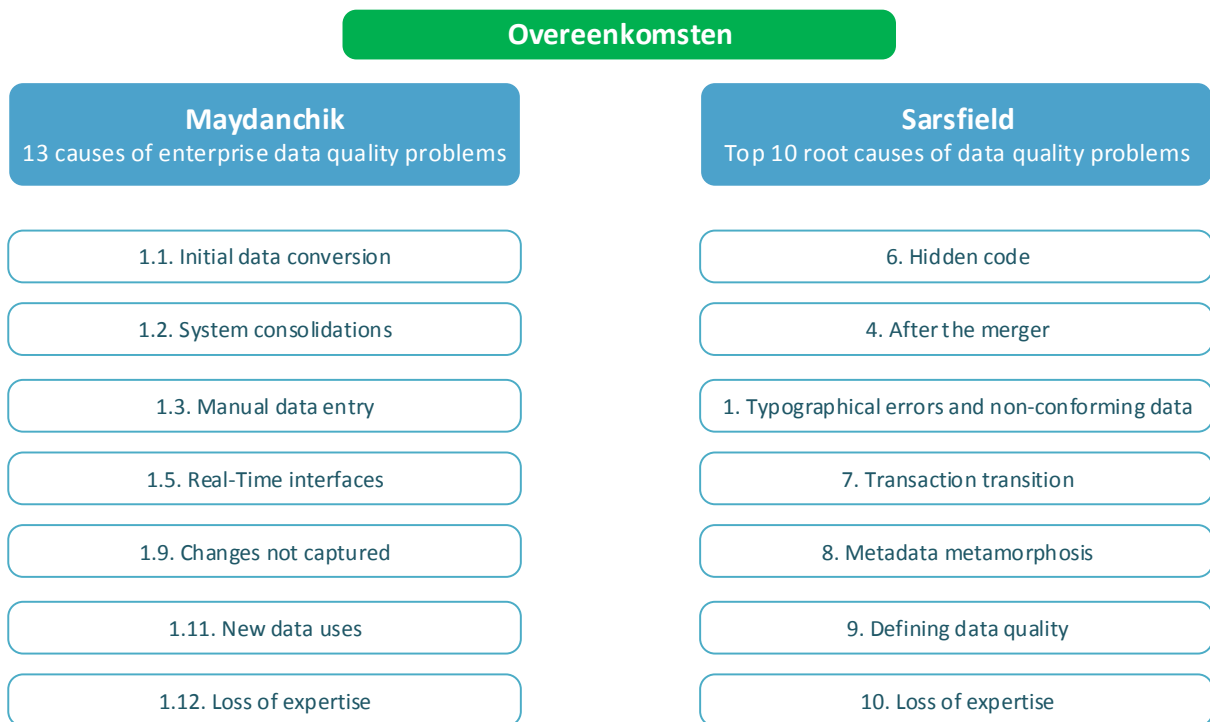
Hoofdstuk 2: Theoretisch kader

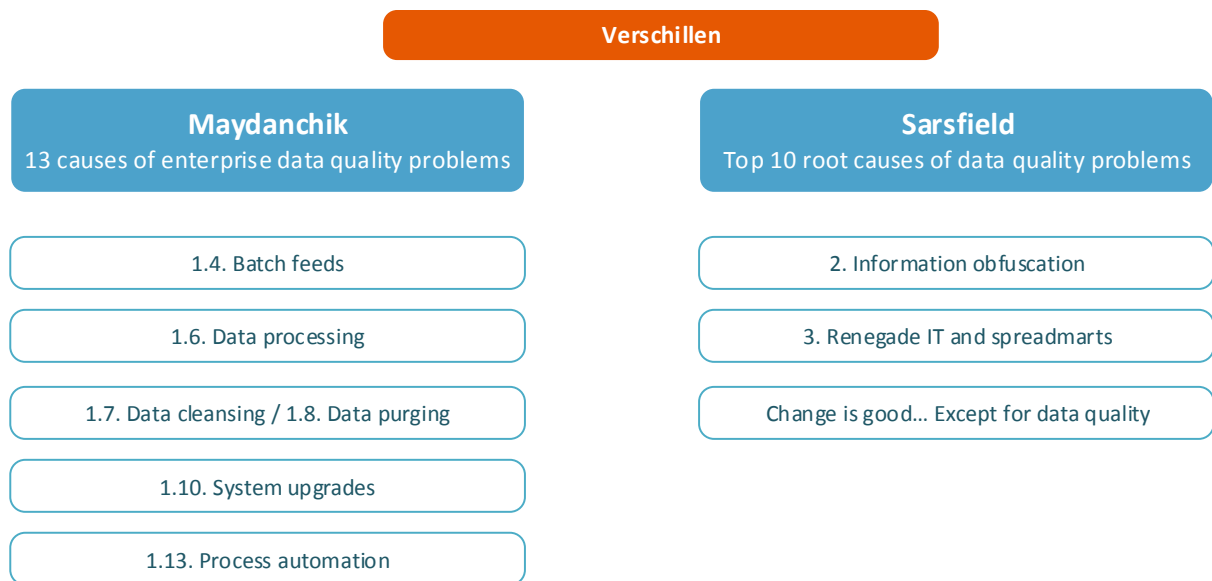
Voor het onderzoek wat binnen dit rapport is gedaan en ter ondersteuning van de eerste deelvraag, is er gekeken naar bestaande theorieën en artikelen welke geschreven zijn ten aanzien van dit onderwerp. Hierbij is gebruik gemaakt van diverse bronsystemen waaronder de UTwente bibliotheek, Google, Google Scholar, boeken etc. Zoektermen voor het zoeken naar de juiste artikelen, in zowel Engels als Nederlandse taal waren onder andere:

- Data accuracy / human interference
- Human error data integrity / systems / organization
- Data quality problems / human interference / systems / information systems
- Data validation / entry error / human

Ondanks het feit dat in veel zoektermen het woord 'human' of 'mens' is opgenomen, zijn de meeste wetenschappelijke artikelen, en daaraan gerelateerde artikelen, inhoudelijk veel gericht op data integriteit in combinatie met de systeem technische factor. Voorbeelden hiervan zijn de artikelen van (Ghaeb & Chebil, 2009), (Nouha, Leneutre, & Roudier, 2011), (Ghaeb, Smadi, & Chebil, 2011) en

(Sivathanu, Wright, & Zadok, 2005) welke zich voornamelijk richten op het opslaan van data en hoe daarbij data integriteit kan worden behouden aan de hand van verschillende technieken. Tevens is er veel onderzoek gedaan naar Big Data en hoe organisaties hiermee dienen om te gaan, zoals de artikelen van (Klein, Tran-Gia, & Hartmann, 2013) en (Tien, 2013). Wetenschappelijke artikelen omtrent de menselijke problemen met data integriteit zijn voornamelijk gericht op de mogelijke manieren om de invoer van data door mensen te verifiëren en controleren. Er is op dit moment geen duidelijk wetenschappelijk model wat concreet op diverse punten deze menselijke factoren toetst en dieper uitlicht wat deze voor invloed hebben op data integriteit. Binnen dit rapport is daarom een combinatie gemaakt van twee gevonden stukken geschreven door (Maydanchik, 2007) en (Sarsfield, 2011) welke beide zeer geruime tijd op data integriteitsprojecten binnen diverse organisaties hebben gewerkt, of met dit dagelijkse vraagstuk omgaan. Deze stukken, een boek en een paper, bevatten een combinatie van systeem & menselijke factoren welke volgens de auteurs de hoofdoorzaken van de problemen met data integriteit beschrijven. Het stuk van Maydanchik: “13 causes of enterprise data quality problems” is naast het stuk van Sarsfield: “Top 10 root causes of data quality problems” gelegd waarna deze aan elkaar getoetst zijn. Hierbij is gekeken naar de overeenkomsten en verschillen van de beschreven hoofdoorzaken. In de figuur 2.1 is een overzicht gegeven van de uitkomst van deze toetsing.





Figuur 2.1 Overeenkomsten & verschillen (Maydanchik, 2007), (Sarsfield, 2011).

Opvallend is dat Maydanchik de uitspraak doet dat *“de grootste oorzaak van data integriteit terug te traceren valt tot de mens”*. Doordat de meeste wetenschappelijke artikelen niet deze specifieke menselijke interventie onderzoeken maar hoe systemen het best data onderling kunnen uitwisselen, is onderzoek hiernaar in combinatie met praktijk onderzoek iets wat van waarde kan zijn voor de huidige literatuur. Er is gevraagd vanuit de organisatie om voornamelijk te onderzoeken wat het effect van de mens is op data integriteit. Dit is de reden dat dit onderzoek hier het meest op zal worden gericht. Omdat dit onderzoek zich richt op de menselijke factor, maar systemen ook een belangrijk rol kunnen spelen is dit waar mogelijk meegenomen. Uit de stukken van de twee auteurs is een nieuw model ontwikkeld wat de leidraad zal zijn voor dit rapport, te weten: het *geïntegreerd model van 4 kwadranten*. Hierbij is een matrixstructuur van toepassing waarbij de interne factoren en daar bijbehorende 4 kwadranten uiteindelijk invloed kunnen hebben op data integriteit. Voor het onderzoek zijn, waar mogelijk, de 4 kwadranten met hun interne factoren in dit rapport ondersteund door aanvullende literatuur. De andere twee kwadranten zullen in mindere mate aanwezig zijn en van belang voor dit onderzoek. Uitspraken en analyses welke binnen dit rapport worden gedaan dienen hierdoor een betere ondersteuning te krijgen. Het betreffende geïntegreerd model van 4 kwadranten is terug te vinden in figuur 2.2.

Omdat het rapport zich richt op slechts twee kwadranten waarvan het menselijk kwadrant de leidraad vormt is voor deze twee kwadranten een hypothese opgesteld voor analyse en het bijdragen van informatie aan de opgestelde deelvragen. Deze hypothesen en verwachtingen zijn als volgt:

Mens

“De grootste problemen veroorzaakt op het gebied van data integriteit ligt aan kennis, kunde en mentaliteit van de mens”. Hierbij wordt verwacht dat door een combinatie van systeemkennis en mentaliteit dit de grondslag is voor de meeste data integriteitsproblemen.

Systeem

“Doordat systemen langere tijd mee gaan en het veel geld kost om deze continue aan te passen, zullen systemen geen volledige aansluiting hebben huidige processen binnen organisaties”. De verwachting is dat de manier van systeemwerking naar huidige processtandaarden een significante invloed kan hebben op data integriteit.

De uitwerking van deze hypothesen zal gebeuren aan de hand van de analyse welke terug te vinden is in hoofdstuk 7.

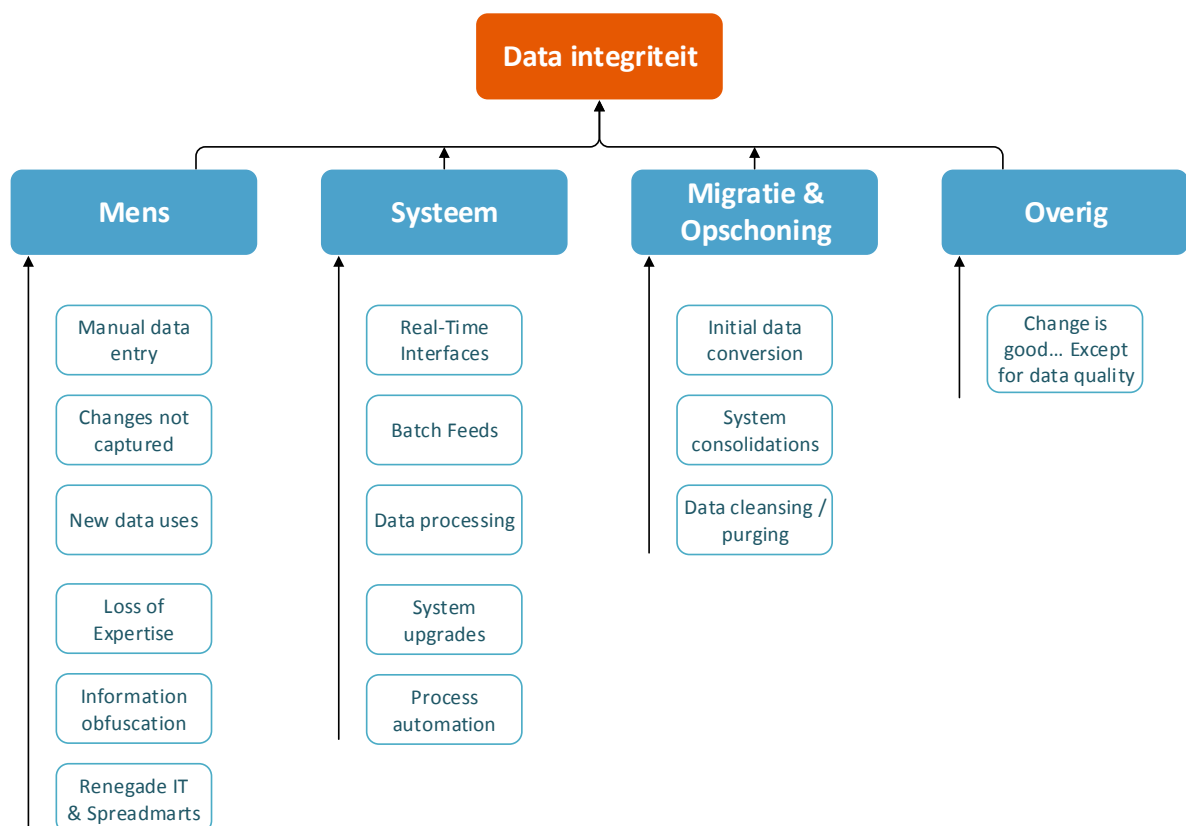
2.1. Het begrip data integriteit

Alvorens er onderzoek gedaan kan worden is het van belang om het begrip van data integriteit te specificeren. Op dit moment zijn er veel uiteenlopende opvatting omtrent dit begrip waarbij hier een uitsplitsing wordt gemaakt van het begrip 'Data' en 'Integriteit'. Data omvat een breed principe waarbij onderscheidt gemaakt kan worden tussen Transactionele data (data welke aan transformatie onderhevig is door het verwerken ervan), Analytische data (data welke gebruikt wordt bij sturing van processen en besluitvorming), Master data (referentie data voor alle gegevens in een systeem), Systeemconfiguratie data (data benodigd voor de werking en logica van het systeem) en Metadata (data over de data zelf, waarbij karakteristieken en informatie wordt opgeslagen). (Getz, 2011)

Het begrip integriteit kan daarnaast worden omschreven als: *"onschendbaarheid, eerlijkheid en onkreukbaarheid"*.

Het combineren van deze twee begrippen leidt tot de volgende omschrijving van het begrip data integriteit welke in dit rapport zal worden gehanteerd:

Data integriteit refereert naar de juistheid en betrouwbaarheid van gegevens. Deze gegevens dienen compleet, zonder variaties of gesloten compromissen vanuit het originele bestand, betrouwbaar en nauwkeurig te worden beschouwd. Pas wanneer aan al deze eisen is voldaan kan data als volledig integer worden beschouwd. (Martinenghi, 2005), (Maydanchik, 2007)



Figuur 2.2 Geïntegreerd model van 4 kwadranten aan de hand van (Maydanchik, 2007) & (Sarsfield, 2011).

Hoofdstuk 3: Methodologie

3.1 Onderzoek strategie

In onderzoeken kan er allereerst onderscheidt gemaakt worden tussen twee verschillende typen onderzoeken, het kwalitatief en kwantitatief onderzoek. Kwantitatieve onderzoeken zijn voornamelijk gericht, zoals de naam wel aangeeft, op onderzoeken omtrent hoeveelheid. Het is gericht op het cijfermatig uitspraken doen omtrent een bepaalde groep. Een voorbeeld hiervan is bijvoorbeeld de uitspraak 'onder de Nederlandse bevolking blijkt dat 34% van de mannen tijdens de lunchpauze buiten de deur eet, in tegenstelling tot 22% van de vrouwen'. Het kwalitatieve onderzoek daarentegen is gericht op diepgaandere informatie waarbij gekeken wordt naar motieven, gedrag en emoties van personen. Het gaat niet zozeer om de generaliseerbaarheid van de uitkomsten, maar meer om het doen van doelgerichte uitspraken waarbij meer rekening wordt gehouden met de doelgroep zelf. Vanwege de aard van dit onderzoek zal in dit rapport gewerkt worden met een kwalitatief onderzoek. (Bryman, 2012) & (Zikmund, Babin, Carr, & Griffin, 2012)

3.2 Onderzoek ontwerp

Voor het doen van onderzoek en het ontwerp wat hieraan ten grondslag ligt, is gebruikt gemaakt van de diverse research designs beschreven door (Bhattacharjee, 2012). Deze designs zijn de volgende:

- Experimenteel onderzoek Testen van oorzaak en effect over tijd aan de hand van twee groepen
- Veld onderzoek Niet-experimenteel onderzoek van momentopnamen
- Secundaire data analyse Analyse uitvoeren op al bestaande data van andere bronnen
- Case onderzoek Diepgaand onderzoek naar praktijk problemen over een lange tijd
- Doelgroep onderzoek 6 – 10 personen bij elkaar zetten om een probleem te bespreken
- Actie onderzoek Onderzoek aan de hand van interventie en observatie
- Etnografie Kwalitatief, waarbij onderzoeker observeert en zich aanpast op de omgeving om informatie te bemachtigen.
- Gefundeerde theorie Inductieve techniek om theorieën te bouwen op nieuwe theorieën.

Uit deze acht onderwerpen, ondersteund door (Zikmund, Babin, Carr, & Griffin, 2012) is de keuze gemaakt om een etnografisch onderzoeksmodel toe te passen bij het doen van het betreffende onderzoek. Doordat er een zeer grote groep van medewerkers binnen KLM E&M werkzaam is en deze niet allemaal persoonlijk kunnen worden geïnterviewd, is een andere aanpak benodigd.. (Bryman, 2012)

3.3 Uitvoering praktijk onderzoek

Voor het praktijk onderzoek is, zoals beschreven in paragraaf 3.2, gekozen voor een etnografische aanpak waarbij er observatie plaatsvindt in de omgeving waar het onderzoek wordt gedaan. Er is vijf maanden lang intern bij KLM E&M door mij persoonlijk onderzoek verricht waarbij er vragen zijn opgesteld. In deze vijf maanden is het proces in kaart gebracht hoe er gewerkt wordt en zijn er gesprekken met een groot aantal van hun medewerkers & managers van verschillende afdelingen gevoerd. Tijdens deze gesprekken is een enquêtestructuur als leidraad aangehouden maar is deze op een informele manier bij de betreffende personen getoetst. Om de correctheid van de onderzoeksresultaten te kunnen waarborgen zijn de uitkomsten in een later stadium wederom getoetst maar ditmaal bij andere afdelingen. Hierdoor is er discussie gevoerd over de aangedragen onderzoeksresultaten om persoonlijke kleuring, welke snel voorkomt bij etnografisch onderzoek, zoveel mogelijk te beperken (Bryman, 2012). Tevens zijn de verkregen gegevens bij een divers aantal

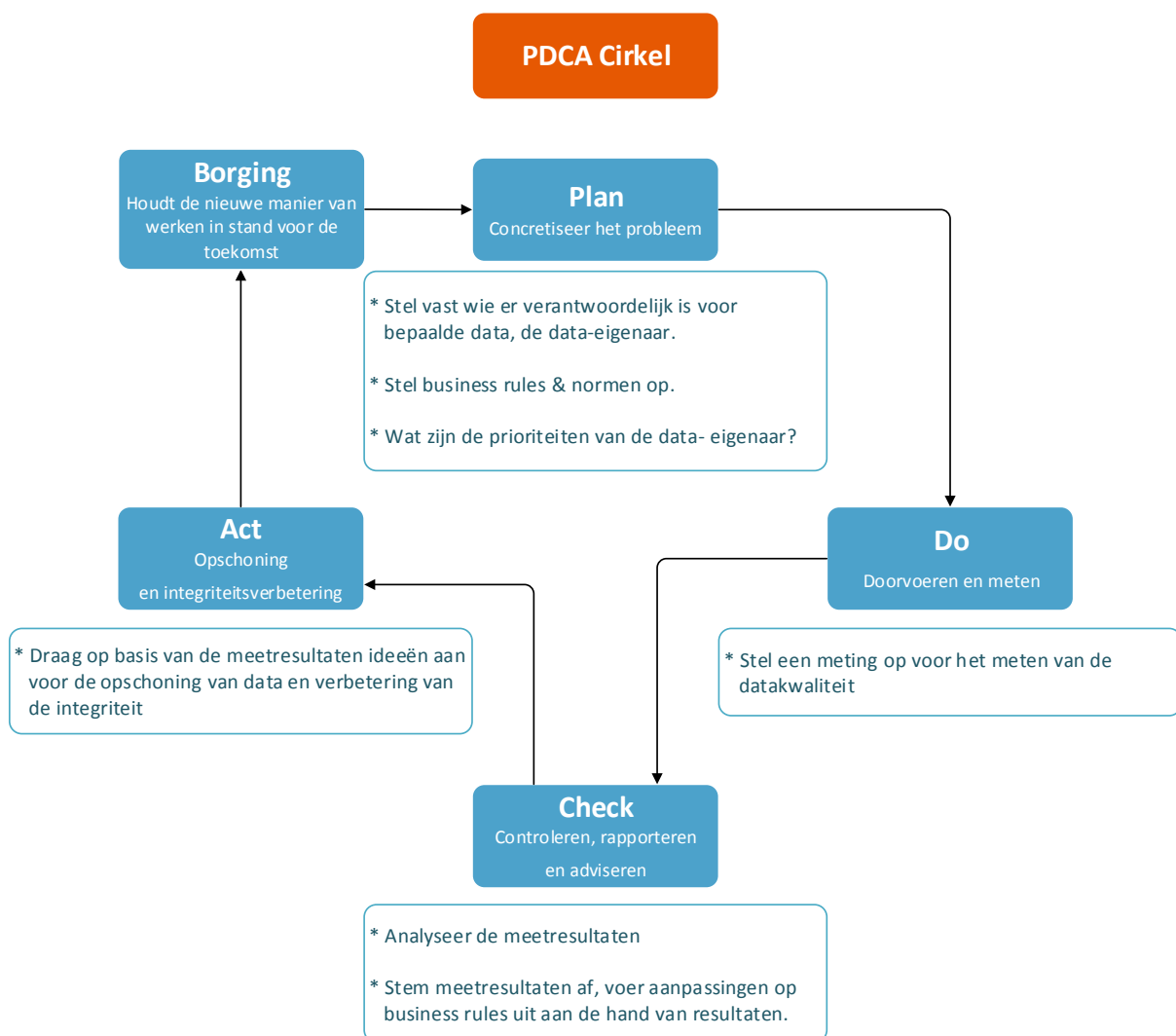
personen binnen de organisatie getoetst om persoonlijke observatie en interventie zoveel mogelijk te beperken. (Zikmund, Babin, Carr, & Griffin, 2012)

3.4 Enquête ontwerp

Voor de enquête waarover gesproken wordt in paragraaf 3.3 is gebruik gemaakt van het 4 kwadranten model weergegeven in figuur 2.2. Hierbij is gekeken naar de redenen waardoor data integriteit in het geding kan komen. Deze aanpak zal leidend zijn in het praktijkonderzoek.

3.5 Aanvulling

Naast het stapsgewijs uitvoeren van het onderzoek aan de hand van de beschreven methode in paragraaf 3.2 en 3.3 zal er tevens, indien mogelijk, gebruik worden gemaakt van de PDCA cirkel beschreven in het artikel van (Zandhuis, 2008), en te vinden in figuur 3.1. Deze Plan, Do, Check, Act cirkel met daarbij de laatste stap Borging, is onderdeel van het Lean Six Sigma principe. Dit principe richt zich op het verminderen van de 7+1 soorten waste welke bij processen te vinden zijn te weten: Transport, Inventory, Motion, Waiting, Overproduction, Overprocessing, Defects en Skills. Skills wordt vaak vergeten maar is een belangrijk onderdeel wat later is toegevoegd aan dit principe. Omdat KLM E&M steeds meer Lean tracht te werken zal, indien mogelijk, dit principe worden gebruikt.



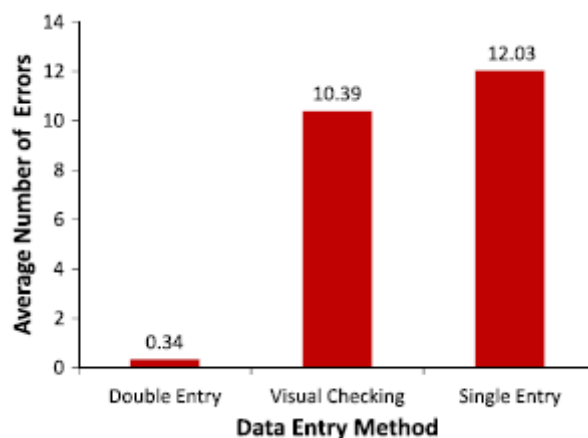
Figuur 3.1 PDCA cirkel inclusief borging beschreven in het artikel van (Zandhuis, 2008).

Hoofdstuk 4: Deelvraag 1

Aan de hand van het opgestelde 4 kwadranten model in figuur 2.2, is het mogelijk om hier gericht onderzoek naar te doen. In dit hoofdstuk wordt er per kwadrant gekeken naar de inhoud hiervan waarbij de onderliggende elementen nader toegelicht zullen worden. Tevens zal er gekeken worden naar extra literaire aanvulling welke de elementen, niet alleen individueel maar ook als geheel, beter kunnen aanvullen. Daarnaast zal er gekeken worden naar interne verbanden welke gelegd kunnen worden tussen de diverse elementen en mogelijk de kwadranten.

4.1 Mens

Binnen het eerste, en voor dit rapport belangrijkste kwadrant, vallen zes elementen welke onder de menselijke factor kunnen worden geschaard op basis van (Maydanchik, 2007) & (Sarsfield, 2011). Het manueel invoeren van data (manual data entry) is volgens beide auteurs een van de belangrijkste bronnen van data integriteitsproblemen. Het komt veelvoudig voor dat mensen de verkeerde data selecteren of juist verkeerd invoeren, zowel onopzettelijk als opzettelijk. Verder blijkt uit de onderzoeken gedaan door (Barchard & Pace, Preventing human error: The impact of data entry methods on data accuracy and statistical results, 2011) & (Barchard & Verenikina, 2013) dat het 'eenmalige invoeren van data' aanzienlijk veel meer fouten oplevert dan wanneer er nog controle op wordt toegepast, tevens te zien in figuur 4.1. Wanneer in het begin van een proces data verkeerd ingevoerd wordt hoeft dit in het begin niet direct voor problemen te zorgen, maar kan het zijn dat dit in een later stadium pas naar boven komt.



Figuur 4.1 Gemiddeld aantal fouten welke gemaakt worden wanneer er gewerkt wordt via 'dubbel invoeren', 'visueel controleren' of 'eenmalig invoeren'.

In het framework wordt daarnaast gesproken over veranderingen welke niet worden vastgelegd (changes not captured). Hierbij wordt de focus gelegd door (Maydanchik, 2007) & (Sarsfield, 2011) op metadata welke niet wordt aangepast en daardoor foutieve informatie over de databron beschrijft. Wanneer data veranderd, dient metadata ook mee te veranderen. Data welke opgeslagen is in verschillende systemen zou representatieve metadata moeten bevatten, maar door veroudering en het niet aanpassen van metadata ontstaan er verschillen tussen de representatie en de werkelijkheid (Morris, 2010). Ter illustratie: wanneer iemand verhuist zou deze bij iedereen waar het oude adres bekend is het nieuwe adres dienen door te geven. De één doet dit wel, de ander niet waardoor wijzigingen deels worden doorgevoerd. Dit leidt tot het behoud van zowel de oude als nieuwe gegevens resulterend in verwarring en een verkeerd gebruik van de betreffende gegevens door

mensen. Tevens relateerd dit aan het element nieuw gegevens gebruik (new data uses) waarbij het gaat om “de toepasbaarheid voor het gebruikersdoel”. Deze doelen verschillen per gebruiker waardoor data voor de een van grotere waarde is dan voor de ander. De manier waarop deze verschillende gebruikers met de betreffende data omgaan maakt dat dit punt in het model is opgenomen. Wanneer de waarde van data wordt onderschat, kan deze verkeerd verwerkt worden, gedupliceerd of verwijderd, waardoor de afhandeling op dat moment in de ogen van een mens correct is verwerkt terwijl dit in latere stappen voor problemen zorgt doordat er conflicten optreden. (Maydanchik, 2007)

Het verloop van kennis binnen een organisatie draagt ook bij aan de menselijke factor (Loss of expertise). Een belangrijk punt wat hierbij centraal staat is algemene kennis van het proces, maar ook diepgaandere kennis over de systemen waar mee gewerkt wordt. Deze systemen zijn in de loop der jaren aan verandering onderhevig geweest waardoor mensen welke langere tijd hiermee gewerkt hebben veel meer weten over hoe zij daar mee om dienen te gaan. Wanneer mensen de organisatie verlaten nemen zij deze opgedane kennis mee waardoor er een ‘kennis kloof’ valt indien hier niet goed op wordt ingespeeld (Cioffi & Dip, 2012).

Daarnaast kan het ook voorkomen dat men wel de correcte informatie wil gebruiken maar deze niet aangeleverd krijgt. Hierdoor wordt er met incorrecte data gewerkt, wat opzettelijk wordt ingevoerd en niet per ongeluk zoals verwerkt bij het eerste element uit dit kwadrant. Het betreft hier het moedwillig achterhouden van informatie (Information obfuscation). Er zijn tal van redenen waarom iemand toch de incorrecte data verwerkt terwijl deze persoon weet dat dit niet de juiste manier van werken is. In het dagelijks leven zou dit principe het best uitgelegd kunnen worden aan de hand van het volgende voorbeeld. Wanneer iemand zich op internet aanmeldt voor een e-mail account wordt bij deze aanmelding om een diversiteit van gegevens gevraagd, bijvoorbeeld het woonadres. Om privacy redenen kan het zo zijn dat de betreffende persoon deze gegevens niet invoert. Er wordt hierdoor opzettelijk foutieve informatie ingevoerd maar heeft dit met privacy te maken.

Het laatste element betreft de eucjes (Renegade IT and Spreadmarts). Een euc (end user computing) is een zelfgeproduceerd tool, bijvoorbeeld een Excel sheet, waarmee gewerkt wordt binnen de organisatie. Eucjes kunnen bijdragen aan een efficiëntere en daardoor snellere manier van werken. De keerzijde aan dit principe is dat een euc in veel gevallen werkt met gegevens uit een of meerdere databases. Deze gegevens worden vervolgens met elkaar gecombineerd en in de euc gebruikt. De euc wordt vaak niet gecontroleerd waardoor deze vol met fouten kan zitten en slechts een of twee personen van de afdeling weten hoe de euc werkt. Zodra deze vertrekken van de afdeling weet niemand meer hoe de euc bijgewerkt dient te worden, waardoor dit euc probleem gerelateerd is aan het verlies van kennis waarover eerder is geschreven in deze paragraaf (Wu, Chen, & Lin, 2006).

4.2 Systeem

Real-Time interfaces zijn tegenwoordig niet meer weg te denken. De mogelijkheid om direct data te synchroniseren tussen diverse systemen zorgt voor procesoptimalisering en tijdwinst. Toch blijken de termen ‘sneller’ en ‘beter’ niet met elkaar samen te gaan. Hoe sneller er uitwisseling plaatsvindt door systemen heen, hoe minder de kwaliteit van data wordt (Maydanchik, 2007). Immers: wanneer incorrecte data in een systeem wordt ingevoerd en deze wordt verwerkt, stroomt deze binnen enkele seconden in alle systemen welke aan het bronsysteem zijn geschakeld. Hoe sneller de verwerking gaat, hoe minder tijd er is voor eventuele verificatie van de ingevoerde data waardoor een database altijd de ingevoerde, mogelijke foutieve data accepteert met alle gevolgen van dien. (Lebdaoui, Orhansou, & Hajji, 2013)

Naast real-time interfaces spelen ook batch feeds een belangrijke rol. Batch feeds zijn, in tegenstelling tot real-time interfaces, grote bestanden waarin een hoop data zit verwerkt welke op vaste tijdstippen van de ene database naar de ander worden verzonden. Wanneer er geen gedegen controle wordt uitgeoefend op deze grote bestanden vol met data, kan het voorkomen dat er in de 'batch' grote fouten zitten. Wanneer dit bestand vervolgens in de database wordt geladen is direct het hele systeem bevuild met foutieve informatie welke zich direct in het systeem nestelt en het herstellen ervan problematisch en tijdrovend is. (Kourti, Nomikos, & MacGregor, 1995)

De twee elementen, real-time interfaces en batch feeds, zijn voorbeelden van het derde element namelijk data verwerking (data processing). Met data wordt 'de staat van data' bedoeld en met processing wordt het 'veranderen van deze staat' bedoeld. Tevens heeft data verwerking nog een tweede punt wat hier aan gekoppeld kan worden, namelijk wanneer de data verwerkt wordt door de verschillende afdelingen. Indien een afdeling handelingen uitvoert voor data verwerking, dit doorstuurt naar de volgende afdeling en geruime tijd later nog extra informatie toevoegt, kan dit een enorm effect hebben op de hoe de verwerking ervan wordt of is gedaan. (Anderson & Randell, 1979)

Een overkoepelende factor binnen dit kwadrant zijn de systeem verbeteringen (system upgrades). Gemiddeld worden systemen binnen een organisatie eens in de vijf jaar geüpdatet naar een nieuwere versie. Hierin zijn de vernieuwingen voortgebouwd op het al bestaande systeem, met de daarbij horende datasets. In theorie zouden systeem updates geen probleem dienen te vormen voor integriteitsproblemen, maar de praktijk blijkt anders. Updates worden getest aan de hand van test data. Deze data omvat vaak niet de volledige of al bestaande data sets waar mogelijk problemen in schuilen. Doordat hier geen rekening mee wordt gehouden kunnen systeem updates voor implementatie perfect werken, maar na implementatie zorgen voor grote problemen, voornamelijk wanneer het systeem ineens moet werken met gebrekkige betrouwbare metadata. (Maydanchik, 2007)

Proces automatisering (process automation) is tevens een element waar rekening mee moet worden gehouden. In huidige organisaties wordt steeds meer gewerkt met het automatiseren van processen vanwege een besparing op tijd en kosten. Hierbij wordt minder rekening gehouden met de problemen welke hierdoor kunnen ontstaan. Wanneer een proces stil komt te liggen is het lastiger om deze weer met de betreffende systemen aan de gang te krijgen. Tijdens dit stil liggen is het onbekend hoeveel data er verloren of veranderd wordt, met daarbij alle gevolgen van dien. Daarnaast zorgt het ervoor dat mensen niet direct problemen kunnen oplossen welke ontstaan en in het oplossen ervan veel tijd gaat zitten. De keerzijde hieraan is dat men vaak een andere, incorrecte, manier van het oplossen van het probleem waardoor integriteit in het geding komt. (Fiske & DePaola, 2006)

4.3 Migratie & Opschoning

Het initieel omzetten van data (initial data conversion) heeft betrekking op het migreren van data. Een nieuwe database begint vrijwel nooit leeg maar wordt gevuld met data afkomstig uit voorgaande databases. Voor de migratie wordt er vanuit gegaan dat de data uit de vorige database correct is, wat nooit zonder goede onderbouwing aangenomen mag worden. Tijdens de migratie kunnen er fouten optreden welke niet direct zichtbaar zijn, daardoor duiden op een correcte migratie van data, maar wederom foutieve aannames zijn. Hoe vaker data uit diverse databases wordt gemigreerd, hoe onbetrouwbaarder de data wordt (Youn & Ku, 1992). Tevens wordt er geen rekening gehouden met de diversiteit van de verschillende systemen waarbij dataopslag en verwerking niet worden meegenomen. Niet alleen speelt de data op zichzelf een belangrijke rol, ook de metadata is vaak aan verandering onderhevig, waardoor een incorrect migratieplan en uiteindelijk migratie leidt tot problemen (Youn & Ku, 1992), (Maydanchik, 2007), (Sarsfield, 2011).

Het samenvoegen van systemen en databases (system consolidations) vindt voornamelijk plaats wanneer oude systemen vernieuwd of gecombineerd worden met nieuwe systemen. In veel gevallen wordt er van te voren goed gekeken naar nieuwe functionaliteiten en veranderingen welke doorgevoerd dienen te worden voor het samenvoegen. Managers willen vaak dat het samenvoegen zo min mogelijk in tijd en geld kost en uiteindelijk snel een besparing voortbrengt. Dit overhaast te werk gaan levert volgens (Maydanchik, 2007) uiteindelijk een nieuw systeem op wat niet meer goed werkt, zoals de oude twee systemen dat wel deden. Tevens vinden binnen systeem samenvoegingen ook database migraties plaats, ware het niet dat dit specifieke element complexer is dan het initieel migreren. Bij systeem samenvoegingen moeten twee of meerdere databases aan elkaar gekoppeld en gemigreerd worden. Hierbij moeten afwegingen worden gemaakt welke data wel of niet dient te blijven bestaan. (Riverbed Technology, Inc, 2006)

Het laatste element uit dit kwadrant beslaat het opschonen of verwijderen van data (data cleansing / purging). Het opschonen en verwijderen van data gebeurt om ruimte te maken voor nieuwe data, en oude niet bruikbare data uit de database te verwijderen. Deze opschoning gebeurt steeds vaker door middel van geautomatiseerde processen waarbij complexe algoritmen worden gebruikt om oude data te traceren en te verwijderen. Doordat databases tegenwoordig heel veel data bevatten is er vrijwel geen andere manier om opschoning toe te passen dan dit geautomatiseerd te doen, tenzij er gericht wordt op een specifiek gedeelte. Binnen deze algoritmen welke gebruikt worden kunnen ook fouten zitten. Deze fouten kunnen leiden tot het verkeerd traceren en geheel of slechts gedeeltelijk verwijderen van data. Ook kan het voorkomen dat de specificaties waar het algoritme naar zoekt verkeerd zijn ingevoerd waardoor het algoritme perfect werkt, maar de verkeerde data verwijderd. (Rahm & Hong Hai, 2002)

4.4 Overig

Fusies tussen organisaties of samenwerking op een andere manier zijn onderdeel van het laatste kwadrant en mogen niet onderschat worden (Change is good... Except for data quality). Doordat twee verschillende organisaties met verschillende systemen werken, wordt er in de meeste gevallen een plan opgesteld om deze combinatie zo goed mogelijk te laten verlopen. Afwegingen worden gemaakt met welk systeem er gewerkt wordt en welke data leidend zal zijn binnen de organisaties. Doordat vaak de ene organisatie de overhand heeft op de ander wordt er vaak gekozen voor de systemen van het 'heersende' bedrijf. Hierbij wordt geen afweging van integriteit van data gemaakt, maar gekeken naar de positie van de verschillende organisaties (Fischer, Greitemeyer, Omay, & Frey, 2007). Tevens kan ook op het interne vlak verandering plaats vinden, waarbij er op een nieuwe manier gewerkt wordt. Wanneer deze verandering niet goed wordt doorgevoerd kan dit leiden tot weerstand van medewerkers. Men zal de oude manier van werken hanteren waardoor data niet meer als belangrijk wordt gezien en deze zaken langs elkaar heen gaan lopen wat ten koste gaat van de integriteit. (Saith, et al., 2013), (Cummings & Worley, 2008)

4.5 Conclusie

In dit vierde hoofdstuk zijn alle elementen van de vier kwadranten besproken van het opgestelde model. Zoals te lezen, zijn er een veel verschillen tussen de elementen welke invloed hebben op data integriteit. Deels zijn er al aanvullingen te vinden van de elementen op elkaar, zoals het element 'changes not captured' op het element 'New data uses' of de twee elementen 'Initial data conversion' samen met 'System consolidations'. Deze samenhang zal verder in het rapport worden besproken. Hierbij zal dieper worden ingegaan op de overeenkomsten en verschillen in verhouding tot de praktijk.

Hoofdstuk 5: Deelvraag 2

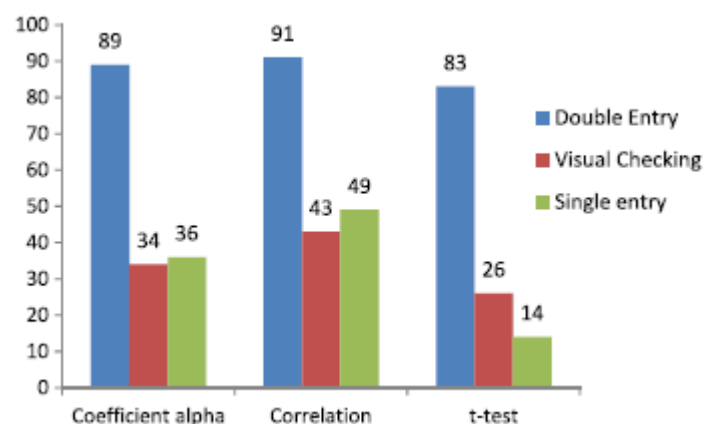
Nu in hoofdstuk vier het theoretische model is opgesteld waarmee gewerkt zal worden binnen dit rapport, kan aan de hand hiervan gericht onderzoek worden gedaan in de praktijk. De uitkomsten van het vierde hoofdstuk dragen bij aan de vragen welke zijn onderzocht in de praktijk. Deze vragen zijn gericht op de elementen welke in het vierde hoofdstuk behandeld zijn. Praktijk onderzoek heeft plaatsgevonden door deze vragen bij diverse afdelingen te stellen en zodoende tot antwoorden te komen. In dit vijfde hoofdstuk zal gekeken worden naar de uitkomsten van dit praktijk onderzoek ten opzichte van de eerste twee kwadranten waarbij Mens en Systeem centraal staan. Omdat de overige twee kwadranten Migratie & Opschoning en Overig zeer gerichte processen zijn waar binnen het onderzochte proces niet mee gewerkt wordt zijn deze buiten beschouwing gelaten bij het praktijk onderzoek. De uitkomsten in dit hoofdstuk zullen tevens de opbouw worden voor het zesde hoofdstuk waarin de analyse op de resultaten wordt toegepast. Dit praktijkonderzoek is inhoudelijk gedaan waarbij de gegevens slechts intern beschikbaar zijn.

Hoofdstuk 6: Deelvraag 3

Nu in hoofdstuk 5 de belangrijkste uitkomsten van het praktijk onderzoek getoetst aan de elementen van het framework naar voren zijn gekomen, kan in dit laatste hoofdstuk gekeken worden naar oplossingen welke de literatuur hieromtrent aandraagt. Stapsgewijs zullen de elementen, voornamelijk gericht op het kwadrant mens, worden behandeld en aangevuld worden met mogelijke aanbevelingen.

6.1. Manual data entry

Aller eerst heeft manueel invoeren van data altijd een kans op fouten. Zoals eerder aangegeven is de kans op fouten bij de invoer van data indien het eenmalig moet worden ingevoerd vele malen hoger dan wanneer dit meerdere keren ingevoerd dient te worden. (Barchard & Verenikina, 2013) & (Barchard & Pace, 2011). Uit deze onderzoeken komt naar voren dat de minste kans op fout marge wordt behaald indien de data twee keer moet worden ingevoerd waarbij het systeem controleert of deze informatie overeenkomt met elkaar, tevens weergegeven in figuur 6.1. Een voorbeeld hiervan is het wachtwoord wat bij een registratie twee keer dient te worden ingevoerd om ervoor te zorgen dat de gebruiker zeker weet dat het correcte wachtwoord is ingevoerd. Door medewerkers tijdens het proces informatie twee keer te laten invoeren kunnen problemen omtrent manuele invoer drastisch verminderen, aldus figuur 6.1.



Figuur 6.1 Percentage correcte invoer van data ten aanzien van de verschillende manieren.

6.2 New data uses & information obfuscation

Het derde element, nieuw gegevens gebruik (new data uses). Door richtlijnen op te stellen welke informatie benodigd is voor alle stappen en welke informatie later kan worden toegevoegd, kan hier integriteit bewaakt worden. Het is dan ook van belang om hier als manager het voortouw in te nemen en daar duidelijk op te sturen. (Block, 1991)

Aan dit element is tevens information obfuscation gerelateerd, het vijfde element. Feedback ten aanzien van werkzaamheden dit punt versterken. (Block, 1991)

6.3 Loss of Expertise

Het verlies van kennis is tevens een van de belangrijkste oorzaken omtrent data integriteit. Door een eenduidige manier van werken te implementeren zal er een meer gestandaardiseerde proces ontstaan waarbij kennis op een effectieve manier kan worden overgedragen. Daarnaast dragen (Maydanchik, 2007) & (Sarsfield, 2011) zelf ook manieren aan om kennis omtrent systemen en processen te bewaren. Door te documenteren wat er in de geschiedenis van een systeem allemaal is veranderd kan men dit naslaan en vervolgens hierdoor beter begrijpen waarom iets werkt zoals het werkt. De uitspraak welke hierboven is weergegeven hoeft dan in principe niet gedaan te worden. Doelgerichte systeemtraining en procestraining dragen ook bij aan niet alleen behoudt van data integriteit, maar ook een efficiënter proces.

6.4 Communication

In het 4 kwadranten framework is dit element niet naar voren gekomen maar door het praktijk onderzoek blijkt dit toch wel een sterk en belangrijk element te zijn. De studies van onder andere (Welch, 2012) & (Hulea, 2010) tonen aan dat communicatie een van de belangrijkste factoren is binnen organisaties. Door onderling meer communicatie te voeren worden medewerkers zich bewust van het totale proces en wat hun invloed hierop kan zijn. Door afdelingen bij elkaar te brengen, met elkaar te laten praten en in te laten zien wat voor invloed ze op elkaar hebben, kan niet alleen de mentaliteit maar ook de efficiëntie van het proces verbeterd worden. (Welch, 2012) Dit zorgt ervoor dat er zorgvuldiger met data wordt omgegaan in de huidige systemen waarbij correctheid een belangrijke rol in neemt.

6.5 System

Dit is wel beschreven in het tweede kwadrant maar niet meegenomen in de elementen welke hierin beschreven worden. Het systeem specifiek kan namelijk een breekpunt zijn. Wanneer er gekeken wordt naar het proces wat verwerkt dient te worden en het systeem kan gekeken worden wat mogelijke gebreken zijn welke aangevuld dienen te worden.

Hoofdstuk 7 Conclusies

De elementen welke in het framework zijn geschaard onder het kwadrant mens zijn allemaal onderzocht. Hierbij zijn tevens per specifiek element oplossingen en mogelijke verbeteringen aangedragen op basis van kennis en onderbouwing van de huidige literatuur. Wat voornamelijk naar voren is gekomen in dit onderzoek is dat de mentaliteit, gebrek aan kennis en het gebrek aan communicatie binnen organisaties tot problemen kunnen leiden. Deze drie specifieke punten kunnen aan alle elementen van het menselijk kwadrant gelinkt worden waarbij deze dan ook kunnen worden beschouwd als de basis wat behoudt en verbetering kan voortbrengen.

In het begin van dit rapport bij hoofdstuk 2 zijn twee hypothesen opgesteld aan de hand de opgestelde deelvragen waarbij focus lag op de kwadranten mens en systeem. Deze hypothesen waren als volgt:

Mens

“De grootste problemen veroorzaakt op het gebied van data integriteit ligt aan kennis, kunde en mentaliteit van de mens”. Hierbij wordt verwacht dat door een combinatie van systeemkennis en mentaliteit dit de grondslag is voor de meeste data integriteitsproblemen.

Systeem

“Doordat systemen langere tijd mee gaan en het veel geld kost om deze continue aan te passen, zullen systemen geen volledige aansluiting hebben op het huidige proces”. De verwachting is dat de manier van systeemwerking naar huidige processtandaarden een significante invloed heeft op data integriteit.

Ten aanzien van de hypothese omtrent het kwadrant mens is de hypothese bevestigd dat de grootste problemen op deze gebieden van kennis, kunde en mentaliteit zorgen voor data integriteitsproblemen. Hier dient wel aan toegevoegd te worden dat een slechte communicatie ook bijdraagt aan deze problemen.

De hypothese ten aanzien van het systeem kan, ondanks minder onderzoek, toch deels worden bevestigd. De reden waarom systemen langere tijd meegaan is inherent aan de kosten welke een vernieuwing met zich meebrengt. Hierbij moet gedacht worden dat ook kennis vermindering en de impact van een nieuw systeem een belangrijke afweging kan zijn.

Tevens is het zo dat door het onderzoek heen het framework aan variatie onderhevig is geweest op zowel het kwadrant mens als het kwadrant systeem. Toch heeft dit framework de basis gevormd van dit onderzoek waarbij wetenschappelijk is aangetoond waar data integriteit op de menselijke kant aan onderhevig is en uiteindelijk verbeterd kan worden. De speerpunten waarop verbeterd kan worden binnen organisaties zijn hieronder kort en bondig weergegeven:

- Het manueel invoeren van data verbeteren aan de hand van het dubbele invoeren van data ter controle.
- Duidelijk sturen op data kwaliteit waarbij feedback en mentaliteit van belang zijn, welke aangestuurd worden door managers.
- Richtlijnen opstellen voor standaardisatie van processen.
- Vastleggen van systeemhistorie voor het verbreden van het kenniscentrum.
- Bekijken van de huidige kennis en aanpak binnen organisaties, deze opnieuw structureren en doorvoeren. Hieruit voortvloeiend gericht training verlenen aan medewerkers omtrent deze vernieuwde werkwijze en systeemomgang.
- Afdelingen bij elkaar zetten en met elkaar laten communiceren voor procesverbetering

Door als organisatie hier mee te beginnen kunnen er stappen worden gezet ten aanzien van de behoud en verbetering van data integriteit op het menselijk vlak. Pas wanneer medewerkers om zijn en de verandering gaan doorvoeren kunnen de andere kwadranten worden aangepakt. (Cameron & Green, 2009)

Hoofdstuk 8 Discussie

Binnen dit rapport is voornamelijk de focus gelegd op het menselijke kwadrant van het opgestelde model. Dit model is opgesteld aan de hand van gevonden literatuur waarbij een selectie is gemaakt van de meest belangrijke elementen. Tijdens het rapport is naar voren gekomen dat niet alle elementen van even groot belang zijn of passen in het onderzoek. Er kan dan ook gekeken worden naar de correctheid van het opgestelde model en of hier wijzigingen op van toepassing zijn. Daarnaast wordt er aanbevolen om in verder onderzoek de andere kwadranten te onderzoeken, waarbij dieper

wordt ingegaan op voornamelijk het systeem gedeelte. Dit vergt duidelijke inzage in de diverse programmatuur en aansluitingen van systemen op elkaar, iets wat niet in dit rapport is opgenomen.

Tevens zou in vervolg onderzoek verder gekeken kunnen worden naar de diverse procesmatige stappen welke binnen een organisatie aanwezig zijn. Hierdoor kan specifieker worden gekeken of de interne handelingen welke worden verricht tevens van invloed zijn op data integriteit.

Daarnaast kan het framework wat is opgesteld onderhevig zijn aan eventuele foutmarge doordat tijdens het onderzoek meer informatie naar boven is gekomen ten aanzien van de werkelijke problemen omtrent data integriteit. Verder onderzoek naar deze specifieke problemen zoals beschreven in dit rapport wordt dan ook als een aanvulling gezien op dit gedane onderzoek.

Daarnaast kan er gekeken worden naar het implementeren en gebruik maken van nieuwe technieken. Op dit moment is RFID (Radio Frequency IDentification) een opkomende techniek waarbij met chips snel en simpel informatie kan worden uitgelezen. Wanneer processen meer naar automatisering overgaan waarbij de menselijke 'manuele factor' kan worden verwijderd, zal dit tot op zekere hoogte leiden tot verbetering. Hier dient wel van te voren onderzoek naar plaats te vinden omtrent de invloed hiervan en hoe dit het beste geïmplementeerd kan worden binnen de processen.

Hoofdstuk 9 Limitatie

Binnen dit rapport zijn diverse limitatie factoren waar rekening mee moet worden gehouden ten aanzien van de onderzoeksresultaten.

Doordat het vraagstuk zeer omvangrijk is, kan er niet met zekerheid worden gesteld dat alle aangedragen elementen van even grote invloed zijn op de data integriteit. Dit vraagstuk beslaat vrijwel alle afdelingen binnen KLM E&M waardoor er diepgaander, langer en toegespitster onderzoek dient te gebeuren om de werkelijke impact van die specifieke oorzaak op een processtap in kaart te kunnen brengen.

Daarnaast is in dit rapport een framework opgesteld aan de hand van gevonden literatuur, gebaseerd op twee auteurs welke ruime ervaring hebben in de data integriteitssector. Dit is vervolgens onderbouwd met extra aanvullende wetenschappelijke literatuur welke de specifieke punten meer heeft onderbouwd. Het blijkt aller eerst dat er op dit moment weinig wetenschappelijke literatuur beschikbaar is over dit specifieke probleem omtrent de menselijke factor. Zoals aangetoond blijkt dat het framework nog verder doorontwikkeld dient te worden ten aanzien van bijvoorbeeld de elementen 'communicatie' en 'systeem'. Doordat op dit moment weinig, specifiek mens gerichte literatuur beschikbaar is, heeft dit rapport zich voornamelijk op dit kwadrant gericht. Vanwege deze focus is er in veel mindere mate gekeken naar het tweede, derde en vierde kwadrant van het opgestelde framework. Indien er op basis van dit rapport meer onderzoek plaatsvindt, zou naar deze overige drie kwadranten aller eerst gekeken kunnen worden om data integriteit op deze aspecten te kunnen uitsluiten of aan te vullen.

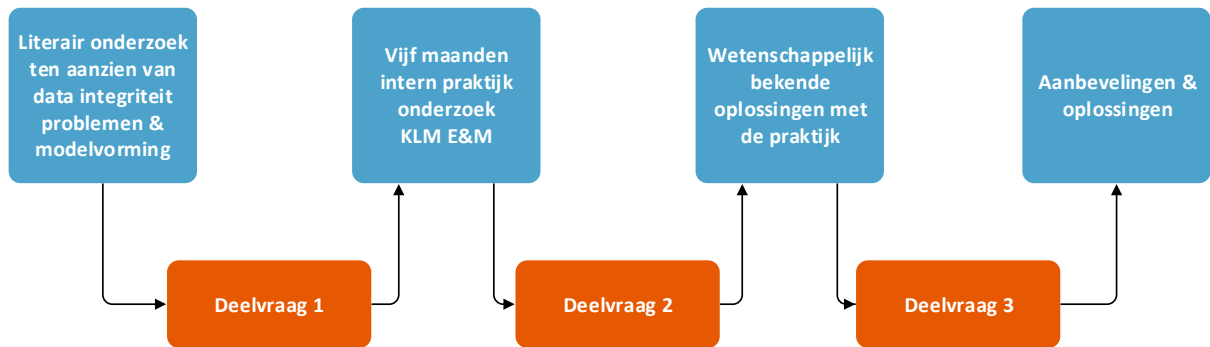
Ook is er geprobeerd het praktijk onderzoek zo objectief mogelijk te verrichten. Hiermee wordt bedoeld dat er getracht is zoveel mogelijk mensen welke in het proces werken vragen te stellen, zonder dat hier een scherpe mening in zou doordringen. Er is gebruik gemaakt van een etnografische aanpak waarbij getracht is zo objectief mogelijk te blijven. Er is getracht de objectiviteit te bevorderen door herhaaldelijk navraag te doen bij de afdelingen omtrent de gegeven antwoorden of door de vragen wederom te laten beantwoorden. Door veel meningen en mensen te ondervragen zijn er conclusies getrokken waarbij de persoonlijke mening van deze medewerkers door de bovenstaande aanpak zoveel mogelijk is getracht te verkleinen. Toch is het mogelijk dat deze filtering niet correct heeft plaatsgevonden waardoor er alsnog een interpretatie van de onderzoeker (Gerring, 2011) in de vergaring van de gegevens verwerkt kan zitten.

Referenties

- Anderson, T., & Randell, B. (1979). *Computing Systems Reliability*. New York: Hamilton Printing Company.
- Barchard, K., & Pace, L. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior - Elsevier*, 1834 - 1839.
- Barchard, K., & Verenikina, Y. (2013). Improving data accuracy: Selecting the best data checking technique. *Computers in Human Behavior - Elsevier*, 1917 - 1922.
- Bhattacharjee, A. (2012). *Social science research: Principles, methods, and practices*. Tampa: The Global Text Project.
- Block, P. (1991). *The empowered manager*. Hoboken: John Wiley & Sons Inc.
- Bryman, A. (2012). *Social Research Methods*. Oxford: Oxford University Press.
- Cameron, E., & Green, M. (2009). *Making Sense of Change Management*. London: Kogan Page Ltd.
- Cioffi, J., & Dip, G. (2012). Loss of clinical nursing expertise:. *International Journal of Nursing Practice*, 423-428.
- Cummings, T., & Worley, C. (2008). *Organization Development and Change*. London: Cengage Learning.
- Fischer, P., Greitemeyer, T., Omay, S., & Frey, D. (2007). Mergers and group status: the impact of high, low and equal group status on identification and satisfaction with a company merger, experienced controllability, group identity and group cohesion. *Community & applied Social Psychology*, 203-217.
- Fiske, T., & DePaola, S. (2006). *Does Your Automation System have Integrity?*. Massachusetts: ARC Advisory Group.
- Gerring, J. (2011). *Social Science Methodology: A Unified Framework*. New York: Cambridge University Press.
- Getz, A. (2011, 07 02). *Types of Enterprise Data (Transactional, Analytical, Master)*. Retrieved from BI-Insider.com: <http://bi-insider.com/posts/types-of-enterprise-data-transactional-analytical-master>
- Ghaeb, J., & Chebil, J. (2009). *An oblique-matrix technique for data integrity assurance*. Jordan: The Hashemite University.
- Ghaeb, J., Smadi, M., & Chebil, J. (2011). A high performance data integrity assurance based on the determinant technique. *Future Generation Computer Systems*, 614-619.
- Hulea, L. (2010). Communication - Organizations' work device. *Annals of the University of Petrosani*, 197-206.
- Klein, D., Tran-Gia, P., & Hartmann, M. (2013). Big Data. *Informatik Spektrum*, 319-323.
- Kourti, T., Nomikos, P., & MacGregor, J. (1995). *Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS*. Canada: McMaster University.

- Lebdaoui, I., Orhansou, G., & Hajji, S. e. (2013). Data integrity in Real-time datawarehousing. *World Congress on Engineering*, 3-5.
- Martinenghi, D. (2005). *Advanced Techniques for Efficient Data Integrity Checking*. Roskilde: Roskilde University.
- Maydanchik, A. (2007). *Data Quality Assessment*. Bradley Beach: Technics Publications, LLC.
- McMichael, P. (2008). *Development and Social Change*. Thousand Oaks: Pine Forge Press.
- Morris, C. (2010). More than one way to annotate metadata. *communications of the acm*, 7.
- Morrow, P. (2011). Managing organizational commitment: Insights from longitudinal research. *Vocational Behavior*, 18-35.
- Nouha, O., Leneutre, J., & Roudier, Y. (2011). Verifying remote data integrity in peer-to-peer data storage: A comprehensive survey of protocols. *Springer Science + Business Media*, 231-243.
- Rahm, E., & Hong Hai, D. (2002). *Data Cleaning: Problems and Current Approaches*. Leipzig: University of Leipzig, Germany.
- Riverbed Technology, Inc. (2006). *5 steps to successful IT consolidation*. San Francisco: Riverbed Technology, Inc.
- Saith, A., Arsel, M., Biekart, K., Chhachhi, A., O'Laughlin, B., & Storm, S. (2013). Development and Change. *Citation Reports*, 12-55.
- Sarsfield, S. (2011). *Top 10 Root Causes of Data Quality Problems*. New York: Talend.
- Sivathanu, G., Wright, C., & Zadok, E. (2005). Ensuring data integrity in storage: techniques and applications. *ACM*, 26-36.
- Tien, J. (2013). *Big Data: Unleashing information*. Miami: University of Miami.
- Todd Rose, L., & Fischer, K. (2011). *Garbage In, Garbage Out: Having Useful Data is Everything*. United States of America: Harvard Graduate School of Education.
- Welch, M. (2012). Appropriateness and acceptability: Employee perspectives of internal communication. *Public Relations Review*, 246-254.
- Wu, J.-H., Chen, Y.-C., & Lin, L.-M. (2006). Empirical evaluation of the revised end user computing acceptance model. *Computers in Human Behavior*, 162-174.
- Youn, C., & Ku, C. (1992). Data migration. *Systems, Man and Cybernetics*, 1255-1258.
- Zandhuis, J. (2008, 03 10). *Datakwaliteitsmanagement: Sturen op de datakwaliteit of blijft het dweilen met de kraan open?* Opgehaald van Sysqa:
http://www.sysqa.nl/images/stories/artikelen/071201_-_spider_koerier_-_datakwaliteitsmanagement_publicatie_1v0.pdf
- Zikmund, W., Babin, B., Carr, J., & Griffin, M. (2012). *Business Research Methods*. Mason: Cengage Learning.

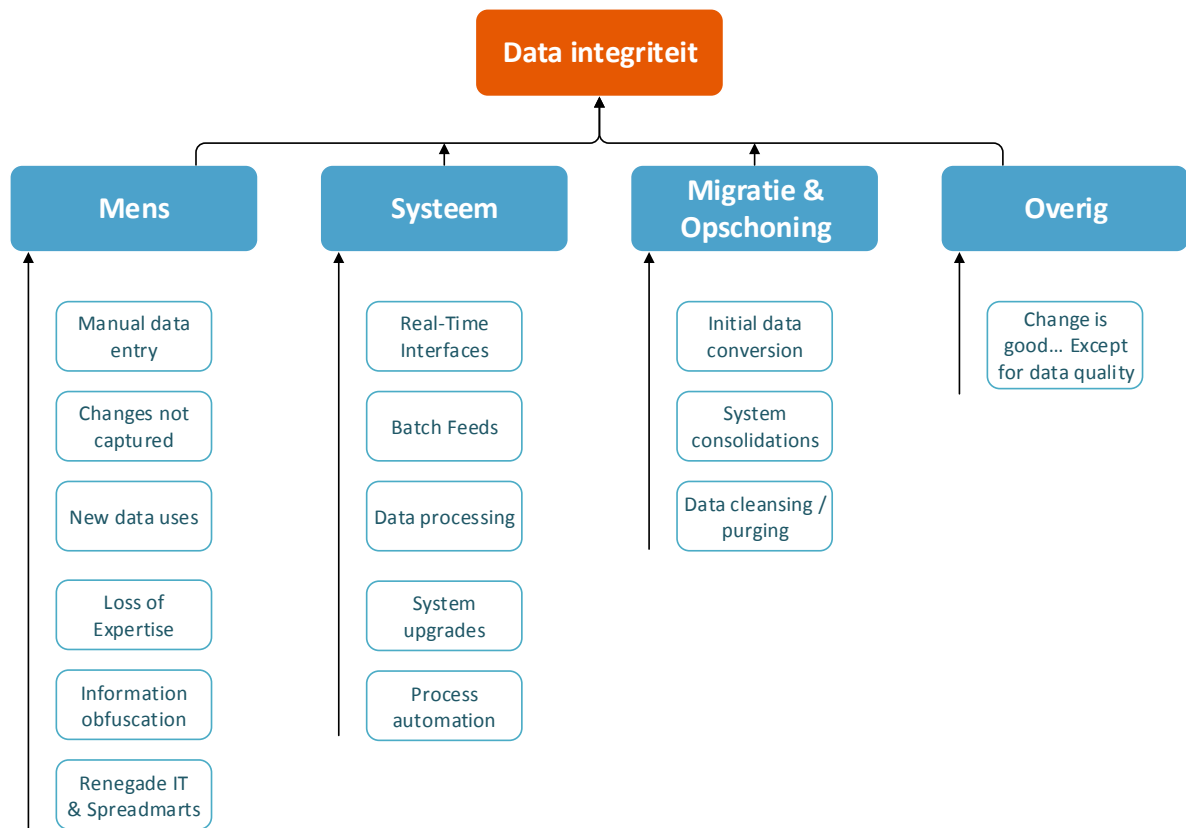
Appendix A figuren



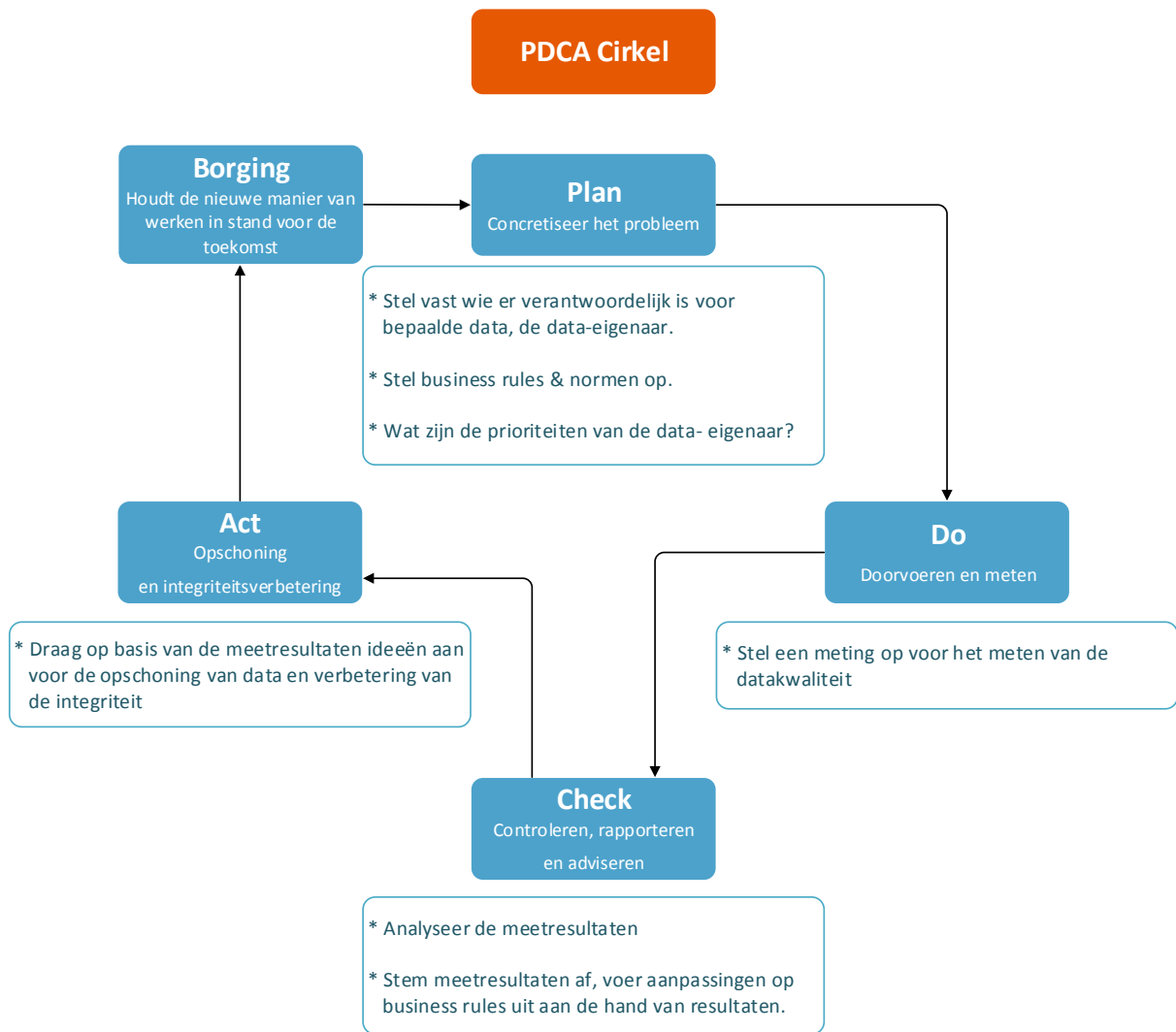
Figuur 1.1 Opbouw van het verslag ten aanzien van onderzoek en deelvragen.



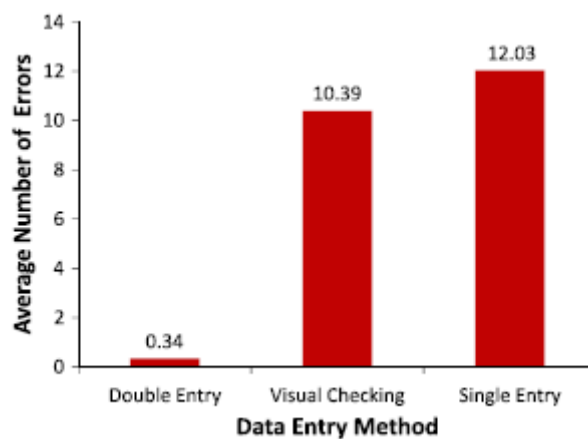
Figuur 2.1 Overeenkomsten & verschillen (Maydanchik, 2007), (Sarsfield, 2011).



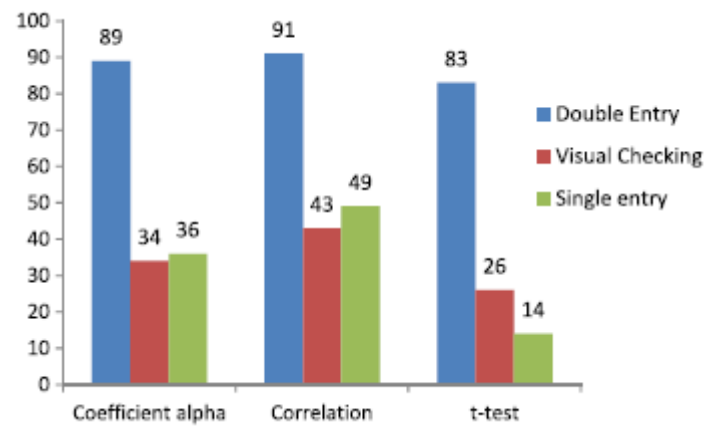
Figuur 2.2 Geïntegreerd model van 4 kwadranten aan de hand van (Maydanchik, 2007) & (Sarsfield, 2011).



Figuur 3.1 PDCA cirkel inclusief borging beschreven in het artikel van (Zandhuis, 2008).



Figuur 4.1 Gemiddeld aantal fouten welke gemaakt worden wanneer er gewerkt wordt via 'dubbel invoeren', 'visueel controleren' of 'eenmalig invoeren'.



Figuur 6.1 Percentage correcte invoer van data ten aanzien van de verschillende manieren.