

Datakwaliteit: een niet te onderschatten onderwerp

47

Trefwoorden:

LIS, zienswijze CBP, datakwaliteit, kwaliteitsmodel

In 2009 heeft het College bescherming persoonsgegevens in een zienswijze¹ geconstateerd dat het Landelijk Informatiesysteem Schulden een registratiesysteem is dat niet voldoet aan de eis van datakwaliteit ingevolge de Wet bescherming persoonsgegevens. Dit artikel wil een lans breken voor meer aandacht voor datakwaliteit met behulp van deze casus. Daartoe ga ik eerst in op wat de LIS-registratie feitelijk inhoudt en bespreek ik kort de zienswijze van het college. Daarna volgt een nadere analyse van de LIS-casus, waarbij in het bijzonder op het punt van datakwaliteit wordt ingegaan. Vervolgens wordt duidelijk gemaakt dat het belang van datakwaliteit groter wordt in onze informatiemaatschappij. Ten slotte besteed ik aandacht aan de verschillende dimensies van datakwaliteit volgens een kwaliteitsmodel. Aan de hand van dat model wordt duidelijk gemaakt wat slechte datakwaliteit veroorzaakt, waarbij ook wordt toegelicht welk kwaliteitsaspect in de LIS-casus ontbrak. De conclusie luidt dat datakwaliteit een lastig en onderschat onderwerp is, waaraan meer aandacht zou moeten worden besteed dan gewoonlijk in de praktijk gebeurt.

1 Aanleiding: het Landelijk Informatiesysteem Schulden en de zienswijze CBP

1.1 Het LIS

Het Landelijk Informatiesysteem Schulden (LIS) is een informatiesysteem dat tot doel heeft het beschikbaar stellen van informatie over de financiële positie van de consument aan LIS-deelnemers en -participanten. Het fungeert als stelsel van (krediet)registratie om voorafgaand aan de totstandkoming van een krediet- of koopovereenkomst en tijdens het beheer daarvan te kunnen beoordelen of die overeenkomst in de gevraagde vorm verantwoord is.² Het LIS registreert daarom de achterstanden op het betalen van de huur en andere rekeningen. Het systeem moet net als het Bureau Krediet Registratie (BKR)³ verhinderen dat mensen met schulden nog

een lening kunnen aangaan. Simpel gezegd beoogt het LIS door middel van registratie van die betalingsachterstanden het ontstaan van (nog meer) problematische schulden bij consumenten te voorkomen.

Het LIS kent een ruime kring van deelnemers en participanten, zoals de leden van Aedes vereniging van woningcorporaties, de leden van de Nederlandse Vereniging voor Volkskrediet (NVVK)⁴, energieretailbedrijven aangesloten bij EnergieNed of Nederlandse Vereniging voor Marktwerking in Energie (VME), de Stichting Leger des Heils Welzijns- en Gezondheidszorg, gemeentelijke sociale diensten en participanten die deelnemen aan het BKR, zoals banken, creditcardmaatschappijen, telefoonbedrijven en thuiswinkels.

Opname in het LIS vindt plaats wanneer op basis van bepaalde criteria in redelijkheid kan worden voorzien dat een natuurlijke persoon niet langer zijn schulden kan betalen of anderszins niet meer aan zijn betalingsverplichting kan voldoen. De LIS-deelnemers hebben bepaald dat zij op basis van in hoofdlijnen de navolgende schulden en achterstallige termijnen zullen registreren:

1. energieschuld; twee maanden aaneengesloten, ten minste € 250;
2. huurschuld, twee maanden aaneengesloten, ten minste € 750;
3. eigen bijdrage Stichting Leger des Heils, twee maanden aaneengesloten, ten minste € 1000;
4. schuld sociale dienst; twee maanden aaneengesloten, ten minste € 250.

Verder vindt registratie in het LIS plaats indien door een NVVK-lid een plan van aanpak wordt opgesteld in het kader van een traject schuldhulpverlening.

1.2 Zienswijze CBP LIS II

In oktober 2009 bericht het College bescherming persoonsgegevens (CBP) dat ook het tweede ontwerp voor een Landelijk Informatiesysteem (LIS) niet voldoet aan de eisen van de Wet bescherming persoonsgegevens (Wbp). Het CBP heeft in een zienswijze⁵ de LIS-registratie beoordeeld op de aspecten doelomschrijving, toetredings- en registratiecriteria. Het CBP constateerde dat het LIS een registratiesysteem is met een stigmatiserend karakter waarbij de criteria voor registratie in het systeem onvoldoende zijn onderbouwd. Het kan daarom niet als een objectief middel beschouwd worden om problematische

* Ellen Hoving is partner bij FPLC, Assurance in Legal & Compliance. Zie <www.fplc.nl>.

1 Zie zienswijze Landelijk Informatiesysteem Schulden (II), z2009-00683 en P&I 2009-6, Berichten van het College bescherming persoonsgegevens, p. 305 en 306.

2 Aldus art. 3.1 Protocol Landelijk Informatiesysteem Schulden.

3 De Wet financieel toezicht verplicht bedrijfsmatige kredietverleners die consumptieve kredieten verstrekken, zich aan te sluiten bij een stelsel van centrale kredietregistratie. In Nederland is dit stelsel ondergebracht bij BKR: het Centraal Krediet Informatiesysteem (CKI).

4 NVVK is de branchevereniging voor schuldhulpverleners.

5 Zie zienswijze Landelijk Informatiesysteem Schulden (II), z2009-00683.

schulden te registreren. Het is een gegevensverwerking waarvan het doel niet duidelijk genoeg is bepaald (art. 7 Wbp) en die niet voldoet aan het proportionaliteitsbeginsel (art. 8 aanhef en onder f Wbp). Het ontbreken van objectieve criteria kan voorts leiden tot een verdere verwerking die valt buiten de oorspronkelijke doelstelling en tot een bovenmatig gebruik van gegevens (art. 9 jo. art. 11 lid 1 Wbp). Bovendien is er een grote foutmarge mogelijk in de te registreren gegevens (o.a. bij achterstanden in betaling van huur en energie). Iedereen kent wel de situatie dat een rekening niet of niet tijdig bij de juiste persoon terecht komt door verhuizing of vakantie. Er is dus een groot risico dat het systeem ‘vervuild’ raakt met schulden die er niet thuishoren, hetgeen leidt tot het verwerken van gegevens die niet juist en nauwkeurig zijn (art. 11 lid 2 Wbp). Hierdoor kunnen veel mensen onterecht te maken krijgen met de negatieve gevolgen van registraties. Zo kunnen onterechte registraties tot gevolg hebben dat mensen verstoken blijven van belangrijke goederen en diensten als gas, water en licht.

1.3 Een nadere beschouwing van de zienswijze

Het CBP heeft in de zienswijze twee fundamentele bezwaren tegen het LIS:

1. Objectieve criteria ontbreken. De doelomschrijving en het ontwerp van het LIS worden gekenmerkt door ruime en subjectieve begrippen. Onvoldoende duidelijk wordt op grond van welke objectieve criteria voorzienbaar is dat er een risico is op een problematische schuldsituatie. Pas dan is registreren volgens de zienswijze gerechtvaardigd.
2. De te registreren gegevens, zoals de achterstanden in huur en energie, kunnen onvoldoende juist en nauwkeurig zijn.

Volgens de zienswijze ontstaat er derhalve het risico dat er mensen in het LIS worden geregistreerd die gezien de doelstelling van het LIS niet in het systeem thuishoren.

Ad 1. Het ontbreken van de objectieve criteria

Het Romeinse adagium ‘Da mihi factum, dabo tibi ius’ geeft aan, dat degene wiens oordeel over een juridische kwestie wordt gevraagd, eerst informatie over de feiten behoeft. Dit geldt niet alleen voor de rechter die een geschil beslecht, maar in zijn algemeenheid voor elke rechtshandeling die consequenties heeft voor een derde. Dit geldt mijns inziens te meer indien een derde mogelijk wordt uitgesloten van een fundamentele dienst zoals de levering van energie of het verkrijgen van huisvesting via huur. Een rechtshandeling kan weliswaar een wilsuïting zijn, maar is geen willekeur door een inzichtelijke motivering aan de hand van toepasselijke regels en relevante feiten. Vertaald naar een LIS-casus: is het feit dat een energieschuld van € 250 die twee maanden aaneengesloten niet wordt voldaan, een voldoende objectief gegeven op grond waarvan voorzienbaar is dat de schuldenaar het risico loopt in een problematische schuldsituatie te geraken?

Ten tijde van het uitbrengen van de tweede zienswijze ontbraken er objectieve maatstaven in de vorm van bedragen of exacte aanduidingen van achterstanden van betaling om te bepalen of er problematische schulden zullen ontstaan. Van problematische schuldsituaties en achterstanden zijn alleen vrij algemene onderzoeksgegevens beschikbaar. Er zijn derhalve geen passende cijfers beschikbaar ter onderbouwing van de registratiecriteria van het LIS als betrouwbare indicatoren voor het dreigen van problematische schulden. Kortom, slechts gezien de specifieke omstandigheden van het geval kan worden beoordeeld of de persoon in kwestie in een problematische schuldsituatie verkeert of dreigt te belanden.

Het doel van het LIS is het voorkomen van problematische schulden door toetsing aan bepaalde van tevoren vastgestelde criteria. Indien deze criteria niet valide worden geacht, dan leidt dat in casu tot een cascade van juridische ‘verwijten’, omdat de grondslag voor de gegevensverwerking in de zin van art. 8 Wbp ontbreekt. Daarmee is de verwerking eigenlijk al onrechtmatig. Immers: de verwerking van persoonsgegevens vindt plaats binnen een onhelder kader waarbij de selectiecriteria onvoldoende zijn om het gewenste doel te bereiken. De gegevensverwerking voldoet dus niet aan het proportionaliteitsbeginsel, dat inhoudt dat de inbreuk op de belangen van de betrokkene noodzakelijk moet zijn en dus niet onevenredig mag zijn in verhouding tot het met de verwerking te dienen doel. Met de verwerking is het beoogde doel niet te bereiken en dit doel kan dan ook beter via een andere weg worden bereikt. Voorts kan een verdere verwerking vallen buiten de oorspronkelijke doelstelling en kan dit leiden tot een bovenmatig gebruik van gegevens. Overtredingen van art. 7, 8, 9 en 11 Wbp zijn derhalve het gevolg van de eerste vaststelling, het gebrek aan objectieve criteria om het gestelde doel te bereiken.

Ad 2 Onvoldoende juist en nauwkeurig

Het tweede bezwaar betreft het niet voldoende juist en nauwkeurig zijn van de te registreren gegevens: het aspect van de kwaliteit van de data. Dit bezwaar hangt nauw samen met het eerste bezwaar. Als de gekozen registratiecriteria onvoldoende betrouwbare indicatoren zijn voor de beoordeling of betrokkene in een problematische schuldsituatie belandt, is er een behoorlijke kans aanwezig dat een betrokkene onterecht in de LIS-registratie wordt opgenomen. Met welke foutmarge in de LIS-registratie hebben we te maken? Informatie daarover ontbreekt en het CBP acht informatie daarover van groot belang, in het bijzonder omdat het LIS een negatief registratiesysteem is.

Wel is door het LIS verwezen naar de Monitor Betalingsachterstanden (2008)⁶. Daaruit blijkt dat er in 2008 581 000 huishoudens waren die in een periode van twaalf maanden een achterstallige rekening hadden. Dit is 8% van het totaal aantal huishoudens in Nederland. Uit de Monitor Betalingsachterstanden blijkt ook, dat in 66,9%

⁶ Monitor Betalingsachterstanden, nulmeting 2008, Panteia 17 november 2008.

van de gevallen de achterstallige rekening een financiële reden had. Dit betekent dat in ongeveer 33% van de gevallen de achterstand een andere oorzaak had: vergeten, achterstand in eigen administratie, een fout in de administratie van een organisatie, of anderszins.

Aangezien de kwaliteit van de registratiecriteria en de foutmarge niet zijn onderbouwd en er uit de Monitor Betalingsachterstanden blijkt dat van de onderzochte betalingsachterstanden ongeveer een derde geen financiële oorzaak heeft, is de kwaliteit van de registratie onvoldoende gewaarborgd. Een betalingsachterstand kan bijvoorbeeld ook ontstaan omdat de schuldenaar een conflict heeft met de schuldeiser over de kwaliteit van het geleverde, óf er bijvoorbeeld een discussie is of er überhaupt geleverd is.

Er bestaat derhalve een gerede kans dat mensen geregistreerd worden die daar gezien de doelstelling van het LIS niet thuishoren. Deze betrokkenen krijgen te maken met de negatieve gevolgen van een registratie, zoals bijvoorbeeld het weigeren van een huur- of energieovereenkomst. Gelet op het voorgaande is de LIS-registratie in strijd met het proportionaliteitsbeginsel en art. 11 lid 2 Wbp, dat bepaalt dat de verantwoordelijke de nodige maatregelen treft opdat persoonsgegevens, gelet op de doeleinden waarvoor zij worden verzameld en verwerkt, juist en nauwkeurig zijn.

2 Belang Datakwaliteit

Het oordeel van het CBP over de datakwaliteit is terecht. De nauwkeurigheid en juistheid van het LIS-systeem liet te wensen over. Nu zal geen enkel systeem honderd procent nauwkeurig kunnen zijn en een zekere foutmarge bevatten. Dit is inherent aan de hoeveelheid gegevens die worden verwerkt. Nu gaat het kwaliteitsvoorschrift van art. 11 Wbp ook niet uit van een absolute verplichting, maar neigt dit veeleer naar een inspanningsverplichting van de verantwoordelijke. Een garantie voor de juistheid van de gegevens kan niet van de verantwoordelijke worden gevergd, maar wel dat hij de nodige maatregelen treft die in redelijkheid van hem kunnen worden gevergd om te bewerkstelligen dat de gegevens juist en nauwkeurig zijn.⁷ Art. 11 legt op de verantwoordelijke een continue verplichting tot toetsing⁸ en draagt hem op ervoor te zorgen dat onnauwkeurige en onvolledige gegevens worden uitgewist of gecorrigeerd.⁹ De genoemde bezwaren van het CBP zijn niet op een simpele manier op te lossen door het Protocol of het LIS-systeem aan te passen, maar vraagt om een fundamentele aanpak van datakwaliteit. Helaas wordt datakwaliteit vaak beschouwd als een puur IT-probleem en het gevolg daarvan is dat organisaties te weinig aandacht geven aan dit on-

derwerp, terwijl het belang van data en de consequenties van incorrecte data voor de burger steeds groter worden.

2.1 De kwetsbare burger

Kenmerkend voor de informatiemaatschappij is dat de besluitvorming steeds meer gebaseerd is op informatie die is gegenereerd door diverse technologieën.¹⁰ De juistheid, rechtmatigheid en doelmatigheid van beslissingen wordt meer en meer afhankelijk van de kwaliteit van de verzamelde data en de daaruit gegenereerde informatie. Hierin schuilt een groot risico, want wat is de kwaliteit van die informatie? Wie waakt er over de juistheid van de gegevens in de vele databestanden, wie controleert of de juiste gegevens wel op het goede moment uit al die bestanden gehaald worden en of er geen ergerlijke of zelfs ernstige problemen voor de betrokkenen ontstaan als de onjuiste gegevens worden verbonden met juiste gegevens? De burger wordt steeds afhankelijker en dus kwetsbaarder. Centralisatie van systemen, stroomlijning van gegevensverkeer en keteninformatisering zijn belangrijke ontwikkelingen bij de overheid en het bedrijfsleven. Een fout in één systeem heeft gevolgen voor heel de keten. Bij welk loket kan de burger in de keten van gebruikers terecht om de fout te laten herstellen? Hoelang moet hij wachten voordat de correctie in alle aangehaakte systemen is doorgevoerd?¹¹

3 Het begrip datakwaliteit

De kwaliteit van data komt niet uit de lucht vallen. Cruciaal is vast te stellen welke eisen daaraan gesteld moeten worden (het 'wat') en hoe die eisen vervolgens in de praktijk moeten worden vertaald naar concrete acties (het 'hoe'). Over het begrip datakwaliteit is in de vakliteratuur het nodige geschreven en er bestaan inmiddels veel definities over dit begrip.¹² Voor dit artikel wordt volstaan met de in de literatuur algemeen geaccepteerde 'fit for purpose'-kwaliteitsdefinitie: '*data has quality if it satisfies the requirements of its intended use*'.¹³ Wat datakwaliteit inhoudt is derhalve afhankelijk van het doel, nut en noodzaak van het beoogde gebruik en van de verwachtingen die de gebruiker ervan heeft. De kwaliteit van data is derhalve subjectief en contextueel bepaald.

In de praktijk worden de kwaliteit en bruikbaarheid van gegevensverzamelingen, zoals het LIS, bepaald door drie elementen:

1. Inherente kwaliteit – de fundamentele eigenschappen van gegevens zelf, ook wel de inherente eigenschappen genoemd. Bijvoorbeeld de juistheid van een gegeven.

7 Kamerstukken II 1997/98, 25 892, nr. 3, p. 97 (MvT).

8 Kamerstukken II 1997/98, 25 892, nr. 3, p. 96 (MvT).

9 Zie in dit verband de tekst van art. 6 lid 1 onder d Privacyrichtlijn.

10 Denk bijvoorbeeld aan biometrie, nanotechnologie, informatietechnologie en cognitieve technologie. Zie voor het gebruik van diverse technologieën en het effect op privacy 'Technologie, cognitie en justitie', *Justitiële verkenningen* 2008-1.

11 Zie hierover C. Prins, 'Technocratie en de toekomstagenda van de Nationale ombudsman', in: A.F.M. Brenninkmeijer e.a., *Werken aan behoorlijkheid, de Nationale ombudsman in zijn context*, Den Haag: Boom Juridische uitgevers 2007, p. 111-134.

12 Mouzhi Ge & Markus Helfert, *A Review of Information Quality Research* (completed paper, School of Computing), Dublin City University.

13 R.Y. Wang, M. Ziad & Y.W. Lee, *Data Quality*, Kluwer Academic Publishers 2001, p. 4.

2. Kwaliteit van de meta-informatie – de mate waarin informatie over de verkregen data voorhanden is binnen de gegevensverzameling. Bijvoorbeeld de herkomst van gegevens.
3. Systeemkwaliteit – de kwaliteit en de functionaliteit van het informatiesysteem waarin de informatie is opgeslagen of waarmee de informatie wordt benaderd.¹⁴

Elk van deze elementen kan doordat zij kwalitatief niet in orde is een bron van fouten vormen.

3.1 Kwaliteits- en verbetermodel

Een belangrijke vraag is: hoe kan datakwaliteit worden beoordeeld? Wetenschappelijk zijn er allerlei denk- en referentiemodellen ontwikkeld. Het model Total Data Quality Management (TDQM) is een belangrijk conceptueel denkmodel op het gebied van datakwaliteit. TDQM is ontwikkeld door het Massachusetts Institute of Technology (MIT) in de Verenigde Staten.¹⁵ Vooral R.Y. Wang, indertijd verbonden als hoogleraar aan het MIT, heeft aan dit model een belangrijke bijdrage geleverd.¹⁶ In het TDQM-model wordt de kwaliteit van de data vanuit het perspectief van de gebruiker beoordeeld op vijftien dimensies die in vier categorieën kunnen worden verdeeld. Zie onderstaande tabel 1.

Categorie	Dimensie
Intrinsieke datakwaliteit	Accuraatheid, objectiviteit, geloofwaardigheid, reputatie
Toegankelijkheid	Autorisatie, beveiliging
Contextuele datakwaliteit	Relevantie, toegevoegde waarde, tijdsafhankelijkheid, volledigheid, hoeveelheid informatie
Datakwaliteit naar gebruikers	Eenduidige interpretatie, gemakkelijk te begrijpen, beknopte voorstelling, consistente voorstelling

Tabel 1. Datakwaliteit, categorieën en dimensies

Problemen met datakwaliteit kunnen worden verbeterd volgens het onderstaande verbetermodel van TDQM, en deze verbetering bestaat normaliter uit vier stappen:

1. Definieer: welke kwaliteitseisen worden er aan de data gesteld? Voorbeeld in de LIS-casus: de informatiemet betrekking tot de betalingsachterstand in het systeem moet voor 95% accuraat zijn.
2. Meet: de huidige situatie in het systeem wordt beoordeeld en de bevindingen worden geregistreerd.
3. Analyseer: de achterliggende oorzaken van de kwaliteitsfouten worden geanalyseerd en de bron van de kwaliteitsproblemen wordt vastgesteld.
4. Verbeter: het probleem wordt verholpen door (a) de vervuilde data te schonen en (b) de bronoorzaken van vervuiling aan te pakken.

Op deze manier kunnen problemen met de datakwaliteit worden opgepakt.

4 Oorzaak slechte datakwaliteit

Zoals hierboven al is aangegeven is het, naast het feit dat er kwaliteitseisen voor de data worden geformuleerd, belangrijk om voortdurend vast te stellen waar de bron van fouten zich bevindt en hoe deze fouten ontstaan. Alleen met dat inzicht kunnen data kwalitatief verbeterd worden. Het inzicht is ook nodig om bijvoorbeeld vast te stellen op welk kwaliteitsaspect de LIS-casus spaak liep.

In alle fasen van het informatieproces kan de kwaliteit van de informatie beïnvloed worden, bij input, transmissie, processing, opslag en output van data. Wang¹⁷ e.a. onderscheiden in hoofdzaak negen keydatakwaliteitsproblemen (DQ-problemen) in informatiesystemen. Zij onderscheiden daarbij DQ-problemen veroorzaakt bij (1) het informatieproductieproces, (2) opslag en toegang tot informatie en (3) de informatie-uitnutting door gebruikers. Hieronder wordt meer in detail een overzicht gegeven van de negen key-DQ-problemen:

1. Diverse bronnen van dezelfde informatie produceren verschillende informatie

Dit DQ-probleem betreft het opslaan van dezelfde soort informatie in verschillende systemen voor verschillende doeleinden.

De betrokken kwaliteitsdimensies zijn onder meer consistentie en geloofwaardigheid.

2. De geproduceerde informatie is gebaseerd op interpretaties, die kunnen leiden tot subjectieve oordelen die niet hoeven te sporen met de werkelijkheid

In de LIS-casus is sprake van dit probleem. Het gegeven dat er sprake is van een bepaalde huurachterstand wil nog niet zeggen dat de betrokkene in een problematische schuldsituatie verkeert. Opname in het LIS-systeem leidt wel tot die suggestie en op basis daarvan wordt betrokkene mogelijk ten onrechte bejegend, omdat hem een bepaalde dienst niet wordt geleverd of tegen andere, nadeliger voorwaarden.

De betrokken kwaliteitsdimensies zijn onder meer consistentie, objectiviteit en geloofwaardigheid.

3. Systematische fouten in informatieproductie leiden tot verkeerde informatie

Belangrijke risico's zijn gelegen in de ontwikkeling van het informatiesysteem. Soms kan bepaalde informatie niet in de velden worden ingevuld, omdat het systeem dat niet toelaat.

De betrokken kwaliteitsdimensies zijn onder andere accuraatheid, volledigheid en consistente voorstelling.

14 E. Bouman, *SmarTest*, 2008, p. 43 e.v.

15 Zie voor meer informatie <<http://web.mit.edu/tdqm/www/about.shtml>>.

16 R.Y. Wang, M. Ziad & Y.W. Lee, *Data Quality*, Kluwer Academic Publishers 2001, zie vooral hfdst. 1.

17 D.M. Strong, Y.W. Lee & R.Y. Wang, 'Data Quality in Context', *Communications of the ACM* 1997, Vol. 40, nr. 5.

4. Meer informatie betekent niet altijd betere informatie

Grotere hoeveelheden informatie vormen een uitdaging om die informatie op een snelle manier te ontsluiten en hebben daarmee ook invloed op de toegevoegde waarde van die informatie. Het beschikken over veel gegevens is op zichzelf niet bijzonder nuttig, zolang deze gegevens niet in context worden geplaatst en niet nader worden geanalyseerd.

De betrokken kwaliteitsdimensies zijn onder meer toegankelijkheid, beknopte voorstelling, tijdsafhankelijkheid en toegevoegde waarde.

5. Gedistribueerde databases leiden tot inconsistente definities, formats en waarden

Waarden in het ene systeem kunnen afwijken van de waarden in het andere systeem, maar hebben wel dezelfde definitie. Het ene systeem kent bijvoorbeeld als aanduiding van het geslacht de waarden 0 (man) en 1 (vrouw) en het andere systeem M en V. Dat wordt pas echt een probleem als de waarden elkaar tegenspreken en ingelezen worden naar een ander overkoepelend systeem.

De betrokken kwaliteitsdimensies zijn onder meer consistente representatie, toegevoegde waarde en autorisatie.

6. Informatie is moeilijk toegankelijk

Tegenwoordig kan allerlei en veel informatie worden verzameld en opgeslagen. Opslag is geen probleem, wel de informatie op een beknopte manier beschrijven met het doel de informatie op een simpele en snelle manier toegankelijk te maken.

De betrokken kwaliteitsdimensies zijn onder meer beknopte representatie, toegevoegde waarde en autorisatie.

7. Analyses van verschillende informatieverzamelingen is lastig

Om (geaggregeerde) informatie samen te stellen uit gegevens van de verschillende (bron)bestanden is het veelal noodzakelijk om gegevens samen te voegen, te ontleden en te analyseren. Zoals hierboven al is aangegeven wordt dit proces bemoeilijkt omdat gedistribueerde databases kunnen leiden tot inconsistente definities, formats, e.d. Informatie kan op deze manier niet structureel verzameld worden.

De betrokken kwaliteitsdimensies zijn onder meer consistente representatie, toegevoegde waarde en autorisatie.

8. Wijzigingen in de organisatie of de gebruikerstaken leiden tot wijzigingen in de informatiebehoefte

Informatie is alleen van hoge kwaliteit indien zij ten minste tegemoet komt aan de eisen van gebruiker. Wijzigingen in de organisatie of in de gebruikerstaken kunnen leiden tot wijzigingen in de informatiebehoefte. Er ontstaat bijvoorbeeld een mismatch tussen de beschikbare informatie en de informatie die nodig is voor de gewijzigde taak.

De betrokken kwaliteitsdimensies zijn te vinden in de categorie contextuele datakwaliteit en betreffen relevantie, toegevoegde waarde en autorisatie.

9. Toegang tot informatie wordt bemoeilijkt door allerlei andere belangen

Voor gebruikers van informatie is het belangrijk dat de informatie gemakkelijk toegankelijk is en verspreid kan worden. Toegang tot informatie wordt bemoeilijkt door allerlei andere belangen, zoals bijvoorbeeld eisen die gesteld worden in verband met de beveiliging van informatie of privacybelangen van burgers.

De betrokken kwaliteitsdimensies zijn beveiliging, autorisatie, toegevoegde waarde en volledigheid.

Uit het bovenstaande blijkt dat een complete en samenhangende aanpak essentieel is voor goede datakwaliteit.

De kwaliteitsbevordering van de data wordt pas echt aangepakt als het probleem bij de bron wordt opgepakt. Het is daarom van belang het hele proces te overzien van begin tot het einde door bij elk nieuw systeem vast te stellen welke data worden verzameld en het systeem te monitoren volgens de vier verbeterstappen van het TDQM-model.

5 Ten slotte

Dit artikel heeft een lans willen breken voor meer aandacht voor datakwaliteit door aan de hand van de LIS-casus en de zienswijze van het CBP een aantal aspecten daarvan te bespreken. Uit de analyse blijkt dat het in de LIS-casus met name ontbrak aan objectieve criteria om vast te stellen of er sprake is van een werkelijk problematische schuldsituatie bij betrokkene. Deze omissie leidt tot een cascade van onrechtmatige verwerkingen, omdat er een gegevensverwerking kan ontstaan waarvan het doel niet duidelijk is bepaald en die niet voldoet aan de proportionaliteitseis. Voorts kan een verdere verwerking leiden tot een bovenmatig gebruik van gegevens en tot het verwerken van gegevens die niet juist zijn.

De genoemde bezwaren zijn niet simpel op te lossen, maar vragen om een gedegen aanpak van de kwaliteit van de data. Zeker in een tijd waarin beslissingen van overheden of bedrijven jegens burgers steeds meer gebaseerd worden op informatie uit datasystemen waarop die burger geen invloed heeft, noch – of nauwelijks – op de gegevens in de systemen.

Conclusie: datakwaliteit is een lastig en onderschat onderwerp, waaraan meer aandacht zou moeten worden besteed. Zo dient voor elk significant systeem bepaald te worden welke eisen aan de data dienen te worden gesteld en vervolgens moet de datakwaliteit daarvan worden beoordeeld. Ten slotte dient het probleem te worden opgepakt door de vervuilde data te schonen en de bronoorzaken van de vervuiling aan te pakken. Er bestaan daartoe verschillende kwaliteitsmodellen, waarvan er één is uitgelegd, namelijk het model Total Data Quality Management.