

Process mining: data analytics voor de accountant die wil weten hoe het nu echt zit

Wil van der Aalst en Angelique Koopman

SAMENVATTING Steeds meer gebeurtenissen (“events”) worden geregistreerd en opgeslagen in IT-systemen. Op dit moment staat “Big Data” volop in de schijnwerpers en denken we vaak aan bedrijven als Google en Facebook. Event data zijn echter in elke organisatie te vinden en op elk niveau. Process mining is de verbindende schakel tussen data en proces. Dankzij process mining is het mogelijk tegelijkertijd prestatie-georiënteerde en compliance-georiënteerde vragen te stellen. Door procesmodellen te koppelen aan event data kunnen knelpunten opgespoord worden en is precies te zien waar en waarom mensen afwijken van het normatieve proces. Dit artikel beschrijft twee basisvormen van process mining: ‘process discovery’ en ‘conformance/compliance checking’.

RELEVANTIE VOOR DE PRAKTIJK Met de toepassing van process mining is het mogelijk om data in de context van onder meer bedrijfsprocessen te analyseren. Hiermee kan de ‘black box’ van gegevensverwerkende processen, inclusief de interne beheersing van organisaties inzichtelijk gemaakt worden. Vervolgens kunnen gedetailleerde analyses en toetsing aan normen worden uitgevoerd. Dit is niet alleen interessant voor organisaties zelf, juist ook voor interne en externe accountants is process mining interessant bij het uitvoeren van audits. Goed controleren gaat onder meer over het stellen van de juiste vragen. Process mining is toepasbaar in de verschillende fasen van het controleproces en biedt op een unieke manier inzicht in wat er echt gebeurt in een organisatie. Om de juiste conclusies te trekken is het echter nodig te weten hoe process mining werkt en hoe het op de juiste wijze toegepast kan worden.

1 Inleiding

Process mining is een opkomende discipline waarbij een uitgebreide set aan tools en technieken beschikbaar is om in uiteenlopende situaties op basis van feiten, vastgelegd in IT-systemen, bedrijfsprocessen te analyseren (Van der Aalst, 2013). Op de markt zijn di-

verse commerciële tools voor process mining beschikbaar. Deze tools zijn vanwege de gebruiksvriendelijkheid ook geschikt om gebruikt te worden bij kleine en middelgrote organisaties. Ook zijn er gereedschappen uit de academische wereld die vaak meer geavanceerde analyses bieden maar minder eenvoudig te bedienen zijn. Veel onderzoek is en wordt gedaan naar de toepassingsmogelijkheden van process mining. Hierover zijn in de afgelopen jaren veel (wetenschappelijke) publicaties beschikbaar gekomen. Een juiste toepassing van process mining is belangrijk om te waarborgen dat het van toegevoegde waarde is en voldoende kwaliteit heeft als bron voor besluitvorming of bij het geven van assurance. Kennis over en ervaring met de beschikbare tools en technieken voor process mining en in welke situatie deze toe te passen, is hierbij belangrijk.

Dit artikel is geschreven voor een brede doelgroep, waaronder interne en externe accountants, want we zijn van mening dat zeker ook voor accountants de toepassing van process mining toegevoegde waarde kan hebben. Accountantscontrole gaat over het geven van zekerheid. Controleren is een proces voor het verzamelen van controle-informatie ter onderbouwing van die zekerheid. De maatschappij verwacht dat de accountant gemaakte fouten in de jaarrekening, en wellicht ook steeds meer fraude, signaleert. Ondanks het feit dat het opsporen van fraude niet de primaire doelstelling is van de accountant bij het controleren van een financiële verantwoording. Van de accountant wordt verwacht dat hij/zij professioneel kritisch is en risico's signaleert (NV Controlestandaard 200.7). In het proces van controleren van bijvoorbeeld een jaarrekening stellen accountants zichzelf vragen als: ‘wat kan er fout gaan’ en ‘wat moet ik weten om in staat te zijn, te concluderen dat iets goed is’. Zoals Leslie, Aldersley, Cockburn & Reiter in 1986 al hebben benoemd, betekent goed controleren nog steeds het stellen van de juiste vragen; daar is niets in veranderd. In de hedendaagse informatiesamenleving met een overvloed aan data

kunnen data-science-technieken (zoals process mining) de accountant ondersteunen bij het stellen van de juiste vragen en het beantwoorden ervan.

Swart, Wille en Majoor (2013) stellen dat data analytics de accountant bij verschillende onderdelen van de controle kunnen ondersteunen, zoals bij ‘understanding the business’, testen van de interne beheersing, uitvoeren van cijferanalyses en detailcontroles. Vanuit een groeiende maatschappelijke behoefte beschrijven zij een model (gebaseerd op Bayesiaanse statistiek) dat de mogelijkheid geeft om zekerheid te kwantificeren waardoor de controle effectiever en efficiënter uitgevoerd kan worden. Data analytics wordt in deze context gedefinieerd als een “kwantitatieve analyse van de betrouwbaarheid van gegevensverzamelingen”. Process mining biedt ook de mogelijkheid om kwantificeerbare zekerheid te geven in de context van processen in een organisatie en de afhandeling van transacties daarbij. Een concreet voorbeeld in deze context is het signaleren dat bij het verwerken van factureren in de realiteit ook wel eens een controle niet plaatsvindt, terwijl dit volgens de opzet van de interne procedures wel altijd het geval zou moeten zijn. Process mining maakt het mogelijk snel zo’n afwijking te vinden.

Jans, Alles en Vasarhelyi (2013) identificeren diverse vormen van toegevoegde waarde voor interne en externe accountants bij de toepassing van process mining. Zij concluderen bijvoorbeeld dat process mining voor de accountant kan bijdragen aan een meer effectieve implementatie van het audit risk model doordat het gebruikt kan worden bij het uitvoeren van walkthroughs en analyses. Hierbij stellen zij dat met process mining de accountant analyses kan uitvoeren die met andere audit tools niet mogelijk zijn, zoals het ontdekken van de manier waarop bedrijfsprocessen werkelijk uitgevoerd zijn en het analyseren van de wijze waarop in de organisatie mensen met elkaar samenwerken en werkzaamheden aan elkaar overdragen.

In dit artikel beschrijven we de twee basisvormen van process mining: ‘process discovery’ en ‘conformance/compliance checking’ aan de hand van enkele eenvoudige voorbeelden. Om process mining in de praktijk toe te passen is het nodig om te weten hoe het werkt, welke tools en technieken in welke situatie nodig zijn en wat er bij komt kijken om de kwaliteit van de uitkomsten te waarborgen.

Dit artikel is als volgt ingedeeld. In paragraaf 2 plaatsen we process mining in de context van data analytics. Hierbij leggen we uit wat process mining is, inclusief een beschrijving van de tools die hiervoor beschikbaar zijn. Deze paragraaf sluiten we af met een betoog over de toepassingsmogelijkheden van process mining voor

accountants. Paragraaf 3 gaat over technieken om allerlei vormen van vastgelegde gebeurtenissen (‘events’) in IT-systemen te visualiseren en te onderzoeken; dit wordt in process mining-terminologie ‘proces discovery’ genoemd. In deze paragraaf besteden we ook aandacht aan de voorbereidingen bij de toepassing van process mining en hoe de kwaliteit van de visualisaties ofwel procesmodellen kan worden onderzocht. Hierbij geven we overwegingen bij de keuze voor de te gebruiken tools. Een tweede basisvorm van process mining noemen we ‘conformance/compliance checking’. Dit beschrijven we in paragraaf 4. Hierbij geven we eveneens overwegingen voor het gebruik van tools hiervoor en benoemen we de mogelijkheden van continuous auditing met de toepassing van process mining. In paragraaf 5 geven we een beschouwing van mogelijkheden voor verder onderzoek en meer geavanceerde toepassingen van process mining. Tenslotte sluiten we in paragraaf 6 af met een samenvatting van dit artikel.

2 Process mining in de context van data analytics

2.1 Process mining als verbindende schakel tussen data en proces

Informatiesystemen ter ondersteuning van bedrijfsprocessen zijn in de hedendaagse maatschappij niet meer weg te denken. Proces en systeem zijn steeds meer met elkaar verweven. Steeds meer gebeurtenissen (‘events’) worden geregistreerd en opgeslagen in IT-systemen en dit gebeurt op allerlei niveaus (Van der Aalst, 2013). Event data zijn in vrijwel elke organisatie te vinden. Denk bijvoorbeeld aan een ziekenhuisinformatiesysteem. Alle afspraken, onderzoeken, onderzoeksresultaten (bloedproeven, röntgenfoto’s, enz.), en betalingen behorende bij een patiënt worden vastgelegd. Dit biedt de mogelijkheid om processen op een meer data-gedreven manier te analyseren.

Alleen organisaties die slim gebruik maken van de stortvloed aan gegevens over het gebruik van producten en diensten zullen overleven. Organisaties hebben medewerkers nodig die weten hoe ze gegevens kunnen omzetten in informatie ofwel ‘waarde’. Hiervoor is een nieuwe beroepsgroep nodig: de data scientist (Van der Aalst, 2014). Zoals aangegeven in figuur 1 (gebaseerd op Van der Aalst, 2014) moet een data scientist¹ voldoende kennis van statistiek, data mining, process mining, visualisatie, databases, algoritmiek, en gedistribueerde systemen (denk aan Hadoop) hebben om data om te zetten in nieuwe inzichten, voorspellingen en aanbevelingen. Het is echter niet voldoende om alleen technische vaardigheden te hebben. Sociologie, psychologie, bedrijfskunde en domeinkennis spelen een belangrijke rol. Voor accountants, managers en analisten gaat het vooral om de processen. Daarom is ken-

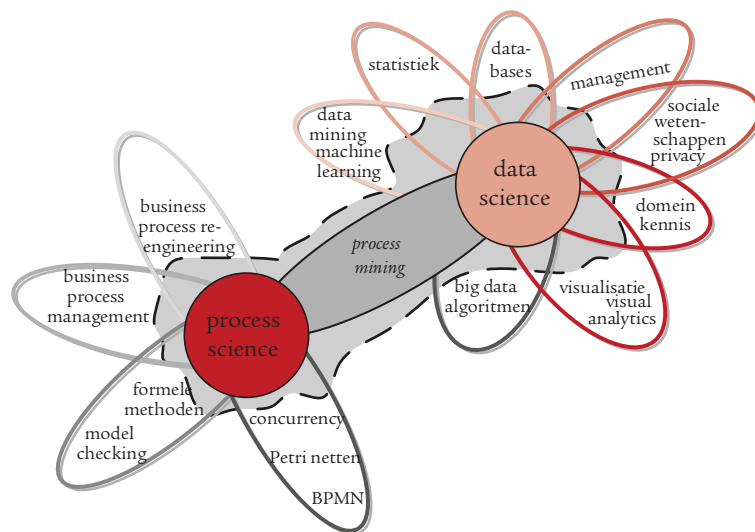
nis van process science essentieel (denk aan procesmodelleren en business process management). Figuur 1 laat zien dat process mining de verbindende schakel is tussen data science en de klassieke procesgeoriënteerde methoden.

Traditioneel is er een ont koppeling tussen procesmodel-gebaseerde analyse (simulatie, verificatie, optimalisatie, gaming, etc.) en data-gebaseerde analyse (data mining, machine learning, business intelligence). Dit maakt dat processen niet geanalyseerd worden op basis van feiten en dat alom aanwezige gegevens niet gebruikt worden om de echte processen zichtbaar te maken (Van der Aalst, 2013). Figuur 2 laat zien dat met de toepassing van process mining het mogelijk is tegelijkertijd prestatie-georiënteerde en compliance-georiënteerde vragen te stellen op basis van feiten. Doordat procesmodellen gekoppeld worden aan event data kunnen knelpunten opgespoord worden en is precies te zien waar mensen afwijken van het normatieve proces en kan onderzocht worden waarom dit is. Ofwel process mining is het gereedschap voor prestatie- en compliance-analyses.

2.2 Tools voor process mining

Voor het toepassen van process mining zijn veel technieken beschikbaar, waaronder tools om bijvoorbeeld procesmodellen te genereren, werkelijke gebeurtenissen te vergelijken met een norm of normatief procesmodel, knelpunten in de afhandeling van transacties te analyseren, om performance-issues te voorspellen, etc. Deze tools maken - eenvoudig gezegd - onder de motorkap gebruik van algoritmes om verbanden tussen gegevens inzichtelijk te maken. Voor de te gebruiken tools maken we voor het gemak onderscheid tussen commercieel ontwikkelde tools met diverse functionaliteiten en eenvoudig in gebruik en tools die zijn ontwikkeld in de academische wereld. Deze laatste tools maken onderdeel uit van de 'pluggable' open source ProM toolset, die bestaat uit een verzameling van functionaliteiten voor diverse doeleinden gericht op het toepassen van process mining. Alle functionaliteiten voor deze toolset zijn ontwikkeld via de TU Eindhoven en zijn allemaal uitvoerig beschreven in wetenschappelijke publicaties (zie www.processmining.org voor een selectie van de vele publicaties over process mining en ProM). ProM blijft in ontwikkeling; wetenschappers zijn continu bezig om nieuwe functionaliteiten (mogelijkheden voor process mining gezien een specifieke vraagstelling) te ontwikkelen en bestaande functionaliteit te verbeteren. Voorbeelden van de commercieel ontwikkelde tools zijn: Perceptive Process Mining, Disco en Celonis. Deze tools zijn gebruiksvriendelijk, intuïtief (zonder kennis van de technieken toch snel te gebruiken) en combineren meerdere functionaliteiten voor de toepassing van process mining en hebben uitgebreide filtermogelijkheden voor het analyseren

Figuur 1 Process mining als de verbindende schakel tussen data science en process science ²

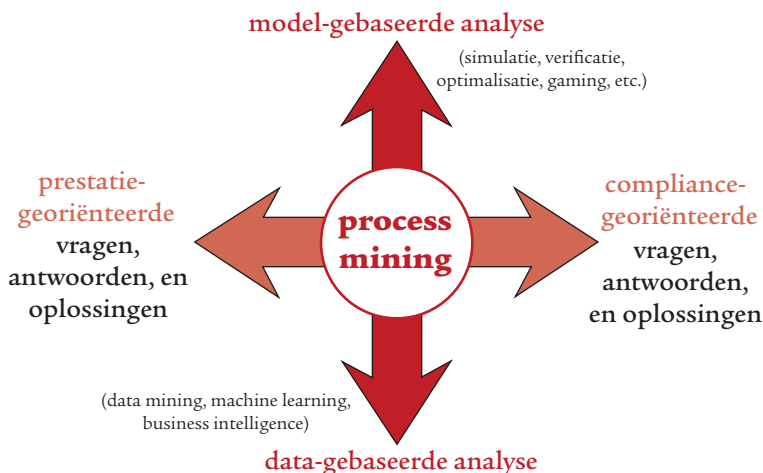


van gebeurtenissen. Een aandachtspunt bij deze tools is dat vanuit het 'geminde'³ procesmodel niet in alle gevallen de afwijkingen in het proces en de kwaliteit van het model goed kunnen worden onderzocht. Hiervoor is wel een oplossing beschikbaar. Daarover later in dit artikel meer.

2.3 Waarom is process mining interessant voor accountants?

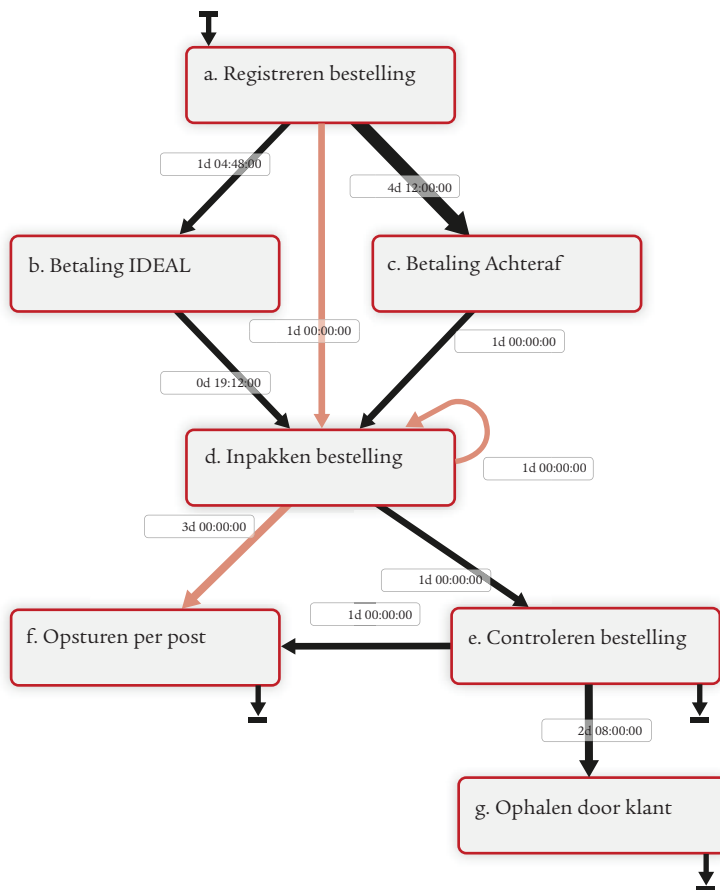
In de praktijk van de accountant is het inmiddels niet meer ongebruikelijk om met behulp van data-analyse-tools (hieronder vatten we voor het gemak ook spreadsheet-toepassingen) opvallende mutaties in het grootboek te selecteren voor nader onderzoek. Dit beeld is gebaseerd op waarnemingen tijdens trainingen aan accountants door een van de auteurs (Koopman) op het gebied van data-analyse en process mining. De cursis-

Figuur 2 Process mining als gereedschap voor prestatie- en compliance-analyses (Van der Aalst & Van Geffen, 2014)



ten tijdens deze trainingen zijn externe en interne accountants en zijn werkzaam bij kleine, middelgrote en grote accountantskantoren. Bij organisaties waarbij de primaire processen in belangrijke mate door ICT worden ondersteund zal de accountant zelfs verder gaan en meer 'advanced' data-analyse toepassen waarbij mogelijk gegevens uit verschillende IT-systemen worden gecombineerd. Bijvoorbeeld, voor het controleren van een geld- en goederenbeweging. Hierbij wordt uitgegaan van vooraf gedefinieerde risico-indicatoren. De praktijk leert dat deze werkwijze kan resulteren in een grote hoeveelheid individuele mutaties die nader onderzocht moeten worden. Mogelijk betreffen deze gesignaleerde mutaties uiteindelijk geen afwijkingen (de 'false positives') of kan het zijn dat afwijkingen niet worden gesignaleerd. Van dit laatste kan sprake zijn als een combinatie van individuele mutaties een signaal kan zijn voor een mogelijke afwijking. Rijnders, Eijken, Fissette en Van Schijndel (2013) stellen in dit kader dat bijvoorbeeld voor het analyseren van grootboekmutaties juist een combinatie van mutaties interessant is, omdat deze mutaties het resultaat zijn van financiële bedrijfsprocessen. Zij merken terecht op dat

Figuur 3 Een eenvoudig procesmodel ontdekt op basis van event data met behulp van een commercieel process mining product (Perceptive Process Mining)



ieder bedrijfsproces leidt tot financiële stromen (denk aan inkoop, verkoop) met als resultaat een aantal opeenvolgende mutaties in het grootboek. Accountants hanteren voor elk type proces met bijbehorende specifieke financiële stroom een norm en mutaties vallen op zodra ze niet in lijn zijn met de vastgestelde norm (Rijnders et al., 2013). Met de toepassing van process mining heeft de accountant een instrument in handen om gegevens in de context van een proces op een visuele manier te analyseren. Of anders gezegd: het verloop van (financiële) stromen die ten grondslag liggen aan bijvoorbeeld mutaties in het grootboek te analyseren en te vergelijken met een norm.

De accountant kan process mining bijvoorbeeld toepassen in de planningsfase van de controle om inzicht te krijgen in het werkelijke verloop van bedrijfsprocessen, inclusief bepaalde proces-gerelateerde aspecten van interne beheersing. Het draagt bij aan 'understanding the business', het uitvoeren van de risico-analyse en het maken van keuzes bij het bepalen van de controle-aanpak. Dit laatste betreft bijvoorbeeld de keuze om op basis van een deelwaarneming/steekproef of met process mining (voor alle transacties in de te controleren periode) vast te stellen of een interne beheersmaatregel heeft gewerkt. Of deze afweging samenhangt met de keuze voor primair systeemgericht ofwel primair gegevensgericht controleren valt buiten de scope van dit artikel. Vanzelfsprekend is dit wel degelijk een interessante vraagstelling voor verder onderzoek. In de hierna volgende paragrafen richten we ons op twee varianten voor de toepassing van process mining, te beginnen bij 'proces discovery'.

3 Process Discovery

3.1 Wat is process discovery?

Met technieken voor 'process discovery' kunnen we op basis van gebeurtenissen (events) in een event-log het verloop van deze gebeurtenissen visualiseren, onderzoeken en voorstellen voor verbetering doen. Hierbij analyseren we dan bijvoorbeeld alle gebeurtenissen rondom de afhandeling van bestellingen, een inkooporder, een klacht of de behandeling van een patiënt. De betreffende bestellingen, inkooporders, klachten of patiënten noemen we 'cases'. In figuur 3 presenteren we een eenvoudig voorbeeld van zo'n 'gemiddeld' procesmodel.

Stel we hebben een proces voor het afhandelen van bestellingen, voorgesteld in figuur 3. De activiteiten worden hier voor de eenvoud voorgesteld door letters. Activiteit **a** is bijvoorbeeld het registreren van de binnenkomende bestelling. Dit is een voorbeeld waarbij de gebeurtenissen volgtijdelijk, verbonden door pijlen worden weergegeven.

Bij het minen van processen kunnen we achteraf het procesmodel visualiseren en evalueren op basis van volledig afgehandelde cases. Process mining kan ook gebruikt worden om vooruit te kijken. Hierbij kan bijvoorbeeld tijdens de uitvoering van een proces (bijvoorbeeld bij continuous monitoring) op basis van gebeurtenissen die al hebben plaatsgevonden, het verdere verloop van de desbetreffende cases voorspeld worden. Vergelijk dit met een navigatiesysteem in een auto waarbij de aankomsttijd zich gedurende de rit aanpast aan de gereden route ten opzichte van het eindpunt en mogelijke filemeldingen.

3.2 Voorbereiden process mining

Voor alle vormen van process mining geldt dat we beginnen met het verzamelen van de event-data waarbij bepaald moet worden op welke wijze de juistheid en volledigheid van de te analyseren events gevalideerd kan worden. Minstens zo belangrijk is om te bepalen welke event-data nodig zijn om bij het ‘minen’ van het proces tot de gewenste resultaten te kunnen komen en analyses op het gewenste detailniveau uit te voeren. Vooraf een heldere en concrete doelstelling bepalen is dus belangrijk voor data-extractie.

Het ‘minen’ van een proces start met het vaststellen dat de event-log een compleet beeld geeft van het te onderzoeken proces in de gewenste tijdsperiode. Denk hierbij aan het maken van aansluitingen als het totaal van de factuurwaarde volgens de event-log met het grootboek en het vaststellen dat alle relevante order- of factuurnummers volgens het bronsysteem aanwezig zijn in de event-log. Vastgesteld (belangrijk voor accountants: en gedocumenteerd!) moet worden dat bij het zogenaamde ‘platslaan’ van de data alle vereiste gegevens op de juiste wijze in de event-log zijn gecombineerd. In de praktijk zien we dat de event-data veelal onttrokken wordt aan een veelheid van tabellen in relationele databases die via sleutels aan elkaar zijn gekoppeld. Inzicht in het datamodel is hierbij onontbeerlijk. Voor een data-analist is dit vanzelfsprekend bekend terrein. Overigens faciliteren moderne informatiesystemen (denk aan BPM-systemen) het loggen van events, zodat het voorbereiden van een event-log relatief eenvoudig is.

Een ander aspect om de betrouwbaarheid van de gebruikte data voor de event-log te beoordelen is bijvoorbeeld de bevestiging dat elke mutatie in het IT-systeem herleidbaar is tot het juiste individu. Accountants besteden hier aandacht aan als onderdeel bij het evalueren van de IT General controls (denk aan: logische toegangsbeveiliging).

Bij process mining kunnen we vanuit verschillende invalshoeken naar een proces kijken (Van der Aalst, Hee, Van der Werf, Kumar & Verdonk, 2011), bijvoorbeeld

hoe de afhandeling van transacties volgtijdelijk in de tijd heeft plaatsgevonden (het ‘Hoe’), wie heeft wanneer wat gedaan (het ‘Wie’), hoe vindt de overdracht tussen verschillende personen in het proces plaats? Hierbij kan ook als invalshoek de kenmerken van de onderzochte transacties genomen worden en welke business rules van toepassing zijn (het ‘Wat’). Business rules zijn feitelijk randvoorwaarden binnen het proces waarbij verschillende varianten mogelijk zijn (Van der Aalst et al., 2011). Denk hierbij aan bestellingen boven een bepaald bedrag, van een specifieke klant of groep klanten, etc., die een extra handeling in het proces vragen of de toepassing van het vier-ogen principe.

3.3 Kwaliteit procesmodellen onderzoeken

Bij het toepassen van ‘process discovery’ is het zaak na te gaan of het automatisch gegenereerde procesmodel geschikt is voor het doel waarvoor het gebruikt wordt. Op basis van dezelfde event-log kunnen verschillende procesmodellen gegenereerd worden (Van der Aalst, 2011). Met een te complex procesmodel, is het lastig om gestructureerd onderzoek te doen en overzicht te krijgen. Aan de andere kant zou een te eenvoudige weergave van het proces interessante afwijkingen kunnen verbergen. Deze kunnen juist vanuit een audit-perspectief interessant zijn. Process mining tools hebben functionaliteiten om bijvoorbeeld aan te geven op basis van welk deel van de te onderzoeken cases een procesmodel moet worden gemaakt of om met behulp van filters activiteiten wel of niet op te nemen in het procesmodel. Ook kunnen filters geplaatst worden op basis van kenmerken van cases. Belangrijk is dus om te weten en te documenteren welke ‘cases’ wel en niet in het gegenereerde procesmodel passen en dat van een procesmodel wordt uitgegaan dat geschikt is voor het doel van de analyse (welke activiteiten zijn wel en niet relevant). Kortom, de kwaliteit van het ‘geminede’ procesmodel moet worden onderzocht om te waarborgen dat het ‘geminede’ procesmodel en de event-log goed op elkaar aansluiten (Rozinat, Alves de Medeiros, Gunther, Weijters & Van der Aalst, 2007).

Er zijn vier dimensies waarop de kwaliteit van procesmodellen beoordeeld kan worden: (1) fitness, (2) generalization, (3) precision en (4) simplicity (Van der Aalst, Adriansyah & Van Dongen, 2012). Als een model een goede ‘fitness’ heeft betekent dit dat alle cases door het geminede ‘procesmodel’ beschreven worden. De geobserveerde paden zijn ook mogelijk in het model. Een model is precies als er niet teveel paden mogelijk zijn die op basis van de event-log onwaarschijnlijk zijn. Daartegenover staat dat een model ook moet kunnen generaliseren. Indien een model alleen laat zien wat er in het verleden is gebeurd, is de voorspellende waarde beperkt. Het model moet van de werkelijkheid “leren” en niet deze domweg reproduceren. De laatste dimen-

sie (simplicity) geeft aan dat we streven naar een zo eenvoudig mogelijke verklaring voor het geobserveerde gedrag. Er zal een juiste balans gevonden moeten worden tussen deze vier kwaliteitsdimensies en dat is niet altijd eenvoudig.

Het ‘geminede’ procesmodel is slechts het startpunt voor analyse. Door de koppeling tussen data en procesmodel kunnen de eerder genoemde prestatie- en compliance-gerelateerde vragen beantwoord worden.

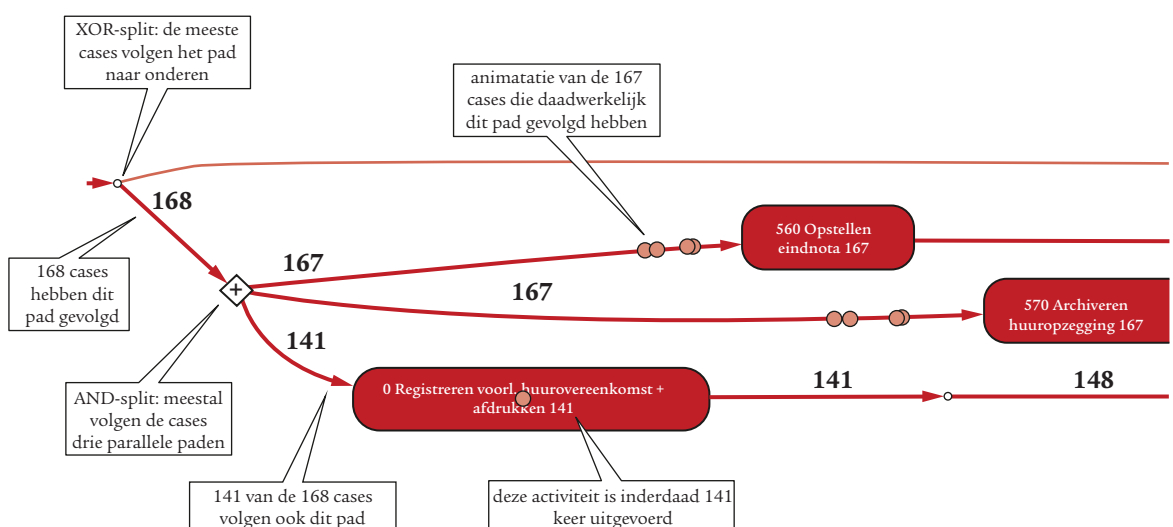
3.4 Overwegingen bij keuze tools voor waarborgen kwaliteit procesmodel

Leemans, Fahland en Van der Aalst (2015) hebben specifiek voor het exploreren van processen verschillende tools (commercieel en open-source) met elkaar vergeleken. Naast de eerder beschreven kwaliteitsdimensies van een procesmodel (fitness, precision, generalization en simplicity) is gekeken naar de mogelijkheden om in en uit te zoomen, het weglaten van minder frequente paden, snelheid, en gebruiksgemak. Commerciële tools zijn vaak eenvoudiger in gebruik, maken het mogelijk snel een model te maken, en bieden animatiemogelijkheden en uitgebreide filteropties. De kwaliteit van de modellen laat echter vaak te wensen over. Commerciële tools maken vooral “plaatjes” en geen procesmodellen met een eenduidige interpretatie en interne consistentie. Hierdoor is het onmogelijk de vier kwaliteitsdimensies te kwantificeren. Indien we eenduidige semantiek proberen te verbinden met de plaatjes, komen we tot de ontdekking dat bijna alle modellen deadlocks (cases komen vast te zitten midden in het proces) of andere problemen hebben. Ook zijn de gemaakte afwegingen tussen fitness, precision, generalization en

simplicity niet zichtbaar waardoor foutieve interpretaties mogelijk zijn. In figuur 4 is bijvoorbeeld te zien dat de weergegeven aantallen niet op elkaar aansluiten omdat niet alle cases in de desbetreffende eventueel perfect passen in het weergegeven procesmodel. Volgens het procesmodel zouden alle drie activiteiten voor alle 168 cases uitgevoerd moeten worden. Dit is echter niet het geval. De activiteiten “Opstellen eindnota” en “Archiveren huuropzegging” zijn 167 keer uitgevoerd. De activiteit “Registreren voorl. Huurovereenkomst” is slechts 141 keer uitgevoerd. Dit voorbeeld laat zien dat het niet voldoende is een ‘mooi plaatje’ te maken. De koppeling met de onderliggende feiten laat zien waar het model knelt en de werkelijkheid weerbarstig is. Dit is essentieel voor compliance-vraagstukken.

De Inductive visual Miner (IvM) die gebruikt is om het model in figuur 4 te creëren geeft wel garanties, dit in tegenstelling tot commerciële tools. IvM ontdekt altijd een model dat intern consistent is. Indien gewenst kan ook een model met 100% fitness ontdekt worden. Vanwege de gemaakte afwegingen tussen de vier kwaliteitsdimensies is dit echter vaak niet het gewenste model (te complex of te generiek). Zoals figuur 4 laat zien is het echter nog steeds mogelijk betrouwbare uitspraken te doen over het onderliggende proces. Ondanks de garanties ten aanzien van consistentie en correctheid, is IvM qua snelheid en gebruiksgemak vergelijkbaar met commerciële tools. Ook is het mogelijk in en uit te zoomen, processen te animeren, en knelpunten zichtbaar maken. Het gebruik van ProM's IvM in combinatie met een commercieel gereedschap is zeker bij het ‘minen’ van complexe processen aan te bevelen.

Figuur 4 Voorbeeld van een klein fragment van een veel groter procesmodel ontdekt met de Inductive visual Miner in ProM



Nadat de gebeurtenissen in de event-log zijn ‘gemined’ en van het procesmodel is vastgesteld dat de kwaliteitscriteria in de gewenste balans zijn, kan het procesmodel gebruikt worden voor het toetsen aan normen. We noemen dit ‘conformance/compliance checking’ waarbij het procesmodel an sich kan worden vergeleken met een normatief procesmodel en waarbij afwijkende paden in het proces kunnen worden geïdentificeerd om nader te onderzoeken. Ook kan worden getoetst of de gebeurtenissen in het ‘geminede’ procesmodel voldoen aan zogenaamde ‘business rules’. Denk hierbij aan de vraagstelling of bijvoorbeeld alle bestellingen boven een bepaald bedrag pas uitgeleverd worden nadat een extra controle op kredietwaardigheid van de afnemer heeft plaatsgevonden.

4 Conformance/compliance checking

4.1 Wat is conformance/compliance checking?

Process mining-technieken kunnen ingezet worden om processen te ontdekken (discovery). Voor audit-doel-einden (bijvoorbeeld het evalueren van de interne beheersing) is het echter minstens zo belangrijk afwijkingen tussen een normatief procesmodel en de werkelijkheid vast te stellen. Gelukkig is het mogelijk event-logs af te spelen op procesmodellen. Hierdoor kunnen afwijkingen snel opgespoord worden en kan er een diagnose uitgevoerd worden om te achterhalen wanneer en waarom die afwijkingen plaatsgevonden hebben. We noemen dit “conformance/compliance checking”.

Het idee achter conformance/compliance checking is het volgende. Stel we nemen hetzelfde proces uit de vorige paragraaf voor het afhandelen van bestellingen, voorgesteld in figuur 5. De activiteiten worden hier voor de eenvoud voorgesteld door letters. Activiteit **a** is bijvoorbeeld het registreren van de binnenkomende bestelling. Na activiteit **a** is er een keu-

ze tussen **b** of **c** en wordt altijd **d** gedaan. Nadat **b** of **c** en **d** zijn gedaan, kan activiteit **e** plaatsvinden, gevolgd door **f** of **g**. Mogelijke uitvoeringen van dit proces zijn **a,b,d,e,f** en **a,d,c,e,g**. Stel we zien een bestelling in de volgende sequentie van activiteiten: **a,c,e,f**. Door middel van conformance checking is eenvoudig vast te stellen dat activiteit **d** ontbreekt. Dit doen we door het opstellen van een zogenaamde “alignment”:

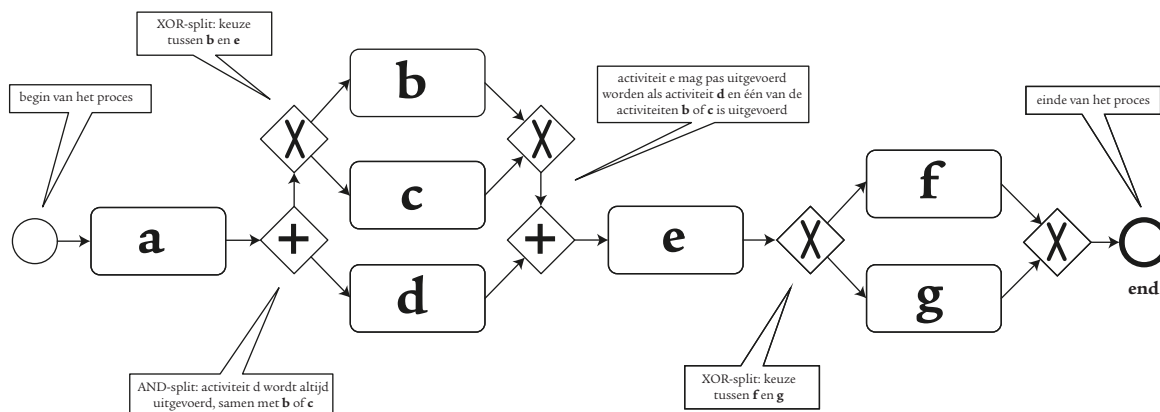
a	c	»	e	f
a	c	d	e	f

De bovenste rij in de alignment (**a,c,e,f**) laat de daadwerkelijk afhandeling van de bestelling zien en de onderste rij laat een pad in het procesmodel (**a,c,d,e,f**) zien dat het dichtste bij de daadwerkelijke afhandeling ligt. Het symbool » heeft aan waar het probleem zit: activiteit **d** had uitgevoerd moeten worden maar is niet uitgevoerd. Stel we zien voor een andere bestelling de volgende sequentie van activiteiten: **a,d,d,e,f**. Een mogelijke alignment is:

a	d	d	»	e	f
a	d	»	c	e	f

Deze alignment laat zien dat activiteit **d** niet twee keer uitgevoerd had mogen worden en dat activiteit **c** ontbreekt. Er kunnen meerdere alignments zijn. In het laatste voorbeeld hadden we ook activiteit **b** kunnen kiezen. Immers activiteit **b** of **c** moet uitgevoerd worden. Dankzij conformance/compliance checking kunnen we dus precies zien waar mogelijke afwijkingen zitten en kunnen automatisch diagnoses uitgevoerd worden. Bijvoorbeeld: is het altijd dezelfde medewer-

Figuur 5 Een simpel fictief procesmodel in BPMN-notatie



ker die activiteit **d** dubbel uitvoert? De interpretatie van deze diagnoses vindt plaats door degene die de process mining uitvoert, dus handmatig.

Activiteiten worden hier voorgesteld door een letter, maar ze verwijzen naar gebeurtenissen die plaatsgevonden hebben op een bepaald tijdstip en wellicht ook naar de persoon die de activiteit uitvoerde of naar informatie over de klant. Conformance checking is immers veel breder dan wat het bovenstaande doet vermoeden. Alignments kunnen ook gebruikt worden om aan te geven dat een activiteit te laat is uitgevoerd of door de verkeerde persoon (bijvoorbeeld een overtreding van het vier-ogen principe).

Dankzij process mining is het dus mogelijk automatisch modellen af te leiden en voortdurend de modellen te koppelen aan de werkelijkheid. Onvolkomenheden in modellen (patronen) kunnen nu genadeloos blootgelegd worden door de confrontatie met de harde data (gebaseerd op de werkelijkheid) aan te gaan. Een gedetailleerde vergelijking van 'Soll' (het normatieve procesmodel) en 'Ist' (het daadwerkelijke proces) is nu mogelijk. Dit is een waardevolle toepassing voor interne en externe audit-doeleinden, zeker in omgevingen waarbij de primaire bedrijfsprocessen een hoge mate van automatisering kennen. Het is bijvoorbeeld in de praktijk denkbaar dat in een situatie waarbij gebruikers te ruime bevoegdheden hebben vanuit het perspectief van interne beheersing door de accountant process mining ingezet kan worden om vast te stellen dat gebruikers wel of niet gebruik hebben gemaakt van deze te ruime bevoegdheden.

4.2 Overwegingen bij het gebruik van tools

Commerciële process mining tools bieden vaak uitgebreide filtermogelijkheden om de cases in de eventlog te toetsen aan business rules ofwel normen. Dit is voor overzichtelijke processen met stap voor stap de evaluatie van individuele afwijkingen van de norm een prima alternatief. Voor de evaluatie van een uitgebreide set met elkaar samenhangende business rules en complexe processen zijn de diverse compliance plug-ins in ProM aan te bevelen. Met bijvoorbeeld een tool als het 'Compliance dashboard' kan conformance checking plaatsvinden waar normen als het ware geprogrammeerd worden in de tool. Het resultaat bestaat uit statistische informatie, grafieken en een overzicht van de gesignaleerde afwijkingen. Vervolgens kan met een andere plug-in van ProM 'Get problem insight' een gedetailleerde analyse van de gesignaleerde afwijkingen worden uitgevoerd. Correlaties tussen kenmerken van cases worden gebruikt om inzicht in compliance te krijgen. Voor het waarborgen van betrouwbare resultaten is dus een zorgvuldige afweging van de te gebruiken tooling en deskundig gebruik daarvan belangrijk.

4.3 Process mining en continuous auditing

Van der Aalst et al. (2011) presenteerden een conceptueel model voor de toepassing van een online auditing tool (OLAT) van bedrijfsprocessen waarmee accountants alle voor hen relevante gebeurtenissen in bedrijfsprocessen van een cliënt kunnen inlezen en vervolgens met onder meer process mining-technieken kunnen controleren. Het gaat hierbij om de toetsing van gedefinieerde business rules; feitelijk een toepassing van continuous auditing waarbij real-time schending van business rules inzichtelijk wordt gemaakt (Chan & Vasarhelyi, 2011). Dit sluit aan bij de stelling van Chan en Vasarhelyi (2011) dat de traditionele audit gedateerd is in een real-time economie. De ontwikkelingen van process mining-technieken en de exponentieel groeiende hoeveelheid event data bij organisaties brengen de toepassing van continuous auditing in de praktijk steeds dichterbij. De stip aan de horizon lijkt misschien nog ver weg voor accountants. Echter, als we kijken naar ontwikkelingen op het gebied van big data-toepassingen is het hoogstwaarschijnlijk een kwestie van afwachten totdat vanuit de markt de vraag naar deze vorm van assurance komt.

5 What's next?

In dit artikel hebben we beschreven hoe de basisvormen 'process discovery' en 'conformance checking' van process mining werken en wat er komt kijken bij de toepassing ervan in de praktijk. De huidige toepassingmogelijkheden van process mining zijn divers en er wordt veel onderzoek gedaan naar nieuwe mogelijkheden. Het onderzoek van een van de auteurs (Koopman) richt zich specifiek op toepassing van process mining vanuit het perspectief van interne beheersing. Denk hierbij ook aan de mogelijkheden van process mining over de grenzen van een individueel proces; het combineren van de geld- en goederenbeweging, ofwel standen met kwantitatieve informatie en stromen (hoe deze informatie tot stand is gekomen). Concreet gaat dit over het vergelijken van het normatieve procesmodel à la Starreveld of andere sector-gerelateerde normatieve procesmodellen en het werkelijke proces gebaseerd op event data. Hiermee kan de effectiviteit van de opzet van processen worden beoordeeld, het bestaan vastgesteld en de werking van interne beheersmaatregelen over een periode worden geëvalueerd.

In het verlengde daarvan ligt de vraag hoe en op welke wijze process mining kan ondersteunen om zogenaamde 'soft risks' (die samenhangen met gedrag en cultuur in een organisatie) te identificeren. Ofwel, de concrete handvatten die process mining de accountant kan geven bij het evalueren van de interne beheersomgeving, die immers het fundament zijn voor een deugdelijke interne beheersing.

Process mining zal zich ook verder blijven ontwikkelen als wetenschapsgebied. Nieuwe technieken zijn nodig om event-logs met miljoenen events nog sneller te analyseren. Ook verschuift de aandacht van de analyse van historische data naar real-time-toepassingen (predictive analytics) en automatische procesverbeteringen.

6 Samenvatting

In dit artikel hebben we process mining gepositioneerd als verbindende schakel tussen op (proces) model-gebaseerde analyse en op data-gebaseerde analyse. Process mining gaat over het analyseren van event-data (gegevens over gebeurtenissen) in de context van processen. Met het gebruik van visualisatietechnieken zijn we in staat om patronen in data zichtbaar te maken die voor het menselijke brein niet zonder meer te onderscheiden zijn.

Process mining-technieken kunnen ingezet worden om processen in organisaties te ontdekken, te analyseren en te verbeteren. Ook voor accountants kan process mining een waardevol hulpmiddel zijn om inzicht in processen te krijgen, ofwel 'understanding the business', bij het uitvoeren van risicoanalyse en bij het maken van keuzes in de controleaanpak. Voor audit-doel-einden is het echter minstens zo belangrijk afwijkingen tussen een normatief procesmodel en de werkelijkheid vast te stellen. Gelukkig is het mogelijk event-logs (verzameling event-data voor een proces) af te spelen op (normatieve) procesmodellen. Hierdoor kunnen afwijkingen snel opgespoord worden en kan er een diagnose uitgevoerd worden om te achterhalen wanneer en waarom die afwijkingen plaatsgevonden hebben. We noemen dit "conformance checking". Vanuit het perspectief van de audit sluit dit aan bij de werkzaamheden voor het evalueren van de interne beheersing of het uitvoeren van gegevensgerichte werkzaamheden. Gezien de diversiteit aan beschikbare technieken en tools voor process mining is het belangrijk dat de gebruiker bij toepassing van process mining op gestructureerde wijze te werk gaat en een weloverwogen keuze maakt voor de te gebruiken gereedschappen. We hebben voor het gemak onderscheid gemaakt tussen commerciële tools en tools ontwikkeld in de academische wereld (veelal open-source), zoals ProM. De commerciële tools bieden uitgebreide functionaliteiten en filtermogelijkheden waarbij in gedachten gehouden moet worden dat een combinatie met een academische

tool nodig kan zijn om bijvoorbeeld de kwaliteit van het 'geminede' procesmodel te onderzoeken. Ook bij het toetsen van complexe business rules en onderzoek naar de oorzaak van afwijkingen kan een academische tool nodig zijn.

Het 'minen' van een proces start met het vaststellen dat de event-log een compleet beeld geeft van het te onderzoeken proces in de gewenste tijdsperiode en dat alle vereiste gegevens op de juiste wijze in de event-log zijn gecombineerd. Het 'geminede' procesmodel is het startpunt voor de echte analyse van het proces waarbij prestatie- en compliance-gerelateerde vragen beantwoord moeten gaan worden. Voor het waarborgen van betrouwbare resultaten op basis van process mining zal de kwaliteit van 'geminede' procesmodellen moeten worden beoordeeld. We hebben vier dimensies (fitness, generalization, precision en simplicity) beschreven waarop de kwaliteit van procesmodellen beoordeeld kunnen worden en het vinden van de juiste balans kan in de praktijk een uitdaging zijn.

Process mining-technieken kunnen ook gebruikt worden voor het creëren van simulatiemodellen op basis van werkelijke data. Hiermee kunnen 'What-if'-vragen worden beantwoord (Van der Aalst, 2015). De toegevoegde waarde voor het management is dat hiermee besluitvorming gebaseerd kan worden op basis van werkelijke data. Ook bij continuïteitsvraagstukken kan simulatie met behulp van process mining voor het management en ook de accountant een waardevolle rol hebben, juist door het gebruik van automatisch gegenereerde procesmodellen. ■

Prof. dr. ir. W.M.P. (Wil) van der Aalst (www.vdaalst.com) is universiteitshoogleraar aan de Technische Universiteit Eindhoven (TU/e) en wetenschappelijk directeur van het Data Science Center Eindhoven (DSC/e). Hij is ook trekker van de IEEE Task Force on Process Mining.

Drs A.J.M. (Angelique) Koopman RE RA is partner bij Coney (angelique.koopman@coney.nl) en promovenda bij Tilburg University (a.j.m.koopman@uvt.nl) op het gebied van process mining in het domein van interne beheersing en continuous monitoring/auditing.

Noten

■ Een data scientist is gespecialiseerd in het verzamelen, bewerken en analyseren van gegevens met als doel deze gegevens om te zetten in informatie ofwel waarde voor de organisatie. Voor big data-toepassingen worden data scientists ingezet.

■ Petri-netten en BPMN zijn veelgebruikte beschrijvingswijzen. Met behulp van een Petri-net kan een proces grafisch weergegeven worden. De formele basis maakt het ook mogelijk om te redeneren over processen en deze te automatiseren. BPMN staat voor Business Process Modeling No-

tation en is een industriestandaard voor het weer-geven van processen. Zie (Dijkman, 2008) voor de relatie tussen BPMN en Petri-netten.

■ Hiermee bedoelen we het door de process mining tool automatisch afgeleide en daarna gevisualiseerde procesmodel.

Literatuur

- Aalst, W.M.P. van der (2015). Business process simulation survival guide. In J. vom Brocke & M. Rosemann (editors), *Handbook on Business Process Management 1, International Handbooks on Information Systems* (pp. 337-370). Berlin: Springer-Verlag.
- Aalst, W.M.P. van der (2014). Data scientist: The engineer of the future. In K. Mertins, F. Benaben, R. Poler, & J. Bourrieres (editors), *Proceedings of the I-ESA Conference, volume 7 of Enterprise Interoperability* (pp. 13-28). Berlin: Springer-Verlag.
- Aalst, W.M.P. van der, & Geffen, F. van (2014). De ontbrekende schakel tussen BI en BPM. *Informatie*, juni/juli.
- Aalst, W.M.P. van der (2013). Business process management: A comprehensive survey. *ISRN Software Engineering*, 1-37.
- Aalst, W.M.P. van der (2011). *Process mining: Discovery, conformance and enhancement of business processes*. Berlin: Springer-Verlag.
- Aalst, W.M.P. van der, Adriansyah, A., & Dongen, B. van. (2012). Replaying history on process models for conformance checking and performance analysis. *WIREs Data Mining and Knowledge Discovery*, 2(2), 182-192. doi: 10.1002/widm.1045.
- Aalst, W.M.P. van der, Hee, K.M. van, Werf, J.M.E.M. van der, Kumar, A., & Verdonk, M.C. (2011). Conceptual model for on line auditing. *Decision Support Systems*, 50(3), 636-647.
- Chan, D.Y., & Vasarhelyi, M.A. (2011). Innovation and practice of continuous auditing. *International Journal of Accounting Information Systems*, 12, 152-160.
- Dijkman, R.M., Dumas, M. & Ouyang, C. (2008). Semantics and analysis of business process models in BPMN. *Information & Software Technology*, 50(12): 1281-1294.
- Jans, M., Alles, M., & Vasarhelyi, M. (2013). The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems*, 14, 1-20.
- Leemans, S.J.J., Fahland, D., & Aalst, W.M.P. van der (2015). Exploring processes and deviations (pp. 304-316). *Business Process Management Workshops 2014*. Lecture Notes in Business Information Processing Volume 202, Berlin: Springer International Publishing, 2015.
- Leslie, D.A., Aldersley, S.J., Cockburn, D.J., & Reiter, C.J. (1986). An assertion-based approach to auditing. *Proc. of the 1986 Touche Ross/University of Kansas Symposium on Auditing Problems* (pp. 31-67). School of Business, Univ. of Kansas.
- Nederlandse Beroepsorganisatie van Accountants, Koninklijke (2015). Controlestandaard 200, Algemene doelstellingen van de onafhankelijke accountant, alsmede het uitvoeren van een controle overeenkomstig de Standaarden. Handleiding Regelgeving Accountancy (HRA) deel 1. Nadere voorschriften controle- en overige standaarden (NV COS).
- Rijnders, Q., Eijken, T., Fissette, M., & Schijndel, J. van (2013). Behoeftte aan visuele technieken voor verbetering controle. *Maandblad voor Accountancy en Bedrijfseconomie*, 87(6), 282-289.
- Rozinat, A., Alves de Medeiros, A.K., Gunther, C.W., Weijters, A.J.M.M., & Aalst, W.M.P. van der (2007). *Towards an evaluation framework for process mining algorithms*. BPM Center Report BPM-07-06, BPMcenter.org.
- Swart, J. de, Wille, J., & Majoor, B. (2013). Het 'push left'-principe als motor van data analytics in de accountantscontrole. *Maandblad voor Accountancy en Bedrijfseconomie*, 87 (10), 425-433.