



Lessons learned bij toepassing van **process mining**

Het onderwerp 'process mining'¹ is al een aantal jaren een hot topic. In dit tijdschrift verscheen drie jaar geleden een artikel van dezelfde auteur en twee medeauteurs [BEZV09]. Dat artikel ging over de toepassing van process mining in IT-auditing. Nu, drie jaar verder, geeft dit artikel de lessen weer die de auteur en haar collega's de afgelopen jaren bij toepassing in de praktijk hebben geleerd.

MARIA HAASNOOT-BEZVERHAYA

Process mining is circa tien jaar geleden ontwikkeld door de werkgroep van prof. W. van der Aalst bij de TU Eindhoven en verspreidt zich steeds meer vanuit de wetenschappelijke wereld naar het bedrijfsleven. Er zijn inmiddels veel artikelen over de ontwikkeling en toepassing van deze techniek geschreven. Ook zijn er verschillende tools ontwikkeld die process mining-technieken gebruiken, zoals het open source tool *ProM* (TU Eindhoven) en commerciële tools, zoals *ReflectOne* (Futura Technology), *BPMOne* (Pallas Athena), *Nitro* en *Disco* (Fluxicon).

De departementale auditdienst van het ministerie van Infrastructuur en Milieu² (IenM) past sinds enkele jaren process mining-technieken toe op verschillende complexe geautomatiseerde bedrijfsprocessen. Op basis van deze onderzoeken wil ik in dit artikel onze ervaringen en *lessons learned* samenvatten en de belangrijkste aandachtspunten en knelpunten bij de toepassing van process mining benoemen. Bovendien wil ik aangeven wat een mogelijke oplossing is voor deze knelpunten.

PROCESS MINING

Process mining is een krachtige techniek om bedrijfsprocessen³ transparant te maken, knelpunten in het procesverloop te ontdekken en verschillende performanceanalyses op bedrijfsprocessen uit te voeren. Het basisprincipe van process mining is in figuur 1 afgebeeld. De implementatie van een bedrijfsproces in een informatiesys-

teem verloopt *top-down*. In hoofdlijnen betekent dit dat de ontwikkelaar eerst een ontwerp van een bedrijfsproces maakt (het *soll*-model, ofwel het proces in opzet) en dat vervolgens in het systeem implementeert.

Process mining is een *bottom-up* techniek waarmee een onderzoeker, bijvoorbeeld een IT-auditor, op basis van de in een informatiesysteem gelogde gegevens een procesmodel opbouwt (*ist*-model, het bestaan en de werking van het proces). Zo kan men ontdekken hoe bedrijfsprocessen werkelijk verlopen. Bovendien kan men het werkelijke procesverloop vergelijken met het ontwerp van het proces en mogelijke afwijkingen constateren. Deze afwijkingen kunnen ontstaan door verschillende oorzaken zoals fouten in de implementatie, verkeerde inrichting of configuratie van het systeem of het verkeerd gebruik van het systeem door de medewerkers.

De basis voor process mining is een *event log*. Dit is een bestand met gestructureerde data over een bedrijfsproces verkregen uit een informatiesysteem. Bijvoorbeeld: wie (welke gebruiker) heeft wat (activiteit in het bedrijfsproces), wanneer (datum en tijdstip) met betrekking tot een bepaald object in het bedrijfsproces in het systeem gedaan? De onderzoeker analyseert de event log vervolgens met een process mining-tool, bijvoorbeeld de open source tool *ProM*. Analyse van de event log is vanuit verschillende

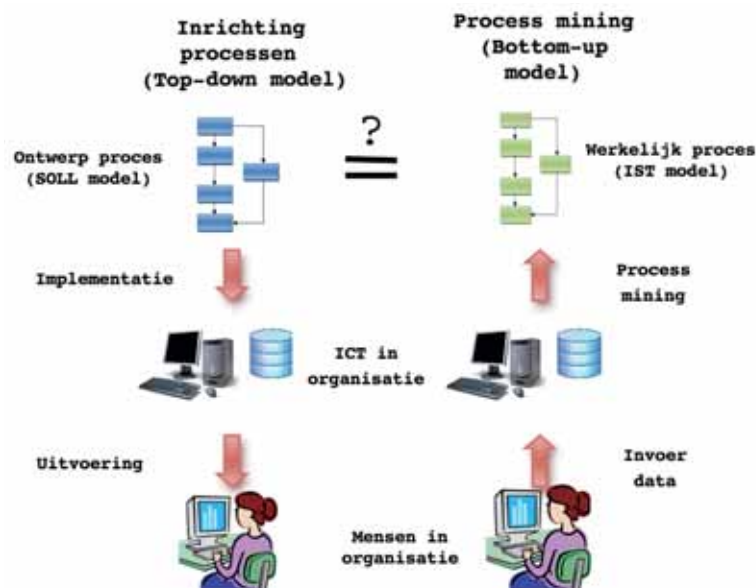


perspectieven mogelijk. Bijvoorbeeld: vanuit procesperspectief kan men kijken hoe het proces in werkelijkheid verloopt op basis van de geregistreerde data over een bepaalde periode. Vervolgens kan de event log vanuit caseperspectief geanalyseerd worden, dat wil zeggen dat men in een bepaald proces op een bepaald object (bijvoorbeeld, inkoopnummer) kan inzoomen en deze case kan ana-

lyseren. Vanuit gebruikersperspectief kan men zien wat de relatie tussen de gebruikers (of de afdelingen) in het proces in deze periode was, en of de functiescheidingen functioneren.

Voor IT-auditors (en ook voor proceseigenaren) is process mining een goede methode om de werking van een proces te visualiseren, te meten en vervolgens de soll- en ist-modellen van het proces met elkaar te

vergelijken (opzet tegenover bestaan en werking). Met process mining kunnen IT-auditors bijvoorbeeld aantonen dat functiescheidingen binnen een bepaalde periode doorbroken zijn, en kunnen ze de werking van *application controls* controleren. In het geval dat een IT-audit als ondersteuning voor een operationele audit wordt uitgevoerd, kunnen ook andere aspecten van een proces van belang zijn, namelijk efficiëntie en effectiviteit. Hierbij kan process mining de bottlenecks van het proces in kaart brengen, waardoor de IT-auditor conclusies over de oorzaken kan trekken en verbeterpunten kan aangeven.



Figuur 1: Inrichting van bedrijfsprocessen en process mining-aanpak

PROBLEMATIEK

Het principe van de process mining-techniek is dat een proces in het systeem achteraf wordt geanalyseerd. Dat wil zeggen dat het proces eerst binnen een informatiesysteem wordt geïmplementeerd en dat daarna op basis van de data in de event log het proces in kaart wordt gebracht. Een vaak voorkomend knelpunt bij de analyse van bestaande systemen is dat de onderzoeker in de voorbereidingsfase van het process mining onder-

Casus 1 Factuurverwerkingsproces

Het factuurverwerkingsproces bij sommige onderdelen van lenM is geautomatiseerd en verloopt in het SAP-systeem via een workflow. Het proces bestaat uit vijf hoofdstappen:

1. Factuur scannen
2. Factuur valideren
3. Prestatieverklaring geven
4. Factuur boeken
5. Factuur betalen

Behalve deze hoofdstappen kunnen binnen de workflow nog veel meer interne acties rond een factuur plaatsvinden, bijvoorbeeld: een factuur kan van status veranderen, worden teruggestuurd naar een andere afdeling, of worden geblokkeerd. Voor het monitoren van het proces werd in het

systeem een rapportage ontwikkeld die de hoofdstappen van het proces uit de workflow weergeeft zoals scandatum, workflow begindatum, workflow einddatum en betaaldatum. Maar in deze rapportage zijn de interne tussenstappen alleen per factuur zichtbaar te maken op het scherm van de computer in een pop-up venster van de rapportage. Dit detailniveau blijft in de programmering van de workflow verborgen. Hierdoor is er geen totaaloverzicht van alle mogelijke activiteiten van het proces.

Om deze tussenstappen uit het systeem te extraheren, was het nodig om een aanvullende rapportage te bouwen die alle (tussen)stappen van de workflow bevat. De ontwikkeling van de rapportage voor het factuurverwerkingsproces met

alle tussenstappen heeft veel tijd in beslag genomen. Er hebben veel discussies met de programmeurs plaatsgevonden omdat het ingewikkeld was om de tekstvelden van de bestaande rapportage met veel gecombineerde informatie in sommige velden op de juiste manier uit te splitsen en te plaatsen in een formaat dat geschikt is voor process mining. Verschillende fouten bij de splitsing van de velden werden telkens tijdens de test van de rapportage in de testomgeving geconstateerd en gecorrigeerd. Ten slotte werd de rapportage met een nieuwe release van het SAP-systeem in de productieomgeving gedraaid. Vervolgens werd de event log voor het factuurverwerkingsproces op basis van deze rapportage gemaakt. Deze event log gebruiken we voor de process mining-analyse van het proces.

Casus 2 Inkoopproces

Dit is een samenvatting van de casus die is beschreven in [BEZ09].

Het inkoopproces is bij lenM in het SAP-systeem geïmplementeerd. Het proces bestaat uit een aantal

stappen, zoals: creëren van de aanvraag tot bestelling (ATB), goedkeuren ATB, creëren bestelling, ontvangen goederen en facturen en afwickelen bestelling. Al deze stappen zijn in verschillende SAP-tabellen geregistreerd. In de analyse van het inkoopproces viel op dat het lastig was om in het

SAP-systeem te bepalen wat de juiste tabellen, data en velden zijn voor de analyse. Dat komt doordat sommige velden van de tabellen maatwerkvelden zijn en juist deze velden nodig zijn en niet de standaardvelden van SAP. Vaak was het tijdstip van de activiteiten ook niet geregistreerd in het systeem.

Casus 3 TBU-proces

Het proces Tegemoetkoming Buitengewone Uitgaven (TBU) is een complex proces bij de Belastingdienst. TBU is het afhandelen van de aftrek van ziektekosten als buitengewone uitgaven voor mensen met een laag inkomen en hoge zorgkosten. De gegevens van dit proces zijn in

verschillende systemen geregistreerd in een groot aantal tabellen met verschillende unieke sleutels. Om een event log voor de analyse te creëren, moet de onderzoeker in eerste instantie alle unieke sleutels voor de hele analyse bepalen. Bijvoorbeeld: in een tabel is de unieke sleutel het sofinummer van de klant, maar voor een andere

tabel is dat niet voldoende, en moet ook nog het fiscale jaar daaraan gekoppeld worden. In een volgende tabel moet er tevens de fiscale partner daaraan toegevoegd worden om de totale sleutel voor de event log uniek te maken. Uiteindelijk wordt een complexe event log met alle benodigde data gemaakt.

zoek ontdekt dat het systeem niet alle vereiste data beschikbaar kan stellen, dat de data onvolledig geregistreerd is of dat de data van slechte kwaliteit is. Een proces dat op deze manier in het systeem geregistreerd wordt, is niet geschikt voor analyse via process mining, tenzij er in het systeem een aantal aanpassingen worden gedaan. Dit is echter vaak vrij kostbaar en complex, omdat deze wijziging de andere processen in het systeem kan beïnvloeden en men een *request for change*-traject moet doorlopen.

LESSONS LEARNED UIT DE PRAKTIJKCASUSSEN

Binnen de rijksoverheid hebben we een aantal pilots met process mining

uitgevoerd (zie de tekstkaders). Hierna volgt een samenvatting van de belangrijkste lessen en aandachtspunten uit deze onderzoeken.

Een kant-en-klaar event log is niet altijd beschikbaar

Er is meestal geen kant-en-klaar event log beschikbaar dat geschikt is voor process mining. In dit geval moet de onderzoeker de event log zelf opbouwen door bepaalde velden uit verschillende tabellen uit een systeem of systemen te combineren. Als een proces via een aantal systemen verloopt, is het soms lastig om de juiste unieke sleutel (case ID) voor de event log te vinden, zoals in casus 3 is aangegeven.

Het is vaak lastig de benodigde data te vinden

In een (ERP-)systeem is het vaak lastig om de data te vinden die voor process mining nodig is. Omdat de onderzochte bedrijfsprocessen al vaker door middel van data mining-technieken zijn geanalyseerd, zijn de benodigde tabellen van het systeem, bijvoorbeeld SAP, voor de analyse op transactieniveau vaak bekend. Maar voor de process mining-analyse is nog aanvullende data noodzakelijk, zoals bijvoorbeeld tijdstip van een activiteit, gebruikersnaam van de persoon die een activiteit heeft uitgevoerd of de naam van de activiteit. Deze gegevens zijn vaak niet gelogd. De registratie van het ▣



tijdstip van een activiteit is heel belangrijk, omdat process mining een procesmodel maakt op basis van datum en tijdstip van de gelogde gegevens. Als alleen datum en geen tijdstip geregistreerd is en er een aantal activiteiten in een proces binnen een dag zijn uitgevoerd, is niet duidelijk in welke volgorde de processtappen moeten worden afgebeeld en kan het process mining-tool een vertekend beeld van het proces geven.

Registratie van gebruikers in het systeem gebeurt ook niet bij elke activiteit. Het in kaart brengen van alle gebruikers van een proces is nodig voor de autorisatiecontroles en voor de organisatorische analyse en de sociale netwerkanalyse van het proces (analyses die inzicht geven welke medewerker bepaalde activiteiten heeft uitgevoerd en hoe vaak, respectievelijk welke medewerkers met elkaar samenwerken). Deze problematiek is in de casussen 2 en 3 aan de orde gekomen.

Aanpassingen in het systeem kunnen tijdrovend en kostbaar zijn

Soms zijn aanpassingen in het systeem nodig om bruikbare data te verkrijgen. Sommige informatiesystemen, zoals SAP, zijn complex. De benodigde data is vaak moeilijk of zelfs onmogelijk te verkrijgen, zoals in casus 1. Bijvoorbeeld: de data van sommige bedrijfsprocessen kan in tientallen tabellen zijn opgeslagen. Deze tabellen zijn vaak heel groot en daarom lastig te downloaden. Bovendien kan de data in het systeem zo 'verborgen' zijn dat het onmogelijk is om de data te bemachtigen zonder het systeem uit te breiden met extra tabellen of door specifieke rapportages te bouwen. Bij grote bedrijven met complexe systemen is het een ingewikkeld proces om een systeem aan te passen. Bij elke aanvraag moeten de *change management*-procedures doorlopen. Dit duurt soms vrij lang, omdat alle wijzigingen in het systeem moeten worden afgestemd

en de toegangsrechten tot de nieuwe objecten bepaald moeten worden. Bovendien zijn aanpassingen in het systeem vaak kostbaar, en daarom moet men rekening houden met deze kosten bij de afweging of process mining gerechtvaardigd is.

Data is niet altijd volledig of is van slechte kwaliteit

Uit onze praktijkervaring met data- en process mining analyses blijkt dat de data in de informatiesystemen vaak niet volledig is of van onvoldoende kwaliteit voor process mining. Dat wil zeggen, als de data handmatig ingevuld moet worden, ontstaan soms typefouten of worden velden overgeslagen. In casus 1 over het factuurverwerkingsproces bij de workflow stap 'Factuur in de wacht gezet' bijvoorbeeld, was het veld met de toelichting bijna nooit ingevuld, hoewel dat veld verplicht is. Maar medewerkers zetten in het veld een spatie en konden in de workflow doorgaan. Voor de analyse van het

Niveau	Beschrijving	Voorbeelden
5	Hoogste niveau: event logs zijn van zeer goede kwaliteit (betrouwbaar en volledig). Gegevens zijn juist gedefinieerd, automatisch en systematisch in het systeem gelogd.	Semantisch geannoteerde logs van BPM systemen.
4	Gegevens zijn in het informatiesysteem automatisch en op een systematische, volledige en betrouwbare wijze gelogd. In tegenstelling tot de event logs van niveau 3 worden de variabelen, zoals process instance (case) en activiteiten, expliciet in het systeem gelogd en bijgehouden.	Event logs van traditionele BPM/ workflow systemen.
3	Gegevens zijn in het informatiesysteem automatisch, maar niet op een systematische en volledige wijze gelogd. Anderzijds, in tegenstelling tot de event logs van niveau 2, is er enige garantie dat de gelogde gegevens betrouwbaar zijn.	Tabellen in ERP systemen, event logs van CRM systemen, transactionele logs van messaging systemen, event logs van high-tech systemen.
2	Gegevens zijn in het informatiesysteem automatisch gelogd, maar de beslissing over welke gegevens gelogd moeten worden is niet op een systematische wijze genomen. Bovendien is het mogelijk het informatiesysteem ongeautoriseerd aan te passen. De gegevens kunnen onvolledig of onbetrouwbaar zijn of ontbreken.	Event logs van document- and product managementsystemen, error logs van embedded systemen, worksheets van service engineers.
1	Laagste niveau. Event logs zijn van slechte kwaliteit. Gelogde gegevens zijn onvolledig en/of corresponderen niet met de realiteit.	Typisch voor de handmatig gemaakte event logs, 'yellow notes', hard-copy van de medische dossiers.
2	Gegevens zijn in het informatiesysteem automatisch gelogd, maar de beslissing over welke gegevens gelogd moeten worden is niet op een systematische wijze genomen. Bovendien is het mogelijk het informatiesysteem ongeautoriseerd aan te passen. De gegevens kunnen onvolledig of onbetrouwbaar zijn of ontbreken.	Event logs van document- and product managementsystemen, error logs van embedded systemen, worksheets van service engineers.
1	Laagste niveau. Event logs zijn van slechte kwaliteit. Gelogde gegevens zijn onvolledig en/of corresponderen niet met de realiteit.	Typisch voor de handmatig gemaakte event logs, 'yellow notes', hard-copy van de medische dossiers.

Tabel 1: Volwassenheidsniveau van event logs [AALS11]

proces is het echter noodzakelijk dat dit veld correct ingevuld is om na te kunnen gaan waarom de facturen in de wacht zijn gezet.

Als de voor de process mining-analyse belangrijke velden niet ingevuld zijn, kunnen deze niet in de event log meegenomen worden en vervolgens kan de event log onvoldoende informatie over het bedrijfsproces leveren. In het 'Process mining manifest' van augustus 2011 [AALS11] heeft de *IEEE Task Force on Process Mining* een volwassenheidsmodel voor de event logs ontwikkeld. De opstellers brengen in dit model verschillende typen event logs in kaart en geven per type een kwaliteitsniveau aan (van 1-laag tot 5 –hoog, zie tabel 1). Onze ervaring met verschillende informatiesystemen binnen de rijksoverheid is dat de in een ERP-systeem gelogde gegevens volgens dit model een 3 scoren. Een SAP-systeem logt bijvoorbeeld gegevens automatisch, maar nog niet op een voor process mining gestructureerde manier.

De beschrijving van de workflowactiviteiten is niet altijd consequent

Tijdens de analyse van de activiteiten van de workflow in casus 1 kwam de bevinding naar voren dat sommige activiteiten twee keer in de workflow voorkomen, maar op twee verschillende manieren zijn benoemd. Bijvoorbeeld de workflowstap 'in wacht gezet' en 'in de wacht gezet' of 'Creditnota geboekt' en 'Credit nota geboekt'. Uit de interviews met de programmeurs van de workflow blijkt dat het geen bewuste keus is geweest om verschillende omschrijvingen voor dezelfde activiteiten te gebruiken. Uit het onderzoek blijkt ook dat deze activiteiten bij verschillende substromen binnen de workflow horen. Ze zijn door twee verschillende programmeurs ontwikkeld en hebben geen negatieve invloed op het verloop en de kwaliteit van de workflow. Maar kennelijk is er tijdens het ontwerp van de workflow geen afspraak gemaakt over de naam-

geving van de activiteiten van het proces en er is geen document met een lijst met standaardnamen. Dit overzicht met alle activiteiten en uniforme naamgeving is noodzakelijk voor de analyse en de monitoring van de workflow. Uit de bovengenoemde bevindingen kunnen we concluderen dat tijdens de implementatie van de bedrijfsprocessen in het informatiesysteem onvoldoende rekening werd gehouden met de mogelijkheid dat het geïmplementeerde proces later met process mining zou worden geanalyseerd. Over de analyse en monitoring van het proces wordt vaak pas later nagedacht. Vervolgens wordt men na de implementatie van het proces met het feit geconfronteerd dat de event log van het proces ontbreekt en data niet volledig of van onvoldoende kwaliteit is.

CONCLUSIE

Over de voordelen en mogelijkheden van process mining is veel gepubliceerd. In dit artikel zijn vooral de aandachtspunten en moeilijkheden bij de toepassing van deze techniek in kaart gebracht.

De belangrijkste obstakels bij het toepassen van process mining zijn de vaak matige kwaliteit van de beschikbare data, de moeite om de juiste data te selecteren en de kosten die gemoeid zijn met het construeren van event logs vanuit de bronsystemen die geschikt zijn voor analyse.

Een van de mogelijke oplossingen voor deze problemen is over de toepassing van process mining na te denken in de beginfase van de implementatie van een bedrijfsproces in een systeem, namelijk in de analyse- en ontwerpfase.

Als de registratie van data op een zorgvuldige manier wordt ingericht en uitgevoerd, produceert het systeem event logs van hoge kwaliteit en kunnen bedrijfsprocessen en systemen (eventueel continu) worden gecontroleerd/gemonitord en geanalyseerd zowel ten behoeve van audits als van procesbeheersing. ■

Noten

1. Process Mining is een methode om bedrijfsprocessen in en rond een informatiesysteem op te sporen en te analyseren met behulp van de event log die het informatiesysteem produceert.
2. De departementale auditdienst van lenM is sinds 1 mei 2012 onderdeel van de Auditdienst Rijk (ADR), die onder het ministerie van Financiën valt.
3. Bedrijfsprocessen: primaire of secundaire (geheel of gedeeltelijk) geautomatiseerde bedrijfsprocessen die door een informatiesysteem worden ondersteund. Voorbeelden zijn een productieproces, een inkoopproces, een betaalproces of een subsidieproces.

Literatuur

- [AALS11] Aalst, Wil van der et al., *Process mining Manifesto*, 2011, <http://www.win.tue.nl/ieeetfpm/downloads/Process%20Mining%20Manifesto.pdf>.
- [BEZV09] Bezverhaya-Haasnoot, M., E. Caron en P. Goeyenbier, Naar een softwarematige analyse van bedrijfsprocessen voor auditing; Process Mining als gereedschap voor (IT-)Auditors, in: *de EDP-Auditor*, nr. 2, 2009.
- Overige literatuur*
- Arts, Th., *Continuous Auditing in een ERP-omgeving*, 2009 (Referaat).
- Brouwers, P., M. op het Veld en A. Lissone, Tool-based monitoring en auditing van ERP-systemen, van hebbeding naar noodzaak, in: *Compact*, nr.2, 2006.
- Dennis, Alan en Barbara Haley Wixom, *Systeemanalyse en ontwerp*, Academic Service, 2004.
- Jans, Mieke *Process mining of event logs in auditing: opportunities and challenges*, 2010.
- Kuijck, J. van, Continuous monitoring & auditing, in: *Audit Magazine*, nr.1, 2008.
- Mundzer, Martina, *Process Mining binnen de Rijksoverheid: Analyse van het factureringsproces binnen het Ministerie van Infrastructuur en Milieu*, 2011 (Afstudeerscriptie).
- Veld, M. op het, S. Biggelaar, van den en H. Daanen, 'Audit Innovation', in: *Compact*, nr.1, 2010.
- Scheeres, W., Continuous auditing, in: *de EDP-Auditor*, nr. 3, 2007.



Dr. M. (Maria) Haasnoot-Bezverhaya EMITA

is sinds 2008 werkzaam bij de Departementale Auditdienst (DAD) van lenM (per 1 mei 2012 onderdeel van de nieuw gevormde Auditdienst Rijk – ADR). In 2012 heeft zij de postinitiële opleiding IT-Auditing & Advisory bij de Erasmus Universiteit Rotterdam afgerond. Haar interesse gaat vooral uit naar process mining, softwarekwaliteit en projectmanagement.