

# Process Mining Manifest

Een *manifest* is een "publieke verkondiging van principes en bedoelingen" door een groep mensen. Dit manifest werd geschreven door leden en sympathisanten van de *IEEE Task Force on Process Mining*. Het doel van deze Task Force is het verbeteren van onderzoek naar de ontwikkeling van, opleiding over, implementatie van, vooruitgang van en begrip over process mining.

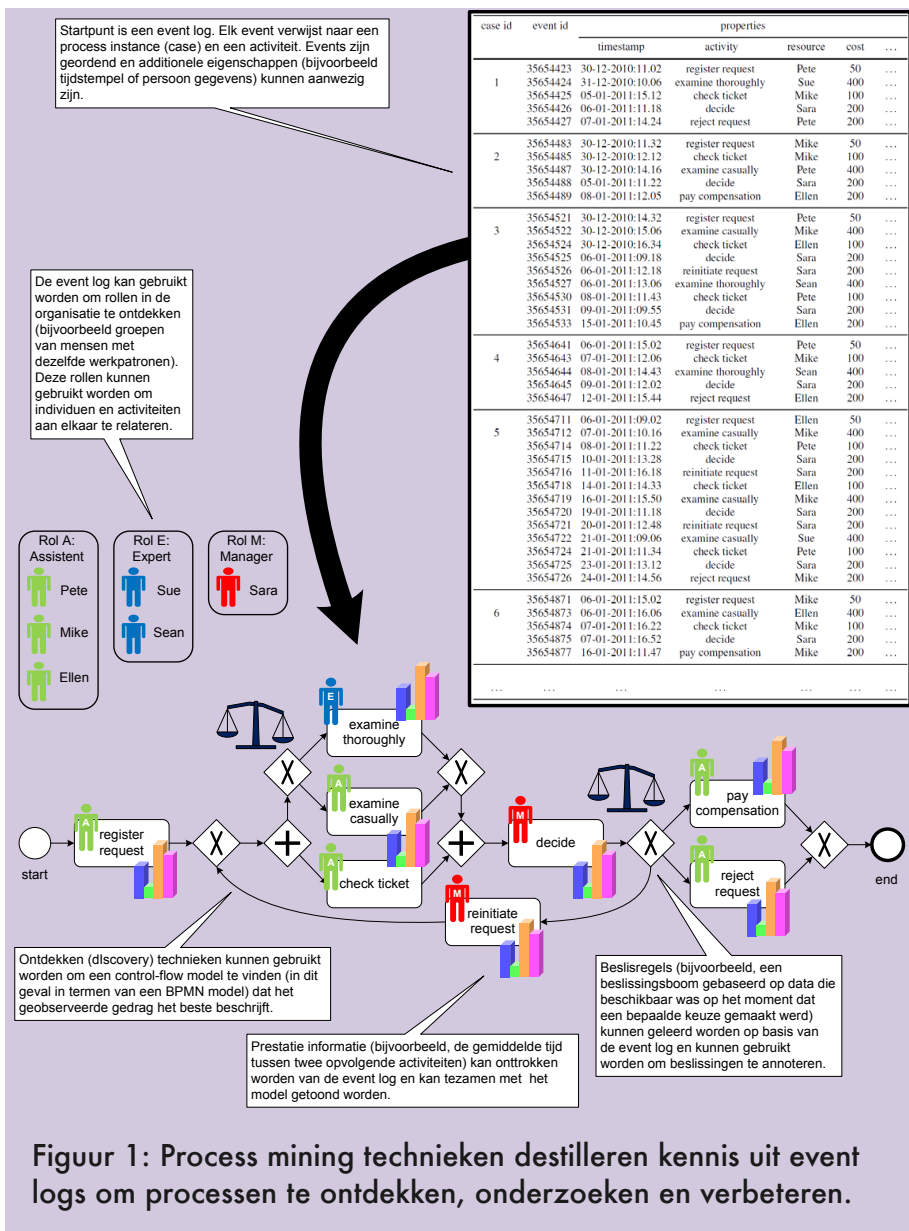
Process mining is een relatief jong onderzoeksdomein en bevindt zich tussen computational intelligence en data mining aan de ene kant en tussen procesmodelleren en procesanalyse aan de andere kant. Het idee achter process mining is het *ontdekken, onderzoeken en verbeteren van de echte processen* (dus niet van de veronderstelde processen) *door kennis te destilleren uit event logs*, welke beschikbaar zijn in hedendaagse (informatie-)systemen (zie Fig. 1). Process mining bestaat uit (automatisch) ontdekken van een proces (procesmodellen opstellen aan de hand van een event log), conformiteit controleren (afwijkingen achterhalen

door het vergelijken van een model en een log), sociale netwerken achterhalen (bijvoorbeeld in bedrijfscontext), automatisch bouwen van simulatiemodellen, modellen

## Inhoud:

Process Mining: Stand van Zaken	3
Richtlijnen	7
Uitdagingen	10
Epiloog	15
Verklarende Woordenlijst	16

Process mining technieken zijn in staat om *kennis uit event logs te destilleren*, welke algemeen beschikbaar zijn in hedendaagse informatiesystemen. Deze technieken voorzien nieuwe manieren om *processen te ontdekken, te bestuderen en te verbeteren* in een waaier van toepassingsdomeinen. Twee factoren sturen de groeiende interesse in process mining. Aan de ene kant worden steeds meer events geregistreerd en wordt dus steeds meer gedetailleerde informatie over de historie van processen beschikbaar. Aan de andere kant groeit de behoefte om bedrijfsprocessen te ondersteunen en te verbeteren in continu veranderende omgevingen. Dit manifest werd gemaakt door de *IEEE Task Force on Process Mining* en heeft als doel het promoten van het begrip process mining. Door bovendien een set van richtlijnen te definiëren en belangrijke uitdagingen in het domein vast te leggen, wil dit manifest een grote hulp betekenen voor *software-ontwikkelaars, wetenschappers, consultants, business managers, en eindgebruikers*. Het doel is om de volwassenheid van process mining als nieuwe tool te verhogen om het (her)ontwerpen, controleren en ondersteunen van operationele bedrijfsprocessen te verbeteren.



Figuur 1: Process mining technieken destilleren kennis uit event logs om processen te ontdekken, onderzoeken en verbeteren.

uitbreiden, modellen aanpassen, voorspellen en op het verleden gebaseerde aanbevelingen voorstellen.

Process mining slaat een belangrijke brug tussen data mining en bedrijfsprocesmodelleren en -analyse. Onder de verzamelterm Business Intelligence (BI) werden heel wat termen geïntroduceerd die eerder handelen over eenvoudige rapportering- en dashboard-tools. Business Activity Monitoring (BAM) gaat over technologieën die het mogelijk maken om bedrijfsprocessen in real-time te bewaken. Complex Event Processing (CEP) beschrijft technologieën om grote hoeveelheden aan events te verwerken, te bewaken, sturen en om een bedrijfsproces te optimaliseren in real time. Corporate Performance Management (CPM) is een andere term voor het meten van de prestaties van een proces of een organisatie. Ook bepaalde managers

praktijken zoals Continuous Process Improvement (CPI), Business Process Improvement (BPI), Total Quality Management (TQM) en Six Sigma zijn hieraan gerelateerd. Deze praktijken hebben als gemeenschappelijke eigenschap dat ze processen "onder de loep nemen" om na te gaan of verbeteringen mogelijk zijn. Process mining is een technologie die CPM, BPI, TQM, Six Sigma en gelijkaardige technieken ondersteunt.

BI tools en management praktijken zoals Six Sigma en TQM proberen de operationele prestaties te verbeteren, bijvoorbeeld flow time en uitval verminderen, maar organisaties leggen meer en meer nadruk op corporate governance, risicobeheer en compliance. Regelgevingen zoals de Sarbanes-Oxley Act (SOX) en het Basel II-akkoord illustreren deze focus op compliance-problemen. Process mining voorziet een manier om compliance

## Concrete doelstellingen van de Task Force zijn:

- 1) Eindgebruikers, ontwikkelaars, consultants, managers en onderzoekers op de hoogte stellen van de laatste ontwikkelingen in process mining,
- 2) Het gebruik van process mining technieken en gereedschappen promoten en het stimuleren van nieuwe toepassingen,
- 3) Een rol spelen in de inspanningen om het loggen van event data te standaardiseren,
- 4) Tutorials, speciale sessies, workshops, panels te organiseren en
- 5) Artikels, boeken, video's en special issues van tijdschriften te publiceren.

veel diepgaander na te gaan en verzekert de validiteit en betrouwbaarheid van informatie over de kernprocessen in een organisatie.

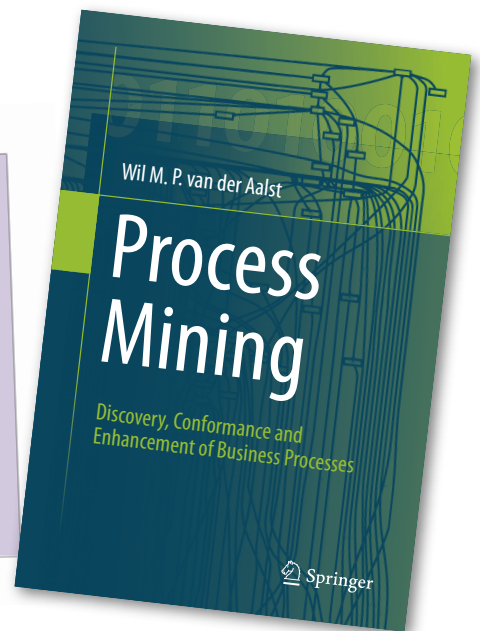
Gedurende de laatste decennia werden steeds meer gegevens over events beschikbaar en konden de process mining technieken rijpen. Bovendien, zoals reeds vermeld kunnen management trends gerelateerd aan procesverbetering (bijvoorbeeld Six Sigma, TQM, CPI en CPM) en compliance (SOX, BAM, enz.) voordeel halen uit process mining. Gelukkig werden de process mining algoritmes geïmplementeerd in verschillende academische en commerciële systemen. Vandaag de dag werkt een actieve groep van onderzoekers rond process mining en is het één van de "hot topics" in onderzoek naar Business Process Management (BPM). Bovendien is er een enorme interesse vanuit de wereld in process mining. Steeds meer softwareleveranciers voegen process mining functionaliteiten toe aan hun producten. Voorbeelden van softwareproducten met process mining mogelijkheden zijn: ARIS Process Performance Manager (Software AG), Comprehend (Open Connect), Discovery Analyst (StereoLOGIC), Flow (Fourspark), Futura Reflect (Futura Process

Intelligence), Interstage Automated Process Discovery (Fujitsu), OKT Process Mining suite (Exeura), Process Discovery Focus (Iontas/Verint), ProcessAnalyzer (QPR), ProM (TU/e), Rbminer/Dbminer (UPC) en Reflect one (Pallas Athena). De groeiende interesse in op logs gebaseerde procesanalyse was de aanzet om een Task Force on Process Mining op te richten.

De Task Force werd opgericht in 2009 in de context van het Data Mining Technical Committee (DMTC) van de Computational Intelligence Society (CIS) van het Institute of Electrical and Electronic Engineers (IEEE). De huidige leden van de Task Force bestaan uit *softwareleveranciers* (bijvoorbeeld Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), *consultancy firma's/ eindgebruikers* (bijvoorbeeld ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chili, BWI Systeme GmbH, Excellentia BPM, Rabobank) en *onderzoeksinstituten* (bijvoorbeeld TU/e, University of Padua, Universitat Politècnica de Catalunya, New Mexico State University, IST - Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universität zu Berlin, Queensland University of

## Process Mining Boek

[www.processmining.org/book/](http://www.processmining.org/book/)  
W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.



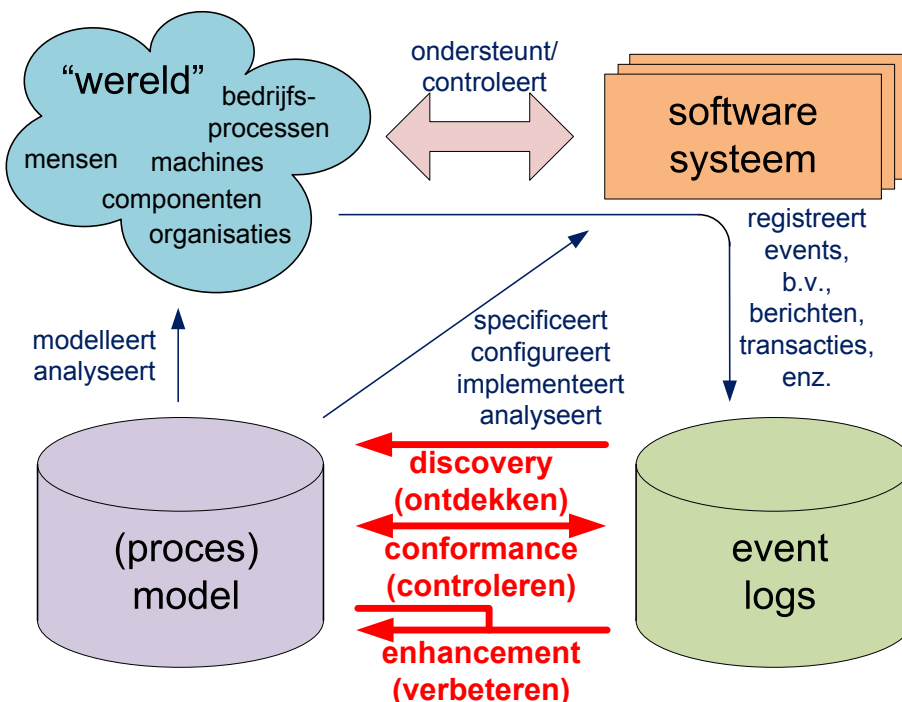
Technology, Vienna University of Economics and Business, Stevens Institute of Technology, University of Haifa, University of Bologna, Ulsan National Institute of Science and Technology, Cranfield University, K.U. Leuven, Tsinghua University, University of Innsbruck, University of Tartu).

Sinds het ontstaan in 2009 werden een reeks activiteiten georganiseerd in het kader van bovenstaande doelstellingen. Zo werden bijvoorbeeld verschillende workshops en speciale tracks ge(co)-organiseerd door de Task Force, bijvoorbeeld de workshops over Business Process Intelligence (BPI'09, BPI'10, en BPI'11) en speciale tracks op de grote IEEE conferenties (bijvoorbeeld CIDM'11). Kennis werd verspreid via tutorials (bijvoorbeeld

WCCI'10 en PMPM'09), summer schools (ESSCaSS'09, ACPN'10, CICH'10, enz.), video's (zie [www.processmining.org](http://www.processmining.org)) en verschillende publicaties, zoals ook het eerste boek over process mining dat recent werd gepubliceerd bij Springer.

De Task Force was ook (co-)organisator van de eerste Business Process Intelligence Challenge (BPIC'11): een wedstrijd waarbij deelnemers betekenisvolle resultaten uit grote en complexe event logs moeten halen. In 2010 standaardiseerde de Task Force ook XES ([www.xes-standard.org](http://www.xes-standard.org)), een standaard bestandsformaat voor logging dat uitbreidbaar is en ondersteund wordt door de OpenXES bibliotheek ([www.openxes.org](http://www.openxes.org)) en door tools zoals ProM, XESame, Nitro, enz.

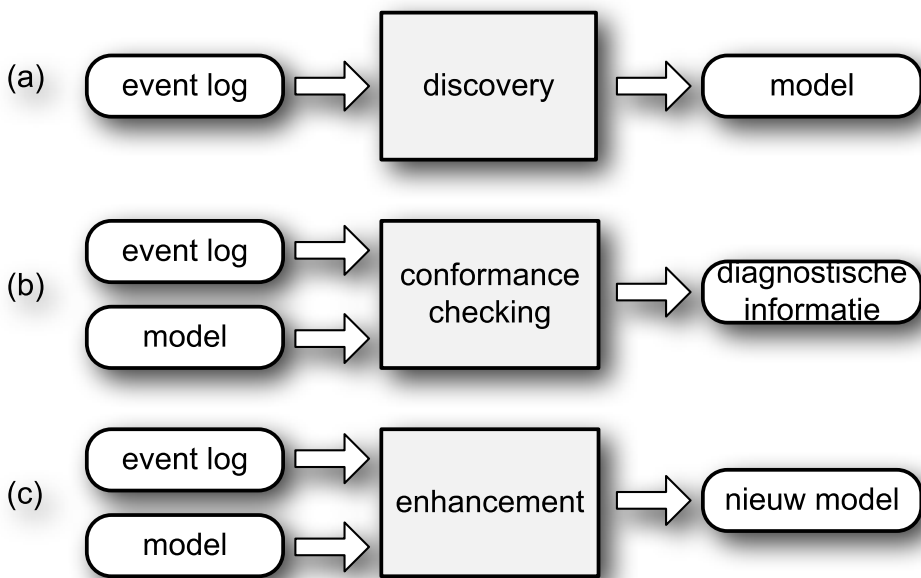
De lezer wordt uitgenodigd om op <http://www.win.tue.nl/ieeetfpm/> te kijken voor meer informatie over de activiteiten van de Task Force.



Figuur 2: Positionering van de drie basistypes van process mining: (a) *discovery* (ontdekken), (b) *conformance checking* (controleren) en (c) *enhancement* (verbeteren).

## 2. Process Mining: Stand van Zaken

De groeiende mogelijkheden van informatiesystemen en andere intensieve systemen worden goed beschreven door de wet van Moore. Gordon Moore, de co-oprichter van Intel, voorspelde in 1965 dat het aantal componenten in geïntegreerde circuits elk jaar zou verdubbelen. Gedurende de laatste vijftig jaar was de groei inderdaad exponentieel, zij het wel aan een iets trager tempo. Deze vooruitgang resulteerde in een spectaculaire groei van het "digitale universum" (alle gegevens die elektronisch worden opgeslagen of



Figuur 3: De drie basistypes van process mining in termen van input en output: (a) discovery (ontdekken), (b) conformance checking (controleren) en (c) enhancement (verbeteren).

uitgewisseld). Bovendien worden de digitale en de echte wereld steeds meer op elkaar afgestemd.

De groei van het digitale universum dat goed overeenkomt met de processen in bedrijven, maakt het mogelijk om events op te slaan en te analyseren. Events kunnen gaan over het verkrijgen van geld bij een geldautomaat, het instellen van een X-ray machine door een dokter, een burger die een aanvraag indient voor een rijbewijs, het indienen van een belastingaangifte en het ontvangen van een e-ticket nummer door een reiziger. De uitdaging bestaat er in om event data op een zinvolle manier te gebruiken: bijvoorbeeld om inzichten te krijgen, bottlenecks te ontdekken, te anticiperen op problemen, inbreuken tegen afspraken te registreren, tegenmaatregelen voor te stellen en processen te stroomlijnen. Dit is exact wat process mining allemaal doet.

Het startpunt voor process mining is een *event log*. Alle process mining technieken veronderstellen dat het mogelijk is om opeenvolgende events te registreren zodat ieder event verwijst naar een *activiteit* (een welgedefinieerde stap in een proces) en hoort bij een specifieke *uitvoering* (een instantiatie van het proces). Event logs kunnen bijkomende informatie over events opslaan. Het is zelfs zo dat waar mogelijk de process mining technieken gebruik maken van extra informatie zoals de *resource* (persoon of machine) die een activiteit uitvoert of veroorzaakt, een *tijdsindicatie* van het

event of *andere opgeslagen data* (bijvoorbeeld grootte van een bestelling).

Zoals te zien is in Fig. 2 worden event logs gebruikt voor drie types van process mining. Het eerste type van process mining is *discovery* (ontdekken). Een discovery techniek produceert vanuit een event log een model zonder enige andere voorkennis. Process discovery is de meest voorkomende process mining techniek. Veel organisaties zijn verrast dat de bestaande technieken inderdaad echte processen kunnen ontdekken voornamelijk gebaseerd op voorbeelduitvoeringen in de event logs. Het tweede type van process mining is *conformance* (controleren). Hier wordt een bestaand procesmodel vergeleken met een event log voor hetzelfde proces. Conformance checking kan gebruikt worden om na te gaan of de realiteit, zoals geregistreerd in de log, overeenkomt met het model en omgekeerd. Merk op dat verschillende soorten modellen in overweging genomen kunnen worden: conformance checking kan toegepast worden op procedurele modellen, organogrammen, declaratieve procesmodellen, bedrijfsregels en -politiecs, wetten, enz. Het derde type van process mining is *enhancement* (verbeteren). Hier wordt getracht een bestaand procesmodel uit te breiden of te verbeteren aan de hand van informatie over het actuele proces dat geregistreerd werd in een event log. Waar conformance checking het

## Kenmerken van Process Mining:

1. Process mining is niet beperkt tot het ontdekken van de control-flow.

Het ontdekken van procesmodellen vanuit event logs is een bron van inspiratie voor zowel mensen in de praktijk als academici. Daarom wordt het ontdekken van de control-flow dikwijls gezien als het meest tot de verbeelding sprekende deel van process mining. Toch is process mining meer dan enkel het ontdekken van control-flow. Aan de ene kant is ontdekken slechts één van de drie basisvormen van process mining (ontdekken, controleren, verbeteren). Aan de andere kant is process mining niet beperkt tot control-flow; organisationele-, case- en tijdsperspectieven spelen ook een belangrijke rol.

2. Process mining is niet zo maar een specifiek type data mining. Process mining kan beschouwd worden als de "missing link" tussen data mining en traditioneel modelgedreven BPM. De meeste data mining technieken zijn helemaal niet procesgericht. Procesmodellen die mogelijke gelijktijdigheid vertonen, kunnen niet vergeleken worden met eenvoudige data mining structuren zoals beslissingsbomen en associatieregels. Daarom is er de behoefte aan compleet nieuwe types van structuren, voorstellingswijzen en algoritmes.

3. Process mining is niet beperkt tot offline analyses.

Process mining technieken destilleren kennis uit historische eventdata. Hoewel de term "post mortem" data gebruikt wordt, kunnen de resultaten toegepast worden op nog lopende gevallen. Zo kan bijvoorbeeld de resterende afhandelingstijd van een gedeeltelijk behandelde bestelling voorspeld worden met behulp van een ontdekt procesmodel.

verschil meet tussen model en realiteit, probeert dit derde type van process mining het vooropgestelde model te veranderen of uit te breiden. Door bijvoorbeeld de tijdstippen in een event log te gebruiken, kan een model aangevuld worden met informatie over bottlenecks, service levels, doorlooptijden en aantal uitvoeringen.

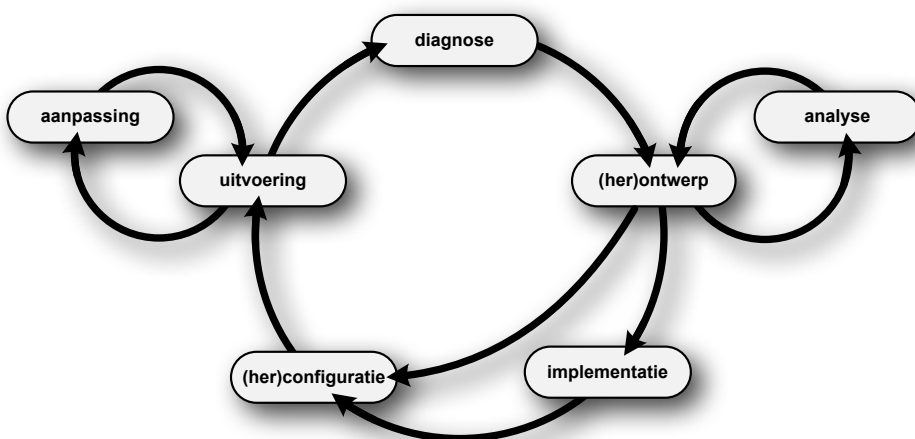
Figuur 3 beschrijft de drie types van process mining in termen van input en output. Technieken voor discovery worden gevoed met een event log en produceren een model. Het ontdekte model is meestal een procesmodel (bijvoorbeeld een Petri net, BPMN, EPC of UML activity diagram), maar het model kan evengoed andere perspectieven belichten (bijvoorbeeld een sociaal netwerk). Conformance checking technieken hebben een event log en een model nodig als input. Hun output bestaat uit diagnostische informatie die de verschillen en gelijkenissen tussen model en log aantonen. Technieken voor model enhancement (herstelling of uitbreiding) hebben ook een event log en een model nodig als input. Hun output is een verbeterd of uitgebreid model.

Process mining kan verschillende perspectieven omvatten. Het *control-flow perspectief* focust op de control-flow, dit is de volgorde van activiteiten. Het doel van mining met dit perspectief is om een goede beschrijving van alle mogelijke paden te vinden. Het resultaat wordt gewoonlijk weergegeven in een Petri net of een andere procesmodelnotatie (zoals bijvoorbeeld EPCs, BPMN of UML activity diagrammen). Het *organisationeel perspectief* focust op

informatie over resources die verborgen zit in de log: welke actoren (bijvoorbeeld mensen, systemen, rollen of departementen) zijn betrokken en wat is hun verhouding. Het doel is ofwel het in kaart brengen van de structuur van de organisatie door mensen in te delen in rollen en departementen, ofwel het blootleggen van een sociaal netwerk. Het *case perspectief* focust op eigenschappen van individuele gevallen. Uiteraard kan een case voorgesteld worden door zijn pad in het procesverloop of door de actoren die er aan werken, maar de gevallen kunnen ook beschreven worden aan de hand van bepaalde waarden van zijn gegevenselementen. Wanneer een geval bijvoorbeeld om een bijbestelling gaat, kan het interessant zijn om de leverancier of het aantal bestelde goederen voor te stellen. Het *tijdsperspectief* richt zich op timing en frequentie van events. Wanneer events tijdsinformatie bevatten, is het mogelijk om bottlenecks te ontdekken, service levels te meten, gebruik van resources te controleren en voorspellingen te doen over de resterende uitvoeringstijd van lopende gevallen.

Er bestaan een aantal algemene misvattingen over process mining. Sommige verkopers, analisten en onderzoekers beperken het bereik van de term process mining tot een speciale data mining techniek voor het ontdekken van processen op een offline basis. Dit is echter *niet* het geval en daarom benadrukken we volgende drie kenmerken van process mining.

Om process mining te positioneren gebruiken we de Business Process



Figuur 4: De BPM levenscyclus identificeert de verschillende fases van bedrijfsprocessen en hun overeenkomende informatiesyste(e)m(en); process mining speelt (mogelijk) een rol in al deze fases (behalve de implementatiefase).

## Richtlijnen:

GP1: Event Data Dienen Gezien te Worden als First-Class Citizens.

GP2: Log Destillatie Dient Gedreven te Zijn Door Vragen.

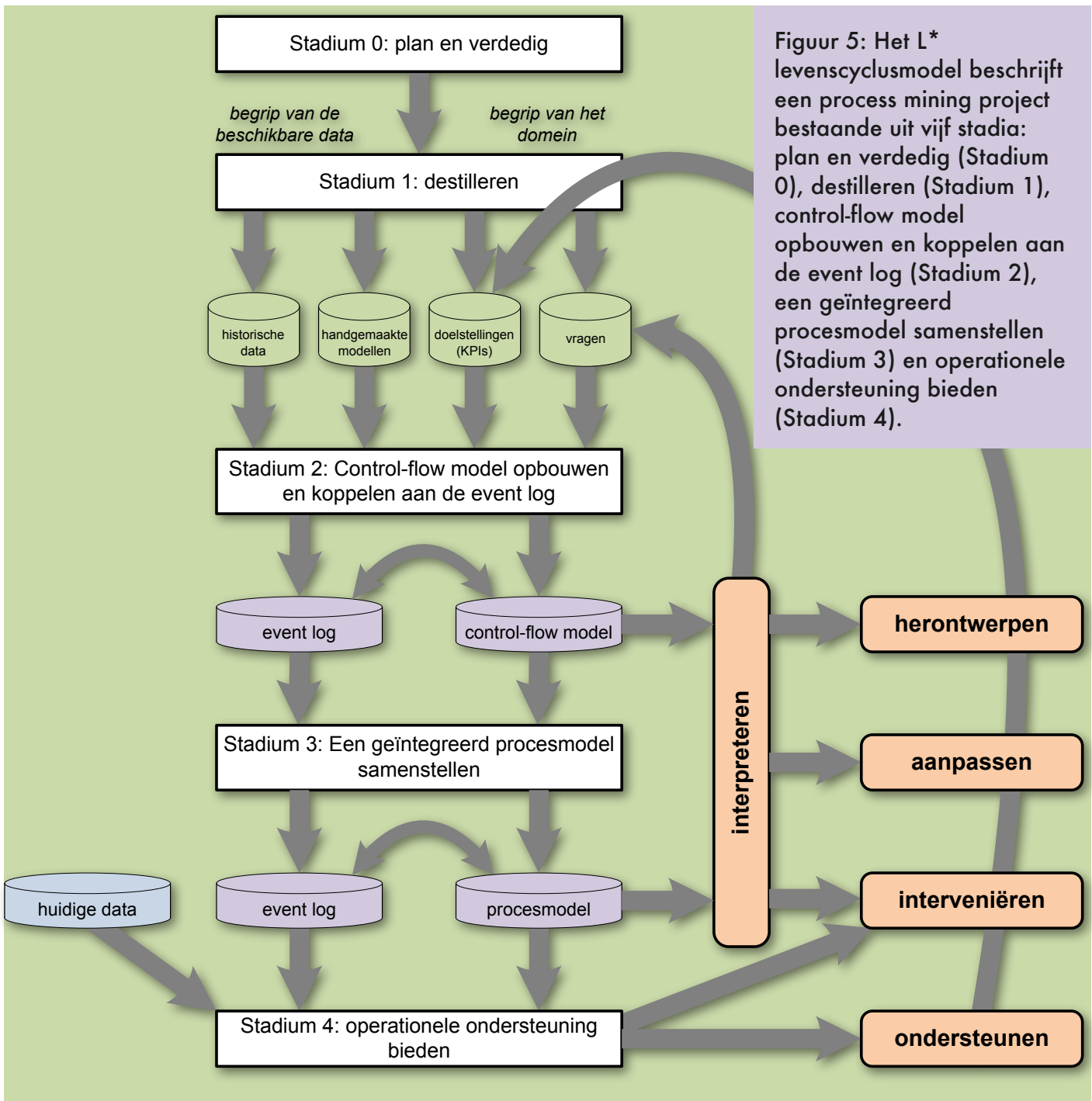
GP3: Gelijktijdigheid, Keuze en Andere Basis Control-Flow Constructies Moeten Ondersteund Worden.

GP4: Events Dienen Gerelateerd te Zijn Aan Model Elementen.

GP5: Modellen Dienen Behandeld te Worden als Betekenisvolle Abstracties van de Realiteit.

GP6: Process Mining Dient Een Continu Proces te Zijn.

Management (BPM) levenscyclus zoals weergegeven in Fig. 4. De BPM levenscyclus toont zeven fases van een bedrijfsproces en hun overeenkomende informatiesyste(e)m(en). In de *(her)ontwerpfase* wordt een nieuw model gemaakt of een bestaand model aangepast. In de *analysefase* wordt een kandidaatmodel en zijn alternatieven geanalyseerd. Na de *(her)ontwerpfase* wordt het model geïmplementeerd (*implementatiefase*) of wordt een bestaand systeem ge(her)configureerd (*herconfiguratiefase*). In de *uitvoeringsfase* wordt het ontworpen model in gebruik genomen. Tijdens de uitvoeringsfase wordt een proces gemonitord. Bovendien kunnen kleine aanpassingen uitgevoerd worden zonder herontwerp in de proces(*aanpassingsfase*). In de *diagnosefase* wordt het in gebruik genomen proces geanalyseerd en de output van deze fase kan een nieuwe herontwerpfase voor het proces veroorzaken. Process mining is een waardevolle tool voor de meeste fases in Fig. 4. Het is duidelijk dat de diagnosefase voordeel haalt uit process mining, maar process mining beperkt zich niet tot deze fase. In de uitvoeringsfase kunnen bijvoorbeeld process mining technieken gebruikt worden voor *operationele ondersteuning*. Voorspellingen en



Figuur 5: Het L\* levenscyclusmodel beschrijft een process mining project bestaande uit vijf stadia: plan en verdedig (Stadium 0), destilleren (Stadium 1), control-flow model opbouwen en koppelen aan de event log (Stadium 2), een geïntegreerd procesmodel samenstellen (Stadium 3) en operationele ondersteuning bieden (Stadium 4).

aanbevelingen gebaseerd op modellen opgebouwd met historische informatie kunnen gebruikt worden om het gedrag van lopende gevallen te sturen. Gelijkaardige vormen voor het ondersteunen van beslissingen kunnen gebruikt worden bij (her)configuratie en (her)ontwerp.

Waar Fig. 4 de algemene BPM levenscyclus toont, focust Fig. 5 op concrete activiteiten en artefacten van process mining. Figuur 5 beschrijft de mogelijke stadia van een process mining project. Elk process mining project start met het plannen en verdedigen van deze planning (Stadium 0). Nadat het project gestart is, moeten event data, modellen,

doelstellingen en vragen worden afgeleid uit het systeem en de bedenkingen van domeinexperts en management (Stadium 1). Hiervoor moet de beschikbare data bekend zijn ("Welke data kunnen we gebruiken in de analyse?") en moet ook het domein voldoende bekend zijn ("Wat zijn de belangrijke vragen?"). Dit zal dan leiden tot de artefacten weergegeven in Fig. 5 (historische data, handgemaakte modellen, doelstellingen en vragen). In Stadium 2 wordt het control-flow model opgebouwd en gekoppeld aan de event log. Hiervoor kunnen ge-automatiseerde process discovery technieken gebruikt worden. Het ontdekte procesmodel kan reeds

antwoorden op sommige van de gestelde vragen geven en leiden tot herontwerp of aanpassingen. Bovendien kan de event log gefilterd of aangepast worden aan de hand van dit model (bijvoorbeeld verwijderen van uitzonderlijke activiteiten of speciale gevallen of het invoegen van ontbrekende events). Soms vraagt het heel wat inspanningen om events die bij eenzelfde procesuitvoering horen aan elkaar te koppelen. De overgebleven events worden dan gerelateerd aan entiteiten van het procesmodel. Wanneer het proces relatief gestructureerd verloopt, kan het control-flow model uitgebreid worden met andere perspectieven (bijvoorbeeld

Niveau	Karakterisering	Voorbeeld
★★★★★	Hoogste niveau: de event log is van excellente kwaliteit (d.w.z. betrouwbaar en compleet) en events zijn duidelijk gedefinieerd. Events worden op een automatische, systematische, betrouwbare en veilige manier geregistreerd. Privacy- en beveiligingsaspecten zijn adequaat in beschouwing genomen. Bovendien hebben de opgeslagen events (en alle bijbehorende attributen) een heldere semantiek. Dit impliceert dat er een of meer ontologieën bestaan waarnaar events en hun attributen naartoe verwijzen.	Semantisch geannoteerde logs van BPM systemen.
★★★★	Events worden automatisch en in een systematische en betrouwbare manier geregistreerd, d.w.z. logs zijn betrouwbaar en compleet. In tegenstelling tot de systemen die tot het *** niveau behoren zijn noties zoals process instance (case) en activiteit expliciet ondersteund.	De events logs van traditionele BPM/workflow systemen.
★★★	Events worden automatisch geregistreerd, maar er wordt geen systematische aanpak gevolgd om deze event logs op te slaan. Echter, in tegenstelling tot de logs die zich op niveau ** bevinden is er een bepaalde garantie dat de geregistreerde events overeenkomen met de realiteit (d.w.z. de event log is betrouwbaar maar niet noodzakelijk compleet). Een voorbeeld hiervan zijn de geregistreerde events van een ERP systeem. Desondanks dat events vanuit verschillende tabellen genomen dienen te worden, kan aangenomen worden dat de informatie correct is (het is betrouwbaar om aan te nemen dat een opgeslagen betaling in het ERP systeem ook daadwerkelijk bestaat en vice versa).	Tabellen in ERP systemen, event logs van CRM systemen, transactie logs van messaging systemen, event logs van high-tech systemen, enz.
★★	Event logs worden automatisch opgeslagen, d.w.z. als bijproduct van een informatiesysteem. Het bereik kan variëren, d.w.z. er is geen systematische aanpak gevolgd om te besluiten welke events geregistreerd worden. Bovendien is het mogelijk om het informatiesysteem te omzeilen. Zodoende kunnen events ontbreken of deze zijn niet op een correcte manier geregistreerd.	Event logs van document management systemen en product management systemen, registraties van foutmeldingen van embedded systemen, worksheets van servicemonteurs, enz.
★	Laagste niveau: event logs zijn van slechte kwaliteit. Het kan voorkomen dat geregistreerde event logs niet overeenkomen met de werkelijkheid en events kunnen ontbreken. Event logs die handmatig geregistreerd zijn, hebben normaliter deze karakteristieken.	Sporen die nagelaten zijn in papieren documenten die door een organisatie heengaan ("yellow notes"), papieren medische dossiers, enz.

Tabel 1: Volwassenheidsniveau's voor event logs.

data, tijd en resources) tijdens Stadium 3. De relatie tussen de event log en het model uit Stadium 2 kan gebruikt worden om het model uit te breiden (bijvoorbeeld tijdstippen of bijhorende events geven een schatting van de wachttijd voor activiteiten). Ook dit kan gebruikt worden om aanvullende vragen te beantwoorden en bijkomende acties te veroorzaken. Uiteindelijk worden de modellen die ontstonden in Stadium 3 gebruikt voor operationele ondersteuning (Stadium 4). Kennis uit historische event data wordt gecombineerd met lopende gevallen. Dit wordt gebruikt om het procesverloop aan te passen, te

voorspellen en aanbevelingen te formuleren. Stadia 3 en 4 kunnen enkel bereikt worden als het proces stabiel genoeg is en voldoende gestructureerd verloopt.

Er zijn nu technieken en tools die alle stadia van Fig. 5 ondersteunen, maar process mining is nog een relatief nieuw gegeven en de meeste van de huidige beschikbare gereedschappen zijn nog onvolwassen. Bovendien zijn beginnende gebruikers zich dikwijls niet bewust van de mogelijkheden en beperkingen van process mining. Daarom worden in dit manifest enkele richtlijnen (zie Sectie 3) en uitdagingen (zie Sectie 4) beschreven voor

gebruikers van process mining technieken, maar ook voor onderzoekers en ontwikkelaars die interesse hebben om de nieuwste technieken verder te ontwikkelen.

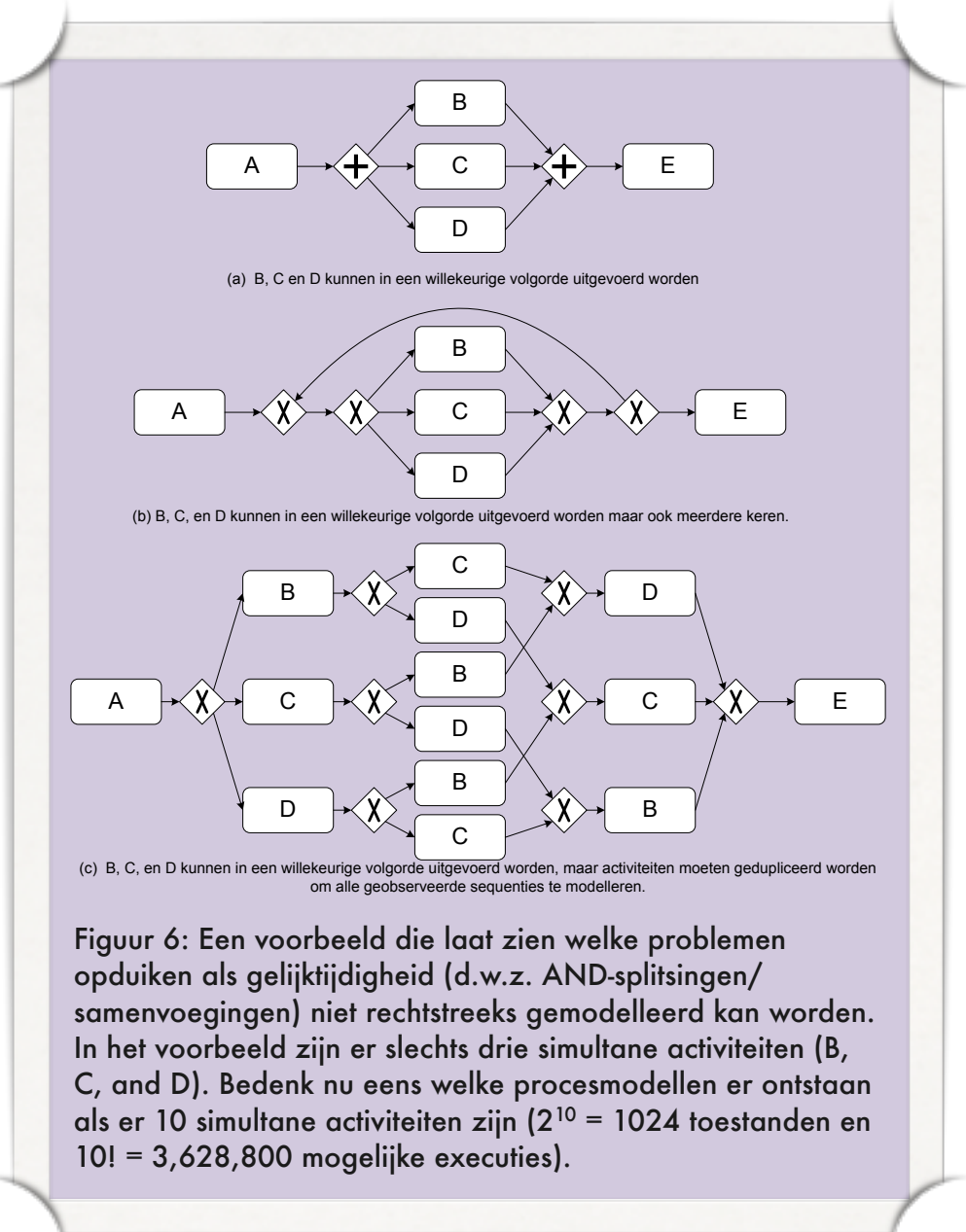
### 3. Richtlijnen

Zoals met elke nieuwe technologie zijn er onmiskenbare fouten die gemaakt kunnen worden als process mining in praktijksituaties toegepast wordt. Daarom presenteren we zes richtlijnen die gebruikers/analisten behoeden voor het maken van zulke fouten.

## GP1: Event Data Dienen Gezien te Worden als First-Class Citizens

Het startpunt voor elke process mining activiteit zijn de events die opgeslagen zijn. Collecties van events beschouwen we als *event logs*, echter, dit betekent niet dat events opgeslagen dienen te zijn in specifieke log bestanden. Deze mogen opgeslagen zijn in databasetabellen, mailarchieven, transactielogs en andere databronnen. Wat belangrijker is, is de *kwaliteit* van de event logs. De kwaliteit van een resultaat dat voortkomt uit process mining, wordt voor een zeer belangrijk deel bepaald door de input. Om deze reden dienen event logs gezien te worden als *first-class citizens* in de informatiesystemen die de processen ondersteunen die geanalyseerd worden. Helaas zijn event logs vaak een "bijproduct" die gebruikt worden voor debuggen of profilering. Bijvoorbeeld, de medische apparaten van Philips Healthcare slaan event logs alleen maar op omdat software ontwikkelaars "print statements" toegevoegd hebben aan de code. Hoewel er enkele informele richtlijnen bestaan voor het toevoegen van zulke statements aan de code is een systematischere aanpak vereist om de kwaliteit van event logs te verbeteren. Event data moeten gezien worden als een first-class citizen (i.p.v. second-class citizens).

Er zijn verschillende criteria om de kwaliteit van event data te beoordelen. Events dienen *betrouwbaar* te zijn, d.w.z. het is betrouwbaar om aan te nemen dat de opgeslagen events ook daadwerkelijk hebben plaatsgevonden en dat de attributen van de events correct zijn. Event logs dienen *compleet* te zijn, d.w.z. gegeven een bepaalde scope mogen er geen events ontbreken. Elk opgeslagen event dient een duidelijke *semantiek* te hebben. Bovendien, de event data dienen *veilig* te zijn wat wil zeggen dat zaken zoals privacy en beveiliging gegarandeerd zijn gedurende de opslag van events. Bijvoorbeeld, betrokkenen dienen op de hoogte te zijn van het type events dat gelogd wordt en de manier waarop deze zullen worden gebruikt. Tabel 1 geeft vijf event log volwassenheidsniveau's welke variëren van excellente kwaliteit (\*\*\*\*\*) tot slechte kwaliteit (\*). Bijvoorbeeld, de



Figuur 6: Een voorbeeld die laat zien welke problemen opduiken als gelijktijdigheid (d.w.z. AND-splitsingen/samenvoegingen) niet rechtstreeks gemodelleerd kan worden. In het voorbeeld zijn er slechts drie simultane activiteiten (B, C, and D). Bedenk nu eens welke procesmodellen er ontstaan als er 10 simultane activiteiten zijn ( $2^{10} = 1024$  toestanden en  $10! = 3,628,800$  mogelijke executies).

event logs van Philips Healthcare bevinden zich op niveau \*\*\*, d.w.z. events worden automatisch geregistreerd en het geregisteerde gedrag komt overeen met de werkelijkheid, maar er is geen systematische aanpak om semantiek toe te kennen aan events en dekking op een bepaald niveau te garanderen. Process mining kan toegepast worden op event logs van niveau \*\*\*\*\*, \*\*\*\* en \*\*\*. In principe is het ook mogelijk om process mining technieken toe te passen op event logs van niveau \*\* en \*. Echter, de analyse van zulke logs is voornamelijk problematisch en de resultaten zijn niet betrouwbaar. Verder heeft het niet veel zin om process mining toe te passen op logs van niveau \*.

Om daadwerkelijk de voordelen van process mining te kunnen ervaren zouden organisaties zich moeten

richten op event logs die zich op het hoogste kwaliteitsniveau bevinden.

## GP2: Log Destillatie Dient Gedreven te Zijn Door Vragen

Zoals weergegeven in Fig. 5, dienen process mining activiteiten gedreven te zijn door vragen. Zonder concrete vragen is het erg moeilijk om zinvolle event data te extraheren. Laten we bijvoorbeeld eens kijken naar de duizenden tabellen in de database van een ERP systeem zoals SAP. Zonder concrete vragen is het onmogelijk om die tabellen te selecteren die relevant zijn voor dataextractie.

Een procesmodel zoals is weergegeven in Fig. 1 beschrijft de levenscyclus van cases (d.w.z. procesinstanties) van een bepaald type. Derhalve dient het te analyseren type



van cases gekozen te worden alvorens process mining technieken toe te passen. Deze keuze moet gedreven zijn door de vragen die beantwoord moeten worden welke een niet-triviale keuze kan zijn. Laten we bijvoorbeeld eens kijken naar de afhandeling van klantorders. Elke klantorder kan bestaan uit meerdere orderregels doordat een klant meerdere producten kan bestellen in één order. Één klantorder kan resulteren in meerdere leveringen. Één levering kan verwijzen naar orderregels van meerdere orders. Zodoende bestaat er een many-to-many relatie tussen orders en leveringen en een one-to-many relatie tussen orders en orderregels. Voor een database met event data die gerelateerd is aan orders, orderregels en afleveringen, is het zodoende mogelijk om verschillende procesmodellen te ontdekken. Bijvoorbeeld, het is mogelijk om data te destilleren met het doel om de levenscyclus van individuele orders te beschrijven. Echter, het is ook mogelijk om data te destilleren die als doel heeft om de levenscyclus van afzonderlijke orderregels te ontdekken of de levenscyclus van afzonderlijke leveringen.

### GP3: Gelijktijdigheid, Keuze en Andere Basis Control-Flow Constructies Moeten Ondersteund Worden

Er bestaat een overvloed aan procesmodelletalen (b.v. BPMN, EPCs, Petri netten, BPEL, en UML activiteitendiagrammen). Enkele van deze talen bevatten vele modelleerelementen (b.v. BPMN biedt meer dan 50 verschillende grafische elementen aan) terwijl anderen erg elementair zijn (b.v. Petri netten bestaan uit maar drie verschillende elementen: plaatsen, transities en pijlen). De control-flow beschrijving is de ruggengraat van elk procesmodel.

Basiswerkstroomconstructies (ook bekend als *patronen*) die ondersteund worden door alle toonaangevende talen zijn sequentie, parallelle routing (AND-splitsingen/samenvoegingen), keuze (XOR-splitsingen/samenvoegingen) en lussen. Het is duidelijk dat deze patronen ondersteund dienen te worden door process mining technieken. Echter,

enkele technieken zijn niet in staat om om te gaan met gelijktijdigheid en ondersteunen alleen Markov ketens/transitie systemen.

Figuur 6 illustreert wat er gebeurt als process mining technieken niet in staat zijn om gelijktijdigheid te ontdekken (geen AND-splitsingen/samenvoegingen). Beschouw nu de event log  $L = \{ \langle A, B, C, D, E \rangle, \langle A, B, D, C, E \rangle, \langle A, C, B, D, E \rangle, \langle A, C, D, B, E \rangle, \langle A, D, B, C, E \rangle, \langle A, D, C, B, E \rangle \}$ .  $L$  die cases bevat die starten met A en eindigen met E. Activiteiten B, C, en D komen in elke volgorde voor tussen A and E. Het BPMN model in Fig. 6(a) laat het onderliggende proces op een compacte manier zien door gebruik te maken van twee AND gateways. Neem nu aan dat AND gateways niet ondersteund worden door de process mining techniek. In dat geval zijn de andere twee BPMN modellen in Fig. 6 logische kandidaten. Het BPMN model in Fig. 6 is compact maar laat te veel gedrag toe. (bijvoorbeeld cases zoals  $\langle A, B, B, B, E \rangle$  zijn toegestaan volgens het model maar zijn niet waarschijnlijk als we naar de log kijken.) Alle cases in  $L$  zijn mogelijk in het BPMN model in Fig. 6(c). Echter, alle sequenties zijn expliciet beschreven, wat dus als nadeel heeft dat de log niet op een compacte manier gerepresenteerd is. Het voorbeeld laat zien dat voor realistische modellen die veel simultane activiteiten bevatten, de bijbehorende modellen ernstig lijden aan een zogenaamde "underfit" en/of extreem complex zijn, in het geval gelijktijdigheid niet ondersteund wordt.

Zoals geïllustreerd in Fig. 6 is het belangrijk dat op zijn minst de basiswerkstroompatronen ondersteund worden. Hiernaast is het ook wenselijk dat OR splitsingen/samenvoegingen ondersteund worden omdat deze het mogelijk maken dat inclusieve beslissingen en gedeeltelijke synchronisaties compact weergegeven kunnen worden.

### GP4: Events Dienen Gerelateerd te Zijn Aan Model Elementen

Zoals aangegeven in Sectie 2, is het een misvatting dat process mining gelimiteerd is tot control-flow ontdekking. Zoals weergegeven in

Fig. 1, kan het ontdekte model verschillende perspectieven bevatten (organisationele perspectief, tijdsperspectief, dataperspectief enz.). Bovendien is "discovery (ontdekken)" maar één van de drie types van process mining zoals weergegeven in Fig. 3. De andere twee process mining types ("conformance checking (controleren)" en "enhancement (verbeteren)") zijn sterk gebaseerd op de relatie tussen *elementen in het model* en *events in de log*. Deze relatie kan gebruikt worden om de event log in het model "na te spelen" (replay). Dit naspelen kan gebruikt worden om discrepanties te ontdekken tussen de event log en een model, bijvoorbeeld, enkele events in de log zijn niet mogelijk volgens het model. Technieken voor conformance checking (controleren) kwantificeren en diagnosticeren deze discrepanties. Tijdstempels in de event log kunnen gebruikt worden om het tijdsgedrag gedurende het naspelen te analyseren. Tijdsverschillen tussen causaal gerelateerde activiteiten kunnen gebruikt worden om verwachte wachttijden toe te voegen aan het model. Deze voorbeelden laten zien dat de relatie tussen events in de log en elementen in het model als startpunt kunnen dienen voor verschillende type analyses.

In sommige gevallen kan het niet-triviaal zijn om zulke relaties te bewerkstelligen. Bijvoorbeeld, een event verwijst naar twee verschillende activiteiten of het is onduidelijk bij welke activiteit het hoort. Deze ambigüiteiten dienen opgelost te worden, om zodoende de process mining resultaten fatsoenlijk te kunnen interpreteren. Naast het probleem van het relateren van events aan activiteiten, is er ook het probleem van het relateren van events aan process instances. Dit wordt gewoonlijk *event correlatie* genoemd.

### GP5: Modellen Dienen Behandeld te Worden als Betekenisvolle Abstracties van de Realiteit

Modellen afgeleid van event data laten een *weergave van de realiteit* zien. Zulke weergaven dienen een betekenisvolle abstractie van het gedrag, bevat in de event log, te laten

zien. Voor een event log kunnen meerder zinvolle weergaven mogelijk zijn. Bovendien, belanghebbenden kunnen geïnteresseerd zijn in andere weergaven. Feitelijk dienen modellen die ontdekt zijn vanuit event logs gezien te worden als "kaarten" (zoals geografische kaarten). Dit richtinggevende principe zorgt voor belangrijke inzichten waarvan twee in het vervolg beschreven worden.

Allereerst, het is belangrijk dat er niet zoiets is als "de kaart" voor een bepaald geografisch gebied. Afhankelijk van het uiteindelijke gebruik zijn er verschillende kaarten: landkaarten, kaarten voor trektochten, fietskaarten, enz. Elk van deze kaarten laat een bepaalde weergave van dezelfde realiteit zien en het zou absurd zijn om aan te nemen, dat er zoiets zou zijn als "de perfecte kaart". Hetzelfde geldt voor procesmodellen: het model moet die dingen benadrukken die relevant zijn voor een bepaald type gebruiker. Ontdekte modellen kunnen zich richten op verschillende perspectieven (control-flow, data stroom, tijd, resources, kosten, enz.) en deze laten zien op verschillende niveaus van granulariteit en precisie, b.v., een manager wil een ruw informeel model zien dat zich richt op de kosten, terwijl een procesanalist een gedetailleerd model wil zien dat zich richt op afwijkingen van de normale flow. Merk op dat verschillende belanghebbenden interesse kunnen hebben in een weergave van het proces op verschillende niveau's: *strategisch niveau* (beslissingen op dit niveau hebben een lange-termijn effect en zijn gebaseerd op geaggregeerde data over een langere periode), *tactisch niveau* (beslissingen op dit niveau hebben een gemiddelde-termijn effect en zijn meestal gebaseerd op recente data), en *operationeel niveau* (beslissingen op dit niveau hebben een direct effect en zijn gebaseerd op event data gerelateerd aan lopende cases).

Ten tweede, het is zinvol om ideeën van cartografie in beschouwing te nemen als het gaat om het produceren van begrijpelijke kaarten. Bijvoorbeeld, landkaarten abstraheren van minder belangrijke wegen en steden. Minder belangrijke dingen zijn eruit gelaten of dynamisch geclusterd in geaggregeerde vormen (b.v. straten en buitenwijken vloeien over in steden).

Cartografen elimineren niet alleen irrelevante details maar gebruiken ook kleuren om belangrijke dingen te benadrukken. Bovendien hebben grafische elementen een bepaalde grootte om hun belangrijkheid aan te geven (b.v. de grootte van lijnen en punten kan variëren). Geografische kaarten hebben ook een duidelijke interpretatie van de x-as en y-as, d.w.z. de layout van een kaart is niet willekeurig aangezien de coördinaten van elementen een betekenis hebben. Voor al het bovengenoemde is er een duidelijke tegenstelling met de huidige toonaangevende procesmodellen. Voor deze modellen wordt normalerwijze geen gebruik gemaakt van kleur, grootte, en locatie-elementen om deze begrijpelijker te maken. Echter, ideeën vanuit de cartografie kunnen makkelijk ingepast worden in de constructie van ontdekte proceskaarten. Bijvoorbeeld, de grootte van een activiteit kan gebruikt worden om de frequentie van de activiteit weer te geven of een andere eigenschap die de belangrijkheid van de activiteit weergeeft (b.v. kosten of gebruik van resources). De breedte van een pijl kan de belangrijkheid van de bijbehorende causale relatie weergeven en de kleur van de pijl kan gebruikt worden om bottlenecks weer te geven.

De bovengenoemde observaties laten zien, dat het belangrijk is om de juiste representatie te kiezen en deze te fine-tunen voor het publiek waarvoor deze bedoeld is. Dit is belangrijk voor de visualisatie van de resultaten naar de eindgebruikers toe en om ontdekkingsalgoritmen naar geschikte modellen toe te laten leiden (zie ook Challenge C5).

## GP6: Process Mining Dient Een Continu Proces te Zijn

Process mining kan helpen in het leveren van betekenisvolle "kaarten" die direct verbonden zijn met event data. Historische event data en huidige event data kan geprojecteerd worden op zulke modellen. Bovendien kunnen processen veranderen terwijl zij worden geanalyseerd. Gegeven de dynamische natuur van processen is het niet aan te raden om process mining te zien als een eenmalige activiteit. Het doel zou niet moeten zijn om een vast model te creëren maar om leven te blazen in procesmodellen zodat

gebruikers en analisten worden aangemoedigd om deze dagelijks te bestuderen.

Vergelijk dit met het gebruik van mashups tijdens geotagging. Via Google Maps zijn er duizenden mashups (b.v. applicaties die informatie over verkeer projecteren, onroerend goed, fastfood restaurants, of filmtijden). Mensen kunnen naadloos in- en uitzoomen op deze kaarten en er mee interacteren (b.v. verkeersopstoppingen worden geprojecteerd op de kaart en gebruikers kunnen een bepaald probleemgebied selecteren om hiervan de details te zien). Het moet ook mogelijk zijn om process mining uit te voeren op basis van real-time data. Door middel van de "kaart metafoor" kunnen we aannemen dat events GPS-coördinaten hebben die real-time geprojecteerd kunnen worden op kaarten. Overeenkomstig met autonavigatiesystemen kunnen process mining gereedschappen gebruikers helpen (a) met navigatie door processen, (b) door dynamische informatie op proceskaarten te projecteren (b.v. visualiseren van "verkeersopstoppingen" binnen procesmodellen) en (c) voorspellingen te leveren over lopende cases (b.v. een schatting van de "aankomsttijd" van een vertrapte case). Deze voorbeelden laten zien dat het zonde is om procesmodellen niet op een actievere manier te gebruiken. Daarom moet process mining als een continu proces gezien worden die navolgbare informatie geeft voor verschillende tijds aanduidingen (minuten, uren, weken, en maanden).

## 4. Uitdagingen

Process mining is een belangrijk gereedschap voor moderne organisaties die niet-triviale operationele processen dienen te beheersen. Aan de ene kant is er een onvoorstelbare groei van event data. Aan de andere kant dienen processen en informatie perfect uitgelijnd te worden om zodoende aan compliance, efficiëntie en customer service voorwaarden te voldoen. Ondanks de toepasbaarheid van proces mining zijn er nog steeds belangrijke uitdagingen die aangepakt dienen te worden; deze illustreren dat process mining een opkomende discipline is. In het vervolg

## Uitdagingen:

C1: Het Vinden van, Samenvoegen van en Opschonen van Event Data

C2: Omgaan met Complexe Event Logs met Verschillende Karakteristieken

C3: Het Creëren van Representatieve Benchmarks

C4: Omgaan met Concept Drift

C5: Verbetering van de Representational Bias (weergave voorkeur) gebruikt voor Process Discovery (ontdekken)

C6: Balanceren Tussen Kwaliteitscriteria zoals Fitness, Eenvoud, Precisie en Generalisatie

C7: Cross-Organizational Mining

C8: Het Bieden van Operationele Ondersteuning

C9: Het Combineren van Process Mining Met Andere Vormen van Analyse

C10: Het Verbeteren van Bruikbaarheid voor Niet-Deskundigen

C11: Het Verbeteren van Begrijpelijkheid voor Niet-Deskundigen

zullen we enkele van deze uitdagingen opnoemen. Merk op dat het niet de intentie is, dat deze lijst compleet is. In de toekomst kunnen nieuwe uitdagingen zich voordoen of bestaande uitdagingen verdwijnen vanwege vorderingen op het gebied van process mining.

### C1: Het Vinden van, Samenvoegen van en Opschonen van Event Data

Aanzienlijke inspanningen zijn nodig om event data te destilleren die geschikt zijn voor process mining. Er

zijn verschillende hordes die genomen moeten worden:

- Gegevens kunnen *verdeeld* zijn over meerdere bronnen. Deze informatie dient samengevoegd te worden. Dit heeft de neiging om problematisch te zijn in het geval verschillende identifiers gebruikt zijn in de verschillende gegevensbronnen. Bijvoorbeeld, een systeem gebruikt naam en geboortedatum om een persoon te identificeren terwijl een ander systeem hiervoor het burgerservicenummer (BSN) / rijksregisternummer (RRNR) gebruikt.
- Event data is vaak "objectcentrisch" in plaats van "procescentrisch". Bijvoorbeeld, afzonderlijke producten, pallets en containers hebben RFID tags en events verwijzen naar deze tags. Echter, om nu een afzonderlijke klantorder te kunnen volgen, dienen deze objectcentrische events samengevoegd en voorberekt te worden.
- Event data kan *incompleteet* zijn. Een algemeen probleem is, dat events niet expliciet verwijzen naar process instances. Het is vaak mogelijk om deze informatie af te leiden, echter dit kan veel inspanning vereisen. Ook kan voor enkele events tijdsinformatie ontbreken. Dit vereist weer dat tijdstempels geïnterpoleerd worden om toch gebruik te kunnen maken van de aanwezige tijdsinformatie.
- Een event log kan *outliers* bevatten, d.w.z. exceptioneel gedrag wat ook wel eens *ruis* wordt genoemd. Hoe nu deze outliers te definiëren? Hoe dienen deze gedetecteerd te worden? Deze vragen dienen beantwoord te worden om zodoende event data te kunnen zuiveren.
- Logs kunnen events op *verschillende niveaus van granulariteit* bevatten. In de event log van een informatiesysteem van een ziekenhuis kunnen events verwijzen naar een eenvoudig bloedonderzoek of naar complexe chirurgische procedures. Ook kunnen tijdstempels verschillende granulariteiten hebben die varieert van een precisie op milliseconden (28-9-2011:h11m28s32ms342) tot ruwe datum informatie (28-09-2011).
- Events vinden plaats in een *specifieke context* (weer, werkbelasting, dag van de week, enz.). Deze context kan bepaalde fenomenen verklaren,

b.v., de responstijd is langer dan gebruikelijk vanwege de huidige hoeveelheden werk of vakanties. Voor de analyse is het gewenst om deze context op te nemen. Dit impliceert de samenvoeging van event data met contextuele gegevens. Echter, hier duidt de "dimensionaliteitsvervloeking" op omdat analyse lastig wordt vanwege het toevoegen van teveel variabelen.

Betere gereedschappen en methodes zijn nodig om bovenstaande problemen aan te pakken. Echter, zoals eerder aangegeven, organisaties dienen event logs te zien als first-class citizens in plaats van als een bijproduct. Het doel is om \*\*\*\*\* event logs te verkrijgen (zie Tabel 1). De opgedane lessen vanuit de context van datawarehouses zijn bruikbaar om zodoende een hoge kwaliteit van event logs te kunnen garanderen. Bijvoorbeeld, eenvoudige controles tijdens het invoeren van gegevens kunnen helpen om de hoeveelheid aan incorrecte event data significant te kunnen reduceren.

### C2: Omgaan met Complexe Event Logs met Verschillende Karakteristieken

Event logs kunnen verschillende karakteristieken hebben. Event logs kunnen zeer groot zijn wat het moeilijk maakt om hier mee om te gaan, terwijl andere event logs zo klein zijn dat niet genoeg data voorhanden is om betrouwbare conclusies te trekken.

In sommige domeinen worden verbazingwekkende hoeveelheden aan data geregistreerd. Daarom zijn additionele inspanningen nodig om prestatie en schaalbaarheid te verbeteren. Bijvoorbeeld, ASML bewaakt continu al haar waferscanners. Deze waferscanners worden gebruikt bij verschillende organisaties (b.v. Samsung en Texas Instruments) om chips te produceren (ongeveer 70% van de chips worden geproduceerd door de waferscanners van ASML). Bestaande gereedschappen hebben moeite om om te gaan met de petabytes aan data die worden vergaard in zulke domeinen. Naast de aantallen events die geregistreerd worden, zijn er andere karakteristieken zoals het gemiddelde aantal events per case,

vergelijkbaarheid tussen cases, het aantal unieke events en het aantal unieke paden. Beschouw een event log L1 met de volgende karakteristieken: 1000 cases, gemiddeld 10 events per case, en weinig variatie (b.v. verscheidene cases volgen dezelfde of zeer vergelijkbare paden). Event log L2 bevat maar 100 cases, maar gemiddeld zijn er 100 events per case en alle cases volgen een uniek pad. Het is duidelijk dat L2 veel moeilijker is om te analyseren dan L1 ondanks dat de logs een vergelijkbare grootte hebben (ongeveer 10.000 events).

Een event log bevat alleen maar voorbeeldgedrag, het is niet zeker dat deze compleet is. Process mining technieken dienen om te gaan met het gegeven dat een event log incompleet is door gebruik te maken van een "open wereld aanname": het feit dat iets niet gebeurt is, betekent niet dat dit niet kan gebeuren. Dit maakt het een uitdaging om om te gaan met kleine event logs die veel variëteit bevatten.

Zoals eerder gezegd, sommige logs bevatten events die zich op een erg laag niveau van abstractie bevinden. Deze logs hebben de neiging om erg groot te zijn en de afzonderlijke laag-niveau events zijn niet erg interessant voor de betrokkenen. Daarom wil men graag laag-niveau events aggregeren tot hoog-niveau events. Bijvoorbeeld, gedurende de analyse van het diagnostische en behandelingsproces van een groep patiënten hoeft men niet persé geïnteresseerd te zijn in de afzonderlijke bloedonderzoeken die opgeslagen zijn in het lab-informatiesysteem in het ziekenhuis.

Op dit moment dienen organisaties een proefondervindelijke methode te gebruiken om uit te vinden of een event log geschikt is voor process mining. Daarom zouden gereedschappen het mogelijk moeten maken om een snelle geschiktheidstest uit te kunnen voeren voor een bepaalde dataset. Zo'n test dient potentiële prestatieproblemen aan te geven en te waarschuwen voor logs die te incompleet zijn of te gedetailleerd.

### C3: Het Creëren van Representatieve Benchmarks

Process mining is een opkomende technologie. Dit verklaart waarom

goede benchmarks nog ontbreken. Bijvoorbeeld, vele process mining technieken zijn beschikbaar en verschillende verkopers bieden verschillende producten aan, maar er is nog geen overeenstemming over de kwaliteit van deze technieken. Ondanks het feit, dat er grote verschillen zijn in functionaliteit en prestatie, is het moeilijk om de verschillende technieken en gereedschappen te vergelijken. Daarom is het nodig om goede benchmarks bestaande uit voorbeeld-datasets en representatieve kwaliteitscriteria te ontwikkelen.

Voor klassieke data mining technieken zijn goede benchmarks beschikbaar. Deze benchmarks hebben providers van gereedschappen en onderzoekers gestimuleerd om de performance van hun technieken te verbeteren. Voor process mining is dit een grotere uitdaging. Bijvoorbeeld, het relationele model van Codd uit 1969 is simpel en geniet veel steun. Als gevolg daarvan kost het weinig moeite om gegevens te converteren van de ene database naar de andere en er zijn geen interpretatieproblemen. Voor processen ontbreekt een dergelijk eenvoudige model. Voorgestelde standaarden voor procesmodellering zijn veel ingewikkelder en slechts enkele leveranciers ondersteunen precies dezelfde set van concepten. Processen zijn gewoon complexer dan gegevens in tabelvorm.

Toch is het belangrijk om representatieve benchmarks voor proces mining te creëren. Enig initieel werk is al beschikbaar. Bijvoorbeeld, er zijn verschillende metrieken voor het meten van de kwaliteit van proces mining resultaten (fitness, eenvoud, precisie en generalisatie). Verder zijn er verschillende event logs voor het publiek beschikbaar (zie [www.processmining.org](http://www.processmining.org)). Zie bijvoorbeeld de gebruikte event log voor de eerste Business Process Intelligence Challenge (BPIC'11) georganiseerd door de Task Force (zie doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54).

Aan de ene kant, moeten er benchmarks op basis van real-life data sets zijn. Aan de andere kant, is er de noodzaak om synthetische datasets met specifieke kenmerken vast te leggen. Dergelijke synthetische datasets helpen om process mining technieken te ontwikkelen, die zijn toegesneden op

onvolledige event logs, ruisgevoelige event logs, of specifieke populaties van processen.

Behalve de ontwikkeling van representatieve benchmarks, moet er ook meer consensus zijn, over de criteria welke gebruikt worden om de kwaliteit van de proces mining resultaten (zie ook Challenge C6) te beoordelen. Bovendien kunnen *cross-validatie* technieken uit data mining worden aangepast om het resultaat te beoordelen. Overweeg bijvoorbeeld, *k-fold* controle. Men kan de event log splitsen in *l* delen. *k-1* onderdelen kunnen worden gebruikt om een procesmodel te leren en controle technieken kunnen worden gebruikt om het resultaat ten opzichte van het resterende deel te beoordelen. Dit kan *k* keer worden herhaald, en zo enig inzicht geven in de kwaliteit van het model.

### C4: Omgaan met Concept Drift

De term *concept drift* verwijst naar de situatie waarin het proces aan het veranderen is, terwijl het wordt geanalyseerd. Bijvoorbeeld, in het begin van het event log kunnen twee activiteiten gelijktijdig worden uitgevoerd, terwijl later in de log de uitvoering van deze activiteiten sequentieel wordt. Processen kunnen veranderen als gevolg van periodieke / seizoensveranderingen (bijvoorbeeld, "in december is er meer vraag" of "op vrijdagmiddag zijn er minder werknemers beschikbaar") of als gevolg van veranderende omstandigheden (bijvoorbeeld, "de markt wordt steeds concurrerder"). Dergelijke veranderingen hebben gevolgen voor processen en het is belangrijk om ze op te sporen en te analyseren. Concept drift in een proces kan worden ontdekt, door de event log op te splitsen in kleinere logs en het analyseren van de "footprints" van de kleinere logs. Een dergelijke "tweede orde" analyse vergt veel meer event data. Toch bevinden weinig processen zich in een "steady-state" en begrip van concept drift is van het grootste belang voor de het managen van processen. Daarom is aanvullend onderzoek en ondersteunend gereedschap nodig, om concept drift adequaat te analyseren.

## C5: Verbetering van de Representational Bias (weergave voorkeur) gebruikt voor Process Discovery (ontdekken)

Een process discovery (ontdekken) techniek produceert een model met behulp van een bepaalde taal (bijvoorbeeld, BPMN of Petrinetten). Toch is het belangrijk om de visualisatie van het resultaat te scheiden van de weergave gebruikt tijdens het eigenlijke discovery (ontdekken) proces. De selectie van een doeltaal omvat vaak een aantal impliciete veronderstellingen. Het beperkt de zoekruimte; processen die niet kunnen worden weergegeven door de doeltaal kunnen niet worden ontdekt. Deze zogenaamde "representational bias" (weergave voorkeur) gebruikt tijdens het discovery (ontdekken) proces moet een bewuste keuze zijn en mag niet (alleen) worden veroorzaakt door de geprefereerde grafische weergave.

Beschouw bijvoorbeeld Fig. 6: de vraag of de doeltaal zorgt voor concurrency of niet, mag geen effect hebben op zowel de visualisatie van het ontdekte model, als de soort beschouwde modellen door het algoritme. Als de "representational bias" (weergave voorkeur) concurrency niet toestaat (Fig. 6 (a) is niet mogelijk) en dat meerdere activiteiten hetzelfde label hebben (Fig. 6 (c) is niet mogelijk), dan zijn alleen problematische modellen zoals die getoond in Fig. 6 (b) mogelijk. Dit voorbeeld laat zien, dat een meer zorgvuldige en nauwkeurige selectie van de "representational bias" (weergave voorkeur) nodig is.

## C6: Balanceren Tussen Kwaliteitscriteria zoals Fitness, Eenvoud, Precisie en Generalisatie

Event logs zijn vaak verre van compleet, dat wil zeggen, alleen voorbeeld gedrag is gegeven. Procesmodellen zorgen meestal voor een exponentiële of zelfs een oneindig aantal verschillende traces (in geval van lussen). Bovendien kunnen sommige traces met een veel lagere kans voorkomen, dan anderen. Daarom is het niet realistisch aan te

nemen, dat alle mogelijke traces aanwezig zijn in de event log. Om te illustreren, dat het onpraktisch is om volledige logs als iets vanzelfsprekend te beschouwen, overweeg een proces dat bestaat uit 10 activiteiten die parallel kunnen worden uitgevoerd en een bijbehorend log dat informatie bevat over ongeveer 10.000 gevallen. Het totaal aantal van mogelijke "interleavings" (combinaties / variaties) in het model met 10 gelijktijdige activiteiten wordt  $10! = 3.628.800$ . Vandaar dat het onmogelijk is, dat elke interleaving (combinatie / variatie) aanwezig is in de log, aangezien er minder cases (10.000) dan mogelijke traces (3628800) zijn. Zelfs als er miljoenen cases in log zitten, is het uiterst onwaarschijnlijk, dat alle mogelijke variaties aanwezig zijn. Een extra complicatie is dat sommige alternatieven minder vaak voorkomen dan andere. Deze kunnen worden beschouwd als "ruis". Het is onmogelijk om een redelijk model te bouwen voor dergelijk ruisgevoelig gedrag. Het ontdekte model moet hiervan abstraheren; het is beter om laag frequent gedrag te onderzoeken met behulp van conformance checking (controleren).

Ruis en onvolledigheid maken process discovery (ontdekken) een uitdagend probleem. In feite zijn er vier concurrerende kwaliteitsdimensies: (a) fitness (b) eenvoud, (c) precisie, en (d) generalisatie. Een model met een goede *fitness*, zorgt voor het grootste deel van het gedrag dat gezien wordt in het event log. Een model heeft een perfecte fitness als alle traces in de log, kunnen worden afgespeeld. Het *eenvoudigste* model dat het geobserveerde gedrag in de log kan verklaren, is het beste model. Dit principe staat bekend als Occam's Razor. Fitness en eenvoud alleen zijn niet voldoende om de kwaliteit van het ontdekte procesmodel te beoordelen. Bijvoorbeeld, het is heel gemakkelijk om een zeer eenvoudig Petrinet te maken ("flower model"), dat in staat is om alle traces af te spelen in een event log (maar ook ieder ander event log met betrekking tot dezelfde set van activiteiten). Ook is het ongewenst, om een model te hebben, dat alleen het exacte gedrag uit de event log toestaat. Vergeet niet dat de log alleen voorbeeldgedrag bevat en dat vele traces die mogelijk kunnen zijn, nog

niet gezien zijn. Een model is *precies* als het niet "teveel" gedrag toestaat. Overduidelijk mist het "flower model" precisie. Een model dat niet nauwkeurig is, is "underfitting". Underfitting treedt op, wanneer het model het voorbeeldgedrag in de log over-generaliseert (dat wil zeggen, het model staat gedrag toe, wat sterk verschilt van wat zichtbaar is de log). Een model mag generaliseren en het gedrag niet beperken tot alleen maar de voorbeelden, welke te zien zijn in de log. Een model dat niet generaliseert is "overfitting". Overfitting is het probleem dat een zeer specifiek model wordt gegenereerd, terwijl het duidelijk is dat de log alleen voorbeeldgedrag in zich heeft (dat wil zeggen, het model verklaart de specifieke voorbeeldlog, maar een volgende voorbeeldlog van hetzelfde proces, kan een volledig ander procesmodel opleveren).

Het balanceren tussen fitness, eenvoud, precisie en generalisatie is uitdagend. Dit is de reden dat de meeste van de meer krachtige process discovery (ontdekken) technieken verschillende parameters bieden. Verbeterde algoritmen moeten worden ontwikkeld, om de vier concurrerende dimensies van kwaliteit beter te balanceren. Bovendien, alle gebruikte parameters moeten te begrijpen zijn door eindgebruikers.

## C7: Cross-Organizational Mining

Traditioneel wordt process mining toegepast binnen een enkele organisatie. Echter, als service technologie, supply-chain integratie en cloud computing steeds verder verspreid raken, zijn er scenario's waar de event logs van meerdere organisaties beschikbaar komen voor analyse. In principe zijn er twee situaties voor *cross-organisatorische process mining*.

Allereerst kunnen we kijken naar de gezamenlijke omgeving, waar verschillende organisaties samenwerken om meerdere procesinstanties te behandelen. Men kan zo'n cross-organisatorisch proces vergelijken met een "legpuzzel": dat wil zeggen, het totale proces wordt opgedeeld in stukjes en verdeeld over organisaties, die moeten samenwerken om de cases succesvol te voltooien. Het analyseren van de event log binnen

een van deze betrokken organisaties is onvoldoende. Om de end-to-end processen te ontdekken, moeten de event logs van verschillende organisaties worden samengevoegd. Dit is geen triviale taak, aangezien events gecorreleerd moeten worden over organisatorische grenzen heen.

Ten tweede kunnen we ook kijken naar een omgeving waarin verschillende organisaties in essentie hetzelfde proces uitvoeren, waarbij ervaringen, kennis of een gemeenschappelijke infrastructuur gedeeld worden. Beschouw bijvoorbeeld Salesforce.com. De verkoopprocessen van veel organisaties worden beheerd en ondersteund door Salesforce. Aan de ene kant, delen deze organisaties een infrastructuur (processen, databases, enz.). Aan de andere kant zijn ze niet gedwongen om een stringent procesmodel te volgen, aangezien het systeem kan worden geconfigureerd, om varianten van hetzelfde proces te ondersteunen. Een ander voorbeeld zijn de basisprocessen uitgevoerd binnen een gemeente (bijvoorbeeld de afgifte van bouwvergunningen). Hoewel alle gemeenten in een land dezelfde basis set van processen moeten ondersteunen, kunnen er ook verschillen zijn. Het is natuurlijk interessant om dergelijke variaties tussen verschillende organisaties te analyseren. Deze organisaties kunnen van elkaar leren, dienstverleners kunnen hun diensten verbeteren en diensten met toegevoegde waarde leveren, op basis van de resultaten van cross-organisatorische process mining.

Voor beide typen van cross-organisatorische process mining moeten nieuwe analysetechnieken worden ontwikkeld. Deze technieken dienen ook rekening te houden met privacy- en veiligheidskwesties. Organisaties zijn niet snel genegen, wegens concurrentieredenen of een gebrek aan vertrouwen, informatie te delen. Daarom is het belangrijk, om proces mining technieken te ontwikkelen gericht op het behoud van privacy.

## C8: Het Bieden van Operationele Ondersteuning

Aanvankelijk lag de focus van process mining op de analyse van historische gegevens. Vandaag de dag echter

worden veel gegevensbronnen (bijna real-time bijgewerkt en is er voldoende rekenkracht beschikbaar om gebeurtenissen te analyseren op het moment dat zij zich voordoen. Daarom moet process mining niet beperkt blijven tot off-line analyse en ook gebruikt worden voor online operationele ondersteuning. Drie soorten operationele ondersteuningsactiviteiten kunnen worden geïdentificeerd: *detect* (detecteren), *predict* (voorspellen) en *recommend* (aanbevelen). Op het moment dat een case (zaak) afwijkt van het vooraf gedefinieerde proces, kan dit worden gedetecteerd en kan het systeem een signaal genereren. Vaak wil men een dergelijke notificatie onmiddellijk kunnen genereren (om nog in staat te zijn het procesverloop te kunnen beïnvloeden) en niet pas achteraf. Historische gegevens kunnen worden gebruikt om voorspellende modellen te maken. Deze kunnen worden gebruikt om lopende procesinstanties te begeleiden. Het is dan bijvoorbeeld mogelijk, om de resterende doorlooptijd van een case (zaak) te voorspellen. Op basis van dergelijke voorspellingen kan men ook recommender (aanbeveling) systemen maken, die bepaalde acties voorstellen om kosten te verlagen of doorlooptijd te verkorten. Het toepassen van process mining technieken in een dergelijke online setting, zorgt voor extra uitdagingen op het gebied van rekenkracht en datakwaliteit.

## C9: Het Combineren van Process Mining Met Andere Vormen van Analyse

Operations management en in het bijzonder operations research, is een tak van management wetenschappen, welke sterk leunt op modellering. Hierbij wordt een verscheidenheid aan wiskundige modellen, variërend van lineaire programmering, projectplanning tot wachtrijmodellen, Markov-ketens en simulatie, gebruikt. Data mining kan worden gedefinieerd als "de analyse van (vaak grote) databestanden om onvermoede relaties te vinden en om de gegevens op nieuwe manieren samen te vatten, zodat deze zowel begrijpelijk als nuttig zijn voor de gegevenseigenaar". Een grote verscheidenheid aan technieken zijn ontwikkeld: classificatie

(bijvoorbeeld decision tree learning (leren via beslisbomen)), regressie, clustering (bijvoorbeeld k-means clustering) en pattern discovery (ontdekken van patronen) (bijvoorbeeld, association rule learning).

Beide gebieden (operations management en data mining) voorzien in waardevolle analysetechnieken. De uitdaging is om deze analysetechnieken te combineren met process mining. Denk bijvoorbeeld aan simulatie. Process mining technieken kunnen worden gebruikt om een simulatiemodel te leren op basis van historische gegevens. Vervolgens kan het simulatiemodel worden gebruikt voor operationele ondersteuning. Vanwege de sterke relatie tussen event log en model, kan het model gebruikt worden om de geschiedenis af te spelen en kan worden gestart met simulaties van de huidige toestand waarmee, op basis van live data, een "fast forward knop" richting de toekomst wordt geboden.

Vergelijkend, is het wenselijk om process mining te combineren met "visual analytics". Visual analytics combineert geautomatiseerde analyse met interactieve visualisaties voor een beter begrip van de grote en complexe datasets. Visual analytics maakt gebruik van de wonderbaarlijke menselijke vaardigheid om patronen te zien in ongestructureerde data. Door het combineren van geautomatiseerde process mining technieken met interactieve visuele analyse, is het mogelijk om meer inzichten te halen uit event data.

## C10: Het Verbeteren van Bruikbaarheid voor Niet-Deskundigen

Een van de doelen van process mining is het creëren van "levende procesmodellen", dat wil zeggen, procesmodellen die dagelijks worden gebruikt, in plaats van statische modellen die eindigen in een archief. Nieuwe event data kan gebruikt worden om nieuw gedrag te ontdekken. De verbinding tussen event data en procesmodellen maakt een projectie van de huidige toestand en recente activiteiten op up-to-date modellen mogelijk. Vandaar dat eindgebruikers kunnen interacteren met de resultaten van het process mining van dag-tot-dag. Dergelijke interacties

Dit Manifest is oorspronkelijk verschenen in "BPM 2011 Workshops proceedings, Lecture Notes in Business Information Processing, Springer-Verlag, 2011," en is vertaald in meerdere talen. Zie de welkomspagina van de IEEE Task Force on Process Mining (<http://www.win.tue.nl/ieeftfpm/>) voor meer informatie.

zijn zeer waardevol, maar vereisen ook intuïtieve user interfaces. De uitdaging is, om de geavanceerde process mining algoritmes te verbergen achter gebruiksvriendelijke interfaces, die automatisch parameters instellen en geschikte soorten analyses voorstellen.

## C11: Het Verbeteren van Begrijpelijkheid voor Niet-Deskundigen

Zelfs als het gemakkelijk is om process mining resultaten te genereren, betekent dit niet dat de resultaten daadwerkelijk bruikbaar zijn. De gebruiker kan problemen hebben met het begrijpen van de output of wordt verleid om verkeerde conclusies af te leiden. Om dergelijke problemen te voorkomen, moeten de resultaten worden gepresenteerd met behulp van een geschikte representatie (zie ook GP5). Bovendien moet de betrouwbaarheid van de resultaten altijd duidelijk worden aangegeven. Er kunnen te weinig gegevens beschikbaar zijn om bepaalde conclusies te rechtvaardigen. In feite, waarschuwen bestaande process discovery (ontdekken) technieken meestal niet voor een onder- of overfitting. Ze tonen altijd een model, zelfs als het duidelijk is dat er te weinig gegevens zijn om conclusies te rechtvaardigen.

## Epiloog

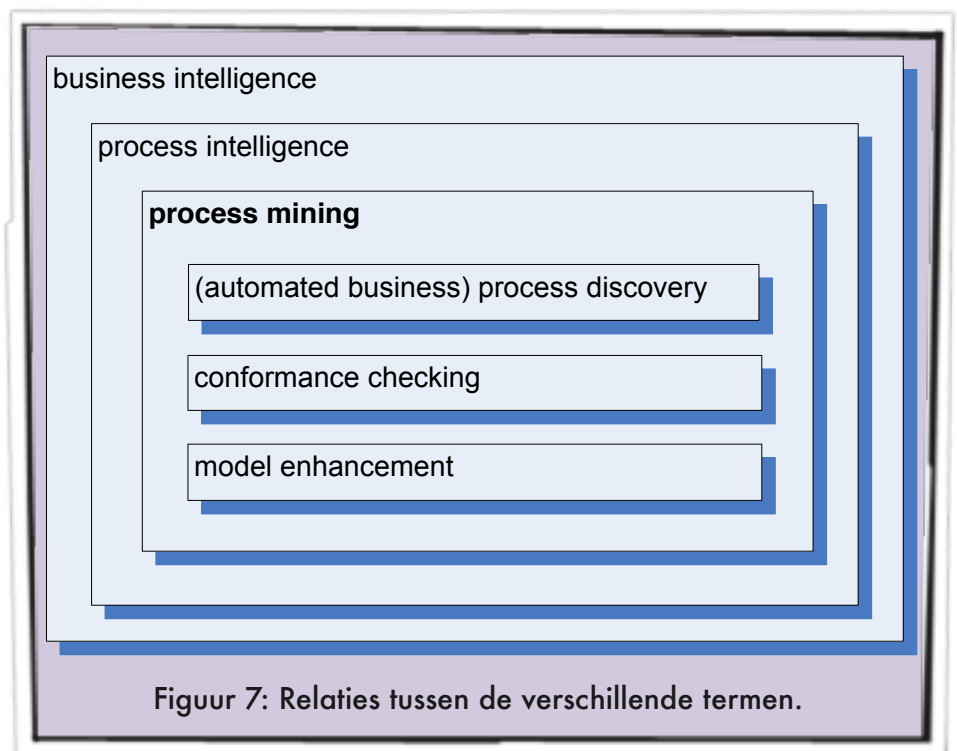
De IEEE Task Force on Process Mining heeft tot doel (a) het bevorderen van de toepassing van Process Mining, (b) het begeleiden van software ontwikkelaars, consultants, business managers en eindgebruikers bij het gebruik van state-of-the-art technieken en (c) het stimuleren van onderzoek op het gebied van process mining. Dit manifest vermeldt de belangrijkste

principes en intenties van de Task Force. Na de introductie van het onderwerp process mining, beschrijft het manifest achtereenvolgens een aantal richtlijnen (Sectie 3) en uitdagingen (Sectie 4). De richtlijnen kunnen worden gebruikt om voor de hand liggende fouten te voorkomen. De lijst van de uitdagingen is bedoeld om inspanningen op gebied van onderzoek en ontwikkeling te sturen. Beide hebben tot doel het volwassenheidsniveau van process mining te verhogen.

Tot slot, nog een paar woorden over terminologie. De volgende termen worden gebruikt in de process mining wereld: workflow mining, (business) process mining, automated (business) process discovery en (business) process intelligence. Verschillende organisaties lijken verschillende termen te gebruiken voor de overlappende concepten. Bijvoorbeeld, Gartner promoot de term "Automated Business Process Discovery" (ABPD) en Software AG gebruikt "Process Intelligence" om te verwijzen naar hun controlling platform. De term "workflow mining" lijkt minder geschikt, aangezien de creatie van workflow modellen slechts een van de vele mogelijke toepassingen is van process mining. Ook de toevoeging van de term "business" verkleint de mogelijkheden op bepaalde toepassingen van process mining. Er zijn tal van toepassingen van process mining (bijvoorbeeld het analyseren van het gebruik van high-

tech systemen of het analyseren van websites) waar deze toevoeging ongepast lijkt. Hoewel process discovery (ontdekken) een belangrijk onderdeel is van het process mining spectrum, is het slechts een van de vele use cases. Conformance checking (conformiteit controleren), voorspelling, organisatorische mining, sociale netwerk analyse, enz. zijn andere use cases, die verder reiken dan process discovery (ontdekken).

Figuur 7 relateert een aantal van de zojuist genoemde termen. Alle technologieën en methoden die gericht zijn op het verstrekken van bruikbare informatie, die gebruikt kan worden als ondersteuning van besluitvorming, kan worden geplaatst onder de paraplu van de Business Intelligence (BI). (Business) process intelligence kan worden gezien als de combinatie van BI en BPM, dat wil zeggen, BI technieken worden gebruikt voor het analyseren en verbeteren van processen en procesmanagement. Process mining kan gezien worden als een concretisering van process intelligence, waarbij event logs als startpunt worden genomen. (Automated business) process discovery is slechts een van de drie basistypen van process mining. Figuur 7 is misschien een beetje misleidend in de zin dat de meeste BI-tools niet voorzien in de process mining functionaliteit zoals beschreven in dit document. De term BI wordt vaak gemakkelijk



Figuur 7: Relaties tussen de verschillende termen.

toegeschreven aan een bepaald instrument of methode die slechts een klein deel van het bredere BI-spectrum bestrijkt.

Er kan om commerciële redenen gekozen worden voor het gebruik van alternatieve termen. Sommige leveranciers kunnen ook nadruk leggen willen op een bepaald aspect (bijvoorbeeld discovery of intelligence). Echter, om verwarring te voorkomen, is het beter om de term "process mining" te gebruiken voor de discipline zoals beschreven in dit manifest.

## Verklarende

## Woordenlijst

**Activiteit:** een goed gedefinieerde stap in het proces. Events kunnen verwijzen naar de start, voltooiing, annulering, etc. van een activiteit voor een specifieke process instance (procesinstantie).

Automated Business Process Discovery: zie Process Discovery.

**Business Intelligence (BI):** brede verzameling van instrumenten en methoden die de gegevens gebruiken ter ondersteuning van besluitvorming. Business Process Intelligence: zie Process Intelligence.

**Business Process Management (BPM):** de discipline die kennis van informatie technologie en kennis van management

wetenschappen combineert en beide toepast voor operationele bedrijfsprocessen.

Case: zie Process Instance.

**Concept Drift:** het verschijnsel dat processen vaak veranderen in de tijd. Het waargenomen proces kan geleidelijk (of plotseling) wijzigen, als gevolg van seizoensgebonden veranderingen of toegenomen concurrentie, wat een analyse moeilijker maakt.

**Conformance Checking:** het analyseren van de vraag, of de werkelijkheid, zoals vastgelegd in een log, voldoet aan de model en vice versa. Het doel is om discrepanties op te sporen en hun ernst te meten. Conformance checking is een van de drie basistypen van het process mining.

**Cross-Organizational Process Mining:** de toepassing van process mining technieken op event logs, die afkomstig zijn van verschillende organisaties.

**Data Mining:** de analyse van (vaak grote) datasets om onverwachte verbanden te vinden en om de gegevens samen te vatten op een manier die nieuwe inzichten biedt.

**Eenvoud:** een maat voor de operationalisering van Occam's Razor, dat wil zeggen, het eenvoudigste model dat het gedrag gezien in de log kan verklaren, is het beste model.

Eenvoud kan op verschillende manieren gekwantificeerd worden, bijvoorbeeld,

het aantal knopen en pijlen in het model.

**Event:** een gebeurtenis opgenomen in een log, bijvoorbeeld, de start, voltooiing, of annulering van een activiteit voor een bepaald process instance (procesinstantie).

**Event Log:** verzameling van gebeurtenissen gebruikt als input voor process mining. Gebeurtenissen hoeven niet te worden opgeslagen in een apart logbestand (bijvoorbeeld gebeurtenissen kunnen worden verspreid over verschillende database tabellen).

**Fitness:** een maat voor het bepalen van hoe goed een bepaald model het gedrag, wat zichtbaar is in het event log, toestaat. Een model heeft een perfecte fitness, als alle traces in het log van begin tot eind, door het model kunnen worden afgespeeld.

**Generalisatie:** een maat voor het bepalen van hoe goed het model in staat is, om ongezien gedrag toe te staan. Een "overfitting" model is niet in staat om voldoende te generaliseren.

**Model Enhancement:** een van de drie basistypen van het process mining. Een procesmodel is uitgebreid of verbeterd met behulp van informatie uit een log. Bijvoorbeeld, knelpunten kunnen worden geïdentificeerd, door het onderzoeken van de tijdstempels tijdens het opnieuw afspelen van een event log op een procesmodel.

## Auteurs

Wil van der Aalst  
Arya Adriansyah  
Ana Karla Alves de  
Medeiros  
Franco Arcieri  
Thomas Baier  
Tobias Blickle  
Jagadeesh Chandra  
Bose  
Peter van den Brand  
Ronald Brandtjen  
Joos Buijs  
Andrea Burattin  
Josep Carmona  
Malu Castellanos  
Jan Claes  
Jonathan Cook  
Nicola Costantini  
Francisco Curbera  
Ernesto Damiani  
Massimiliano de Leoni

Pavlos Delias  
Boudewijn van  
Dongen  
Marlon Dumas  
Schahram Dustdar  
Dirk Fahland  
Diogo R. Ferreira  
Walid Gaaloul  
Frank van Geffen  
Sukriti Goel  
Christian Günther  
Antonella Guzzo  
Paul Harmon  
Arthur ter Hofstede  
John Hoogland  
Jon Espen Ingvaldsen  
Koki Kato  
Rudolf Kuhn  
Akhil Kumar  
Marcello La Rosa  
Fabrizio Maggi

Donato Malerba  
Ronny Mans  
Alberto Manuel  
Martin McCreesh  
Paola Mello  
Jan Mendling  
Marco Montali  
Hamid Motahari  
Nezhad  
Michael zur Muehlen  
Jorge Munoz-Gama  
Luigi Pontieri  
Joel Ribeiro  
Anne Rozinat  
Hugo Seguel Pérez  
Ricardo Seguel Pérez  
Marcos Sepúlveda  
Jim Sinur  
Pnina Soffer  
Minseok Song  
Alessandro Sperduti

Giovanni Stilo  
Casper Stoel  
Keith Swenson  
Maurizio Talamo  
Wei Tan  
Chris Turner  
Jan Vanthienen  
George Varvaressos  
Eric Verbeek  
Marc Verdonk  
Roberto Vigo  
Jianmin Wang  
Barbara Weber  
Matthias Weidlich  
Ton Weijters  
Lijie Wen  
Michael Westergaard  
Moe Wynn



**MXML:** een XML-gebaseerd formaat voor het uitwisselen van event logs. XES vervangt MXML als het nieuwe tool-onafhankelijk process mining formaat.

**Operational Support:** on-line analyse van event data, met als doel om lopende process instances (procesinstanties) te monitoren en te beïnvloeden. Drie ondersteunende operationele activiteiten kunnen worden geïdentificeerd, namelijk: detecteren (het genereren van een melding als het waargenomen gedrag afwijkt van het gemodelleerde gedrag), voorspellen (te voorspellen toekomstig gedrag op basis van gedrag in het verleden, bijvoorbeeld, voorspellen van de resterende verwerkingstijd) en recommend (aanbevelen) (voorstellen van passende maatregelen, om een bepaald doel te realiseren,

bijvoorbeeld om de kosten te minimaliseren).

**Precision:** Maat om te bepalen of het model gedrag verbiedt dat zeer verschillend is van het gedrag, dat te zien is in de log. Een model met weinig Precision (precisie) is "underfitting".

**Process Discovery:** een van de drie basistypes van process mining. Op basis van een event log, wordt een procesmodel geleerd. Bijvoorbeeld, het alpha-algoritme is in staat om een Petri net te ontdekken, door het identificeren van procespatronen in collecties van gebeurtenissen.

**Process Instance:** de entiteit, die wordt behandeld door het proces, welke wordt geanalyseerd. Events verwijzen naar process instances. Voorbeelden van process instances zijn orders van klanten, verzekeringsclaims, aanvragen voor leningen, etc.

**Process Intelligence:** een tak van Business Intelligence gericht op Business Process Management.

**Process Mining:** technieken, tools en methoden om werkelijke processen te ontdekken, bewaken en verbeteren, (dat wil zeggen, niet veronderstelde processen) door extractie van kennis uit event logs, welke over het algemeen beschikbaar zijn in de huidige (informatie) systemen.

**Representational Bias:** de gekozen doeltaal voor de presentatie en constructie van process mining resultaten.

**XES:** is een op XML gebaseerde standaard voor event logs. De standaard is goedgekeurd door de IEEE Task Force on Process Mining als het standaard uitwisselingsformaat voor de event logs (zie [www.xes-standard.org](http://www.xes-standard.org)).