

# Manifest eksploracji procesów

Niniejszy manifest jest publiczną deklaracją przewodnich zasad eksploracji procesów oraz deklaracją intencji członków i sympatyków *IEEE Task Force on Process Mining* (grupy roboczej do spraw eksploracji procesów utworzonej przy IEEE). Celem grupy jest dążenie do prawidłowego zrozumienia i stosowania eksploracji procesów oraz promowanie badań, rozwoju oprogramowania, edukacji, wdrażania rozwiązań i rozwoju dziedziny eksploracji procesów.

Eksploracja procesów jest relatywnie młodą dziedziną badań, z pogranicza kilku nauk: inteligencji obliczeniowej, eksploracji danych oraz modelowania i analizy procesów. Idea eksploracji procesów polega na *odkrywaniu* wiedzy z danych o zdarzeniach, dostępnych we współczesnych aplikacjach biznesowych, infrastrukturze IT oraz systemach bezpieczeństwa, na potrzeby tworzenia modeli procesów, ich *monitorowania* i *doskonalenia*. Eksploracja procesów obejmuje *automatyczne odkrywanie* struktury przepływu pracy (tworzenie nowych modeli procesów na podstawie danych pobranych z dziennika zdarzeń), *badanie zgodności* (monitorowanie odchylenia wykonania procesu od założonego modelu), *analizę*

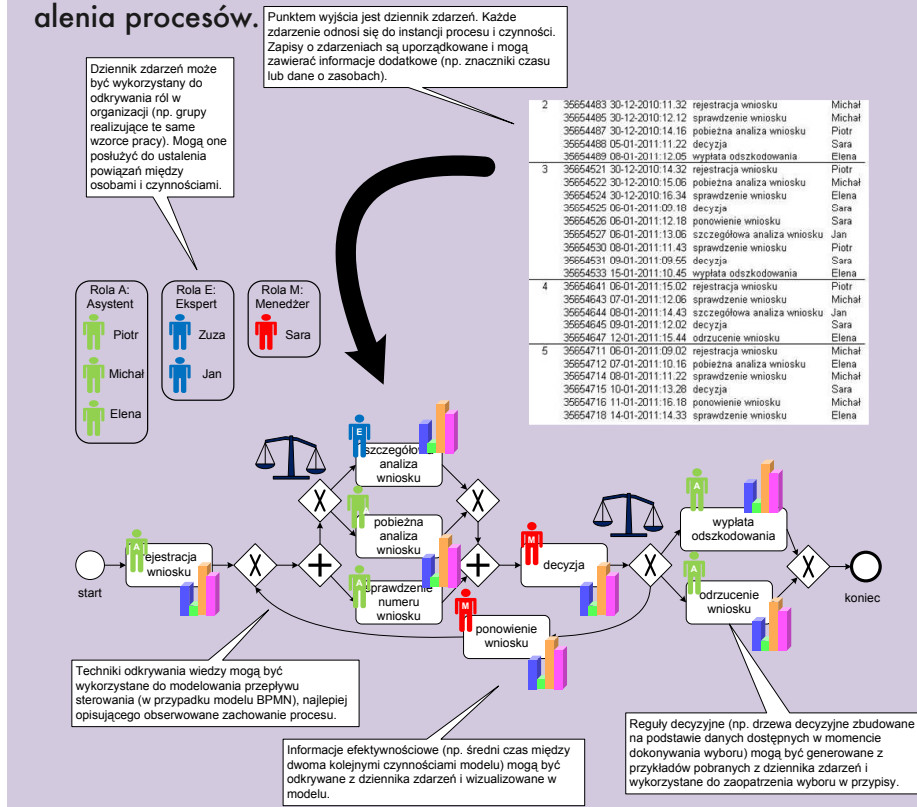
*sieci społecznych w organizacji, automatyczne tworzenie modeli symulacyjnych, rozszerzanie, modyfikowanie i optymalizację modeli już*

## Spis treści:

Aktualny stan wiedzy	3
Przewodnie zasady	7
Wyzwania	11
Epilog	15
Słownik	16

Techniki eksploracji procesów pozwalają na odkrywanie wiedzy z dzienników zdarzeń, które powszechnie występują we współczesnych systemach informatycznych. Techniki te tworzą nowe znaczenie terminów odkrywanie, monitorowanie i doskonalenie procesów. Do głównych przyczyn wzrostu zainteresowania eksploracją procesów należą: gwałtowny wzrost danych o zdarzeniach we współczesnych systemach oraz rozwój narzędzi ukierunkowanych na doskonalenie i wspomaganie procesów biznesowych. Celem grupy roboczej, która opracowała niniejszy manifest, jest promowanie inicjatyw związanych z eksploracją procesów. Poprzez wskazanie zbioru zasad oraz nierozwiązanych problemów, manifest może służyć jako inspiracja dla twórców oprogramowania, naukowców, konsultantów, menedżerów i użytkowników, przyczyniając się do wzrostu stopnia dojrzałości eksploracji procesów jako narzędzia doskonalenia planowania i sterowania procesami operacyjnymi.

## Rysunek 1: Techniki eksploracji procesów odkrywają wiedzę z dzienników zdarzeń w celu budowy, monitorowania i doskonalenia procesów.



istniejących oraz analizę predykcyjną i generowanie rekomendacji w czasie rzeczywistym.

Eksploracja procesów stanowi ważny pomost między eksploracją danych, a modelowaniem i analizą procesów biznesowych oferując nowe spojrzenie na dziedzinę business intelligence. Termin business intelligence odnosi się do systemów analiz i raportowania oraz kokpitów

### Konkretne cele grupy roboczej:

- 1) Budowa świadomości użytkowników, twórców oprogramowania, konsultantów, menadżerów i naukowców w zakresie eksploracji procesów.
- 2) Promocja technik i narzędzi eksploracji procesów oraz stymulowanie nowych zastosowań.
- 3) Odgrywanie istotnej roli w standaryzacji zapisów dzienników zdarzeń.
- 4) Organizacja szkoleń, specjalnych sesji, warsztatów i paneli.

menedżerskich, które obejmują następujące systemy: konfigurowalny pulpit menedżerski (ang. *Business Activity Monitoring - BAM*), dzięki któremu możliwe jest monitorowanie procesów biznesowych w czasie rzeczywistym, system kontroli procesów (ang. *Control Event Processing - CEP*), który umożliwia monitorowanie, sterowanie i optymalizację procesów biznesowych w czasie rzeczywistym, czy system wspólnego zarządzania produkcją (ang. *Corporate Performance Management - CPM*), który służy do pomiaru działania procesu lub organizacji. Obserwowane jest też przenikanie koncepcji eksploracji procesów do koncepcji zarządzania organizacją w zakresie metod ciągłego doskonalenia procesów (ang. *Continuous Process Improvement - CPI*), doskonalenia procesów biznesowych (ang. *Business Process Improvement - BPI*), *Total Quality Management - TQM* i *Six Sigma*.

Obecnie, coraz większe znaczenie przypisuje się do nadzoru korporacyjnego, analizy ryzyka związanej z funkcjonowaniem systemów informatycznych oraz badania zgodności narzędzi informatycznych z wymogami prawa, co jest uregulowane odpowiednimi przepisami, między innymi ustawą Sarbanes-Oxley Act (SOX), czy Basel II

Accord. Narzędzia business intelligence i koncepcje zarządzania, takie jak Six Sigma, czy TQM mające za zadanie doskonalenie działalności operacyjnej poprzez ograniczenie kosztów i czasu koordynacji typowych procesów biznesowych, zapewniają wiarygodność i rzetelność informacji o głównych procesach biznesowych, stając się przydatnym narzędziem wspomagającym wprowadzenie ładu korporacyjnego w organizacji.

Ostatnia dekada charakteryzuje się wzrostem danych o zdarzeniach, rosnącym stopniem dojrzałości technik eksploracji procesów, stałym zainteresowaniem trendami w zakresie doskonalenia procesów (SixSigma, TQM, CPI, CPM) i badaniem zgodności narzędzi informatycznych z wymogami prawa (SOX, BAM, itd.). Sprzyja to rozwojowi rynku eksploracji procesów, czego efektem jest pojawienie się wielu akademickich i komercyjnych systemów eksploracji procesów: ARIS Process Performance Manager (Software AG), Compend (Open Connect), Discover 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), Reflect|one (Pallas Athena).

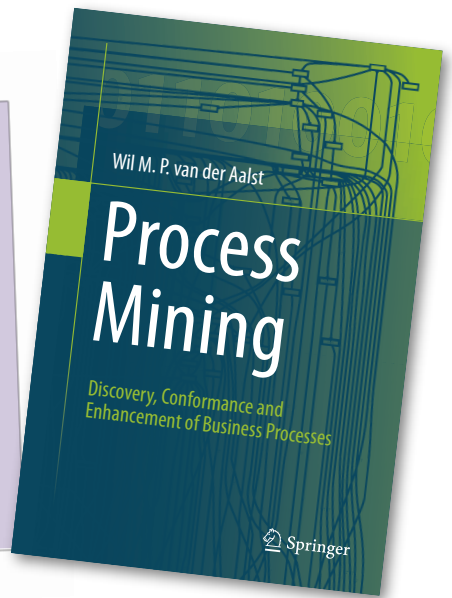
Obserwowany wzrost zainteresowania analizą danych o zdarzeniach wiąże się z utworzeniem w 2009 roku, w ramach Data Mining Technical Committee (DMTC) (komitetu technicznego eksploracji danych przy IEEE) funkcjonującego w IEEE Computational Intelligence Society (CIS) (Stowarzyszenia Inteligencji Obliczeniowej przy IEEE), IEEE Task Force on Process Mining (grupy roboczej do spraw eksploracji procesów przy IEEE). Obecnie, członkami grupy są dostawcy oprogramowania (np. Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), firmy doradcze i konsultingowe oraz użytkownicy (np. ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chile, BWI Systeme GmbH, Excellentia BPM, Rabobank), a także instytuty badawcze (np. 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 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).

Od roku 2009 podjęto szereg działań wynikających z przyjętych celów grupy roboczej, m.in. zorganizowano warsztaty na temat BPI (np. BPI'09, BPI'10 oraz BPI'11) i ścieżki tematyczne w ramach konferencji IEEE (np. CIDM'11) (współ) organizowane przez grupę roboczą. Na rzecz upowszechniania wiedzy o eksploracji procesów zorganizowano seminaria (WCCI'10 i PMPM'09), szkoły letnie (ESSCaSS'09, ACPN'10, CICH'10, itd.), opracowano materiały wideo (por. [www.processmining.org](http://www.processmining.org)) i inne publikacje, w tym, pierwszą książkę na temat eksploracji procesów, wydaną przez wydawnictwo Springer. Grupa robocza (współ)organizowała także pierwszy konkurs BPI Challenge (BPI'11), który poświęcony był odkrywaniu wiedzy z danych o zdarzeniach metodami eksploracji procesów. W roku 2010 grupa robocza zaproponowała rozszerzalny hipertekstowy język XES ([www.xes-standard.org](http://www.xes-standard.org)), bazujący na języku XML i wspomagany przez bibliotekę OpenXES

Książka dotycząca eksploracji procesów  
[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.



([www.openxes.org](http://www.openxes.org)) oraz narzędzia typu ProM, XESame i Nitro, jako standard zapisu danych w dzienniku zdarzeń.

Więcej informacji na ten temat znajduje się na stronie [www.win.tue.nl/ieeefpm/](http://www.win.tue.nl/ieeefpm/).

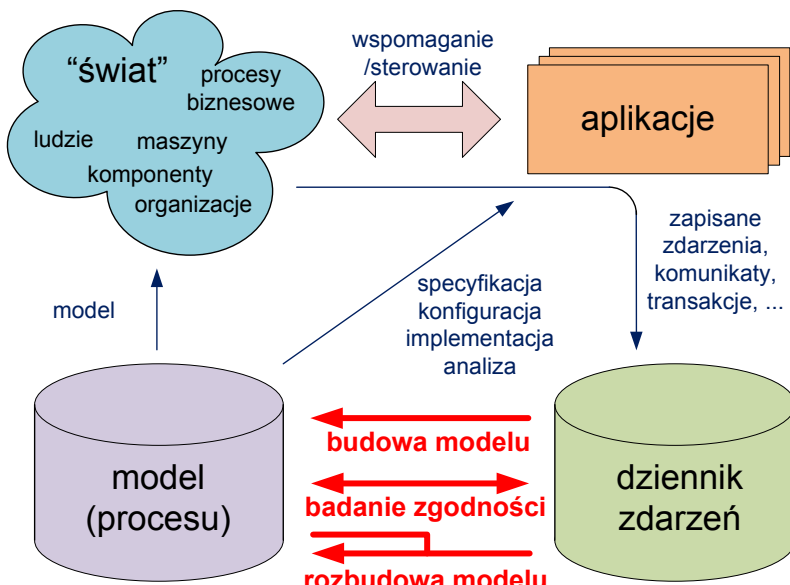
## 2. Aktualny stan wiedzy

Prawo Moore'a dość dokładnie definiuje tempo wzrostu wydajności systemów informatycznych i komputerowych systemów zbierania i przetwarzania danych pomiarowych. Gordon Moore - współzałożyciel Intelu - zaobserwował w roku 1965, że liczba tranzystorów, które mogą pomieścić się w układzie scalonym podwaja się co dwa lata i przewidywał

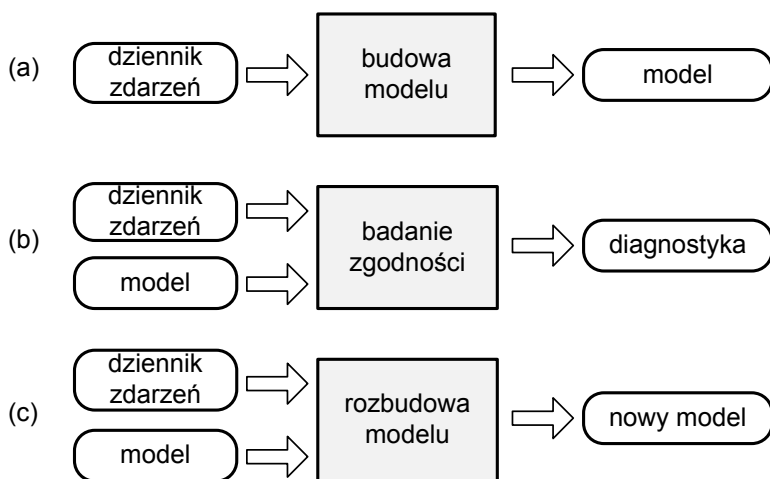
wykładniczy wzrost gęstości upakowania elementów półprzewodnikowych w układach pamięciowych, co zostało w późniejszym okresie rozszerzone na upakowanie tranzystorów w mikroprocesorach oraz ich wydajność. W ciągu ostatnich 50 lat, wzrost był wykładniczy, aczkolwiek tempo zostało spowolnione. Ten postęp spowodował spektakularny wzrost poziomu cyfryzacji w organizacjach (tzn. nasycenia danymi przechowywanymi i/lub wymienianymi drogą elektroniczną), tworząc strukturę, w ramach której poruszają się użytkownicy procesów.

Wzrost stopnia cyfryzacji w organizacjach w odniesieniu do procesów oznacza, że zdarzeniom realizowanym w procesie towarzyszy ślad w postaci zapisu cyfrowego samego zdarzenia oraz danych związanych z jego realizacją. Zdarzenia mogą być różnego rodzaju, od pobrania pieniędzy z bankomatu, czy skonfigurowania maszyny rentgenowskiej, do złożenia wniosku o wydanie prawa jazdy, złożenia zeznania podatkowego, czy potwierdzenia zakupu e-biletu. Wykorzystanie danych o zdarzeniach w celu zapewnienia wglądu w proces, identyfikacji wąskich gardeł, antycypowania przyszłych problemów, rejestrowania naruszenia przepisów prawa lub polityki bezpieczeństwa w systemie informatycznym, rekomendowania działań zaradczych i usprawniających procesy jest sporym wyzwaniem. Eksploracja procesów służy dokładnie tym celom.

Punktem wyjścia do eksploracji procesów jest dziennik zdarzeń. Wszystkie techniki eksploracji procesów zakładają, że możliwy jest sekwencyjny zapis zdarzeń w taki sposób, żeby każde zdarzenie odpowiadało wykonanej czynności (tj. dobrze zdefi-



Rysunek 2: Umieszczenie trzech typów eksploracji procesów: (a) budowy modelu, (b) badania zgodności i (c) rozbudowy modelu.



Rysunek 3: Trzy podstawowe typy eksploracji procesów przedstawione wraz z informacjami wejściowymi i wyjściowymi: (a) budowa modelu, (b) badanie zgodności i (c) rozbudowa modelu.

niowanemu etapowi procesu) i było powiązane z konkretnym przypadkiem procesu (instancją procesu). Dzienniki zdarzeń mogą zawierać dodatkowe informacje o zdarzeniach, tzw. dane kontekstowe, jak zasoby (ludzie lub urządzenia), informacje o wykonywaniu lub inicjowaniu czynności, znacznik czasu lub inne dane zapisywane wraz ze zdarzeniem, np. wielkość zamówienia. Techniki eksploracji procesów wykorzystują te dodatkowe informacje.

Jak pokazano na rys. 2, dzienniki zdarzeń mogą być wykorzystane do realizacji trzech rodzajów eksploracji danych. Pierwszy z nich to budowa modelu procesu. Metody (techniki) budowy modelu przekształcają dane wejściowe pochodzące z dzienników zdarzeń w model procesu bez wykorzystania informacji a priori. Budowa modelu procesu jest najbardziej obiecującą techniką eksploracji procesów. Dane o przykładowym wykonaniu procesów z dziennika zdarzeń są wystarczające do budowy modeli tych procesów, co dla wielu osób może być zaskoczeniem. Drugi rodzaj eksploracji procesów to weryfikacja modelu procesu. W tym przypadku, istniejący model procesu jest porównywany z zapisami z dziennika zdarzeń dotyczącymi rzeczywistego procesu. Weryfikacja modelu może być wykorzystana do sprawdzenia, czy wykonane czynności procesu zapisane w dzienniku zdarzeń są zgodne z modelem i na odwrót. Warto zauważyć, że mogą tu być rozważane różne typy modeli: modele proceduralne, organizacyjne, dekla-

ratywne, reguł biznesowych, itd. Trzeci rodzaj eksploracji procesów to rozbudowa modelu. W tym przypadku, celem jest rozszerzenie i udoskonalenie istniejącego modelu procesu przez wykorzystanie informacji kontekstowych o rzeczywistych procesach zapisanych w dzienniku zdarzeń. Podczas gdy weryfikacja modelu procesu dokonuje oceny zgodności modelu z danymi rzeczywistymi o przebiegu procesu, rozbudowa modelu procesu ma na celu zmianę lub rozszerzenie modelu istniejącego. Np. wykorzystując znaczniki czasu z dziennika zdarzeń można wskazać na wąskie gardła procesu, poziom usług, przepustowość, częstotliwość.

Rysunek 3. opisuje trzy rodzaje eksploracji procesów z uwzględnieniem danych wejściowych i wyjściowych. Techniki wykorzystywane przy budowie modelu procesu pracują na danych wejściowych z dziennika zdarzeń i uzyskują z nich dane wynikowe w postaci modelu procesu. Zbudowany model zazwyczaj ma postać grafu (np. sieć Petriego, BPMN, EPC lub diagram aktywności UML), jakkolwiek może przedstawiać również proces z innej perspektywy (np. jako sieć społeczną). Techniki weryfikacji modelu oraz jego rozbudowy wykorzystują dwa rodzaje danych wejściowych: zapisy z dziennika zdarzeń i model procesu. Dane wynikowe technik weryfikacji modelu zawierają informacje diagnostyczne dotyczące różnic i podobieństw pomiędzy danymi z dziennika zdarzeń i modelem. W przypadku technik roz-

## Charakterystyka Eksploracji Procesów

1. Eksploracja procesów nie jest ograniczona do zdefiniowania przepływu sterowania.

Budowa modelu przepływu sterowania jest uważana za najbardziej ekscytującą część eksploracji procesów. Nie jest jednak do niej ograniczona. Z jednej strony, budowa modelu jest jednym z trzech podstawowych rodzajów eksploracji procesów (budowa, weryfikacja, rozbudowa modelu). Z drugiej strony, duże znaczenie mają również perspektywy organizacyjna oraz czasu.

2. Eksploracja procesów nie jest wyłącznie specyficznym rodzajem eksploracji danych.

Eksploracja procesów może być postrzegana jako brakujące ogniwo między eksploracją danych, a tradycyjnym BPM. Większość technik eksploracji danych nie odwołuje się do procesów. Nie można wogóle porównać modeli procesów współbieżnych z tradycyjnymi technikami eksploracji danych, takimi jak drzewa decyzyjne czy reguły asocjacyjne. Z tego powodu, potrzebne są inne rozwiązania w zakresie reprezentacji zależności i algorytmów.

3. Eksploracja procesów nie jest ograniczona do analiz off-line.

Techniki eksploracji procesów czerpią wiedzę z historycznych danych o zdarzeniach. Jednak, pomimo wykorzystania danych historycznych, analizy mogą być realizowane w czasie rzeczywistym na uruchamionych przypadkach procesów. Np. na podstawie zbudowanego modelu procesu, można ustalić czas zakończenia realizowanego zamówienia klienta.



budowy (naprawy lub rozszerzenia), dane wynikowe zawierają udoskonalony lub rozszerzony model procesu.

Eksploatacja procesów uwzględnia różne konteksty. Kontekst przepływu sterowania koncentruje się na kolejności, w jakiej czynności są wykonywane. Celem eksploatacji procesów, w tym kontekście, jest znalezienie odpowiedniego zapisu wszystkich możliwych ścieżek przebiegu procesu. Dane wynikowe są generowane w postaci sieci Petriego lub w postaci innego grafu (np. EPC, BPMN, diagramu aktywności UML). Kontekst organizacyjny koncentruje się na informacji o zasobach ukrytych w danych o zdarzeniach, tj. kto jest zaangażowany w proces (osoby, systemy, role, działy), jaka jest ich rola i wzajemne powiązania. Celem jest, albo odkrywanie struktury organizacyjnej poprzez grupowanie osób według kryterium roli i jednostki organizacyjnej albo ukazanie sieci społecznej w organizacji. Kontekst przypadku procesu koncentruje się na właściwościach tego przypadku.

Oczywiście, przypadek procesu może być scharakteryzowany przez ścieżkę procesu lub aktorów procesu. Jednakże, przypadki procesu mogą być charakteryzowane także przez wartości danych o zdarzeniach. Np. jeśli przypadek procesu reprezentuje uzupełnienie zapasów, warto wiedzieć, kto jest dostawcą lub jaka jest ilość zamówionego produktu. Kontekst czasu dotyczy koordynacji w czasie i częstotliwości zdarzeń. Gdy zdarzenia mają znaczniki czasu, wówczas możliwe jest odkrywanie wąskich gardeł, mierzenie poziomu usług, monitorowanie wyko-

rzystania zasobów i przewidywanie czasu zakończenia wykonywanego procesu.

Można zaobserwować występowanie kilku typowych błędnych terminów dotyczących eksploatacji procesów. Niektórzy dostawcy oprogramowania, analitycy i naukowcy ograniczają zakres eksploatacji procesów do szczególnych technik eksploatacji, które mogą być wykorzystywane tylko do analiz off-line. Jest to nieporozumienie, dlatego w manifestie opisujemy trzy rodzaje eksploatacji wymienione na poprzedniej stronie.

W celu pokazania miejsca eksploatacji procesów w cyklu życia zarządzania procesem biznesowym wykorzystamy rysunek 4. Widać na nim 7 faz procesu biznesowego wraz z powiązaniem systemami informatycznymi. Faza (prze)projektowania zajmuje się budową lub modyfikacją modelu procesu. W fazie analizy, można wykorzystać elementy analizy dotyczące także modeli alternatywnych. Po zakończeniu etapu (prze)projektowania, model jest implementowany w systemie (faza wdrażania) lub istniejący system jest (re)konfigurowany (faza rekonfiguracji). W fazie uruchomienia systemu, model jest wykonywany, a w systemie realizowane jest śledzenie procesu. Drobne udoskonalenia modelu mogą być wprowadzane do systemu bezpośrednio w fazie regulacji, bez potrzeby powrotu do fazy (prze)projektowania modelu procesu. W fazie diagnozy uruchomiony w systemie proces poddawany jest analizie, a jej wynik może spowo-

## Przewodnie zasady:

PZ1: Dane o zdarzeniach powinny być traktowane jako dane główne

PZ2: Pobieranie danych o zdarzeniach powinno być rezultatem zapytań

PZ3: Model powinien obsługiwać podstawowe konstrukcje przepływu sterowania, m.in. współbieżność i wybór

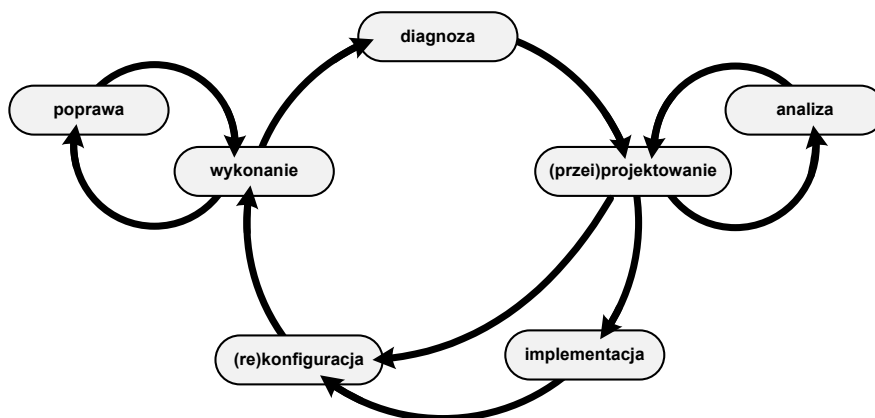
PZ4: Zdarzenia powinny być powiązane z elementami modelu

PZ5: Modele powinny być traktowane jako celowa abstrakcja rzeczywistości

GP6: Eksploatacja procesów powinna być realizowana w sposób ciągły

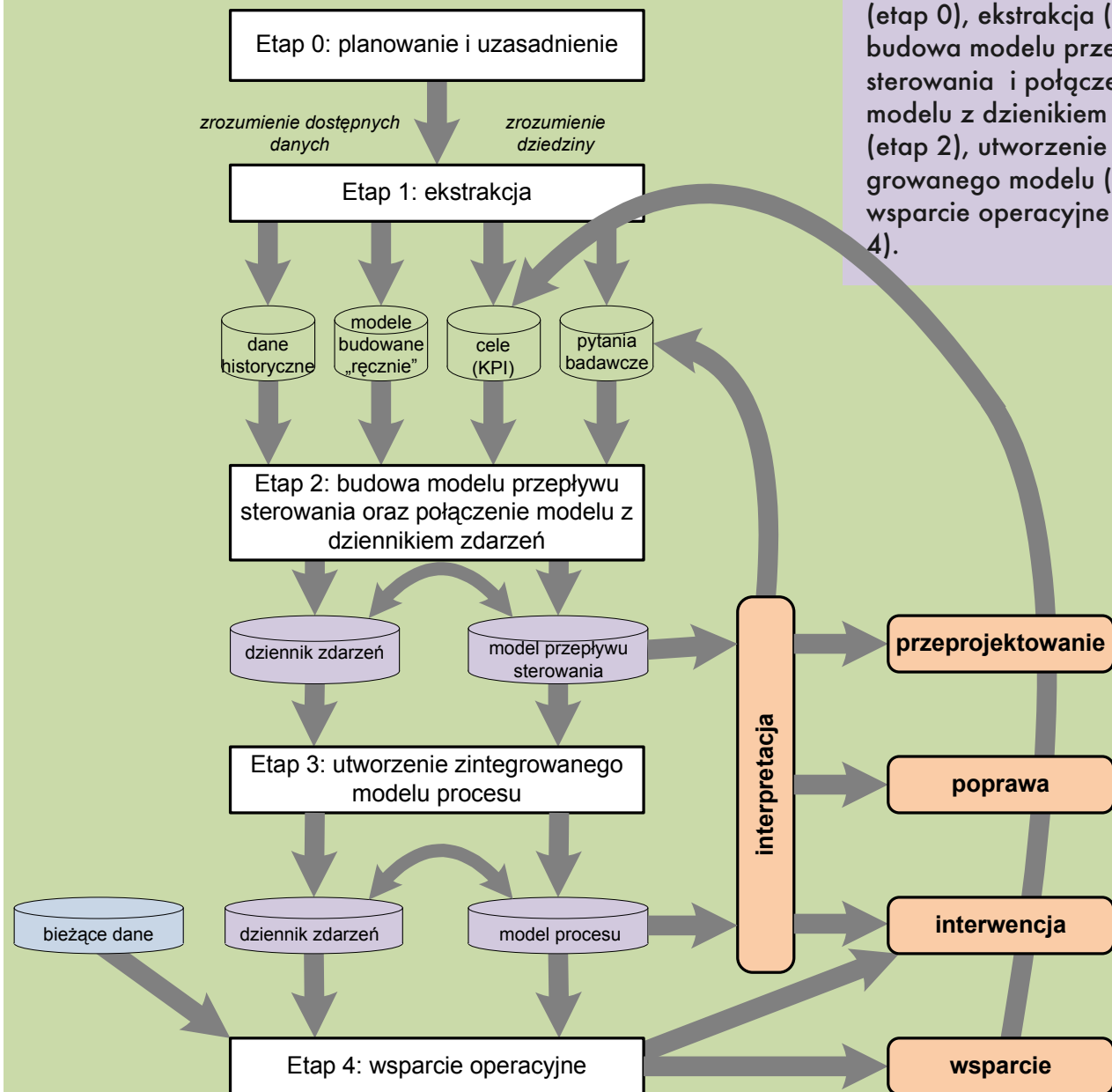
dować powrót do fazy (prze)projektowania modelu. Eksploatacja procesu jest wartościowym narzędziem w większości faz pokazanych na rysunku 4. Wspiera fazę diagnozy w sposób oczywisty, aczkolwiek nie jest do niej ograniczona. Np. w fazie uruchomienia techniki eksploatacji procesów mogą być wykorzystane do wspierania procesów operacyjnych. Analiza predykcyjna i rekomendacje generowane przez modele bazujące na danych historycznych mogą oddziaływać na przebieg procesu. Podobne formy wspierania podejmowania decyzji mogą być wykorzystywane do korekty procesów i do sterowania (re)konfiguracją procesu.

Podczas gdy rysunek 4. przedstawia cały cykl życia BPM, rysunek 5. skupia się na konkretnych czynnościach eksploatacji procesów i artefaktach. Opisuje możliwe etapy projektu eksploatacji procesu. Projekt eksploatacji procesów rozpoczyna się od planowania i uzasadnienia biznesowego (etap 0). Po rozpoczęciu projektu, dane o zdarzeniach, modele procesów, kluczowe wskaźniki efektywności procesów i pytania badawcze muszą być pobrane z systemów, od ekspertów dziedzi-



Rysunek 4: Cykl życia BPM identyfikujący różne fazy procesu biznesowego i odpowiadające mu systemy informatyczne: eksploatacja procesów (potencjalnie) ma znaczenie we wszystkich fazach (z wyjątkiem implementacji).

Rysunek 5: Model cyklu życia eksploracji procesów: planowanie i uzasadnienie (etap 0), ekstrakcja (etap 1), budowa modelu przepływu sterowania i połączenie modelu z dziennikiem zdarzeń (etap 2), utworzenie zintegrowanego modelu (etap 3), wsparcie operacyjne (etap 4).



nowych i kadry zarządzającej (etap 1). Wymaga to zrozumienia danych wejściowych (jakie dane mają być wykorzystane do analizy?), dziedziny (jakie zagadnienia są istotne?) i danych wynikowych w postaci artefaktów zaprezentowanych na rysunku 5. (dane historyczne, modele zbudowane ręcznie, cele i pytania). Na etapie 2 konstruowany jest, w powiązaniu z dziennikiem zdarzeń, model przepływu danych. Mogą tu być zastosowane techniki automatycznej budowy modeli procesów. Zbudowane modele procesów mogą dać już odpowiedź na niektóre pytania badawcze, spowodować przeprojektowanie lub skory-

gowanie czynności. Co więcej, zbudowany model może zostać wykorzystany do przefiltrowania lub wstępnego przetworzenia danych z dziennika zdarzeń (np. usunięcie rzadkich lub nietypowych czynności, dodanie brakujących przypadków). W modelu procesu występują dwa rodzaje zależności: powiązanie zdarzeń należących do tego samego przebiegu procesu oraz powiązanie zdarzeń z jednostką organizacyjną. Na etapie 3, gdy zdarzenia procesu są względnie uporządkowane, model przepływu sterowania może być rozszerzony o inne perspektywy (dane, czas, zasoby). Powiązanie między dziennikiem zda-

zeń i modelem, zdefiniowanym na etapie 2, jest wykorzystywane do rozbudowy modelu (np. znaczniki czasu powiązanych zdarzeń służą do wyeliminowania czasu oczekiwania na kolejną czynność). Może to dać odpowiedź na kolejne pytania badawcze i uruchomić dodatkowe działania. W końcu, model skonstruowany w etapie 3 może być wykorzystywany do wsparcia operacyjnego (etap 4). Wiedza wydobyta z historycznych danych o zdarzeniach jest łączona z informacjami o uruchomionych przypadkach procesu. Może to być użyte do ingerowania w proces, predykcji i rekomendacji. Do etapów 3

Poziom	Charakterystyka	Przykłady
★★★★★	Najwyższy poziom: dziennik zdarzeń jest doskonałej jakości (kompletny i wiarygodny), a zdarzenia dobrze zdefiniowane. Zdarzenia są zapisywane w sposób automatyczny, systematyczny, wiarygodny i bezpieczny. Zapewniona jest ochrona prywatności i bezpieczeństwo zapisów. Co więcej, zapisane zdarzenia (i wszystkie ich atrybuty) mają czytelną semantykę, co powoduje konieczność zdefiniowania jednej lub wielu ontologii. Zdarzenia i ich atrybuty odnoszą się do tych ontologii.	Dzienniki zdarzeń systemów BPM z opisem sematyki zdarzeń.
★★★★	Zdarzenia są zapisywane w sposób automatyczny, systematyczny, wiarygodny, tzn. dane o zdarzeniach są kompletne i wiarygodne. W przeciwieństwie do systemów operujących na poziomie *** pojęcia takie jak instancja procesu (przykład) i czynność są jednoznacznie interpretowane.	Dzienniki zdarzeń tradycyjnego BPM/system przepływu pracy
★★★	Zdarzenia są zapisywane w sposób automatyczny, ale nie jest stosowane podejście systematyczne do zapisywania zdarzeń. Jakkolwiek, w przeciwieństwie do dzienników zdarzeń z poziomu ** istnieje pewien poziom gwarancji, że zapisywane zdarzenia odpowiadają rzeczywistości (dziennik zdarzeń jest wiarygodny, ale nie koniecznie kompletny). Rozważmy np. zdarzenie zapisywane przez system ERP. Chociaż zdarzenia muszą być pobierane z różnych tabel, można założyć, że informacja jest prawdziwa (bezpiecznie jest założyć, że płatność zapisana przez system ERP fizycznie została zrealizowana i vice versa).	Tabele baz danych systemów ERP, dzienniki zdarzeń systemów CRM, logi transakcyjne pochodzące z systemów do wysyłania i odbierania wiadomości, dzienniki zdarzeń systemów high-tech, itd.
★★	Zdarzenia są zapisywane w sposób automatyczny tj. jako dodatkowa funkcjonalność systemów informatycznych. Zakres danych o zdarzeniach nie jest zdefiniowany, tj. nie jest stosowane podejście systematyczne i nie wszystkie zdarzenia są zapisywane. Co więcej, możliwe jest obejście systemu informatycznego (nie zapisanie danych o zdarzeniu). Tym samym, może brakować zdarzeń w dzienniku lub zapis może być błędny.	Dzienniki zdarzeń systemów zarządzania produktem, dane o błędnych logowaniach systemów wbudowanych, inżynierskie arkusze kalkulacyjne, itd.
★	Najniższy poziom: dzienniki zdarzeń są słabej jakości. Zapisane zdarzenia mogą nie odpowiadać rzeczywistości i mogą nie być kompletne. Dotyczy to dzienników zdarzeń, które zawierają zdarzenia zapisywane ręcznie.	Dokumenty papierowe, notatki, itd.

Table 1: Poziom dojrzałości dzienników zdarzeń

i 4 można przejść pod warunkiem, że proces jest wystarczająco stabilny i uporządkowany.

Aktualnie dostępne są narzędzia i techniki, które mogą wspomagać wszystkie etapy przedstawione na rysunku 5. Jakkolwiek, eksploracja procesów jest stosunkowo młodym paradygmatem i większość narzędzi jest ciągle na etapie rozwoju. Co więcej, potencjalni użytkownicy często nie są świadomi możliwości i ograniczeń eksploracji procesów. Dlatego

też, niniejszy manifest porządkuje przewodnie zasady (por. następny punkt) i wyzwania (por. strona 10) stojące przed tą dziedziną, zarówno na potrzeby użytkowników technik eksploracji procesów, jak i naukowców oraz twórców oprogramowania zainteresowanych rozwojem tej dziedziny wiedzy.

### 3. Przewodnie zasady

Każda nowa technologia niesie ze sobą ryzyko błędów. Istnieje niebezpieczeństwo wystąpienia problemów w praktycznym stosowaniu eksploracji procesów. Poniżej definiujemy listę 6 zasad, które pomogą użytkownikom/analitikom ich uniknąć.

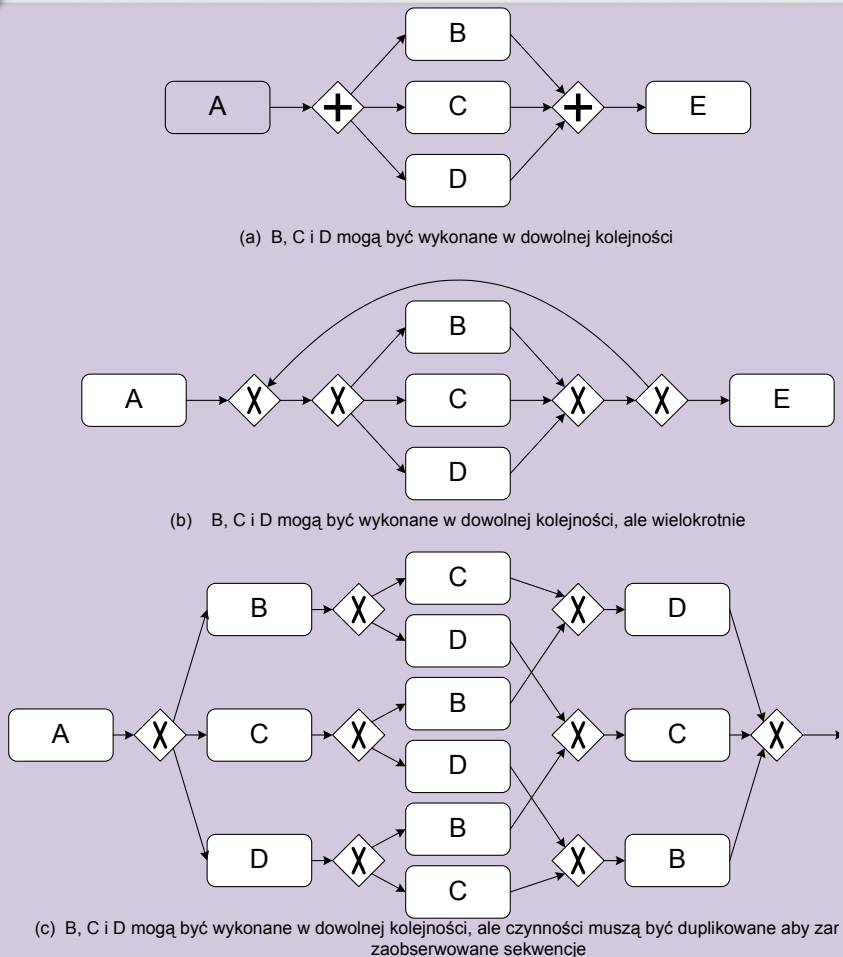


## PZ1. Dane o zdarzeniach powinny być traktowane jako dane główne.

Punktem wyjścia do eksploracji procesów są zapisane zdarzenia. Zauważmy, że dane o zdarzeniach traktowane są jako zapisy *dziennika zdarzeń*, podczas gdy w rzeczywistości nie ma potrzeby, aby były one zapisywane w dedykowanych plikach. Mogą być przechowywane w tabelach baz danych, dziennikach komunikatów, plikach archiwów pocztowych, dziennikach transakcyjnych i in. źródłach danych. Ważniejsza, niż format danych jest ich *jakość*. Jakość danych wyjściowych eksploracji procesów zależy w dużym stopniu od jakości danych wejściowych. Z tego powodu, dzienniki zdarzeń powinny być traktowane w systemach informatycznych wspomagających analizę procesów, jako *dane główne*. Niestety, w systemach informatycznych, dzienniki zdarzeń są często danymi drugorzędowymi, które wykorzystywane są do nadzorowania wykonania programu (debugowania) i profilowania aplikacji. Np. w urządzeniach medycznych firmy Phillips Healthcare, pliki z danymi o zdarzeniach tworzone są tylko do celów "drukowania raportów". Normalizacja nieformalnych praktyk w zakresie zapisywania danych o zdarzeniach jest widoczną potrzebą ze względu na poprawę jakości zapisywanych danych o zdarzeniach. Dane o zdarzeniach powinny być zatem traktowane jako dane główne, a nie drugorzędne.

Do oceny jakości dziennika zdarzeń można stosować kilka kryteriów. Dane o zdarzeniach powinny być *wiarygodne*, tj. bezpiecznie jest założyć, że zdarzenia opisywane przez dane o zdarzeniach zaistniały i ich atrybuty są poprawne. Dzienniki zdarzeń muszą być *kompletne*, tj. w przyjętym okresie czasu nie zostało pominięte żadne zdarzenie. Każde zapisane zdarzenie powinno mieć dobrze zdefiniowaną *semantykę*. Co więcej dane o zdarzeniach muszą być *zabezpieczone*, tj. powinna być zapewniona ochrona i bezpieczeństwo danych. Np. aktorzy procesu powinni wiedzieć o rodzajach zapisywanych zdarzeń i sposobach ich wykorzystania.

Tabela rankingowa (Tabela 1.) określa pięć poziomów dojrzałości dzienników zdarzeń, od najlepszej



Rysunek 6: Przykład pokazuje problemy, gdy nie można zapisać współbieżności (np. bramka AND). Trzy czynności (B, C i D) są współbieżne. Wyobraźmy sobie model, w którym jest 10 czynności równoległych ( $2 \text{ do } 10 = 1024$  możliwości i  $10! = 3.628.800$  możliwości do wykonania sekwencji).

(\*\*\*\*\*) do najniższej jakości (\*). Np. dziennik zdarzeń firmy Phillips Healthcare znajduje się na poziomie (\*\*\*) , tj. zdarzenia są zapisywane automatycznie, a zapisane zachowania odwzorowują wiernie rzeczywistość, niemniej brak jest systematycznego podejścia przy definiowaniu semantyki zdarzeń. Techniki eksploracji procesów mogą być stosowane do dzienników zdarzeń z poziomów (\*\*\*\*\*), (\*\*\*) i (\*\*). Zasadniczo, możliwe jest również zastosowanie eksploracji do poziomów (\*\*) i (\*). Jakkolwiek analiza takich zapisów jest problematyczna, a rezultaty niewiarygodne. W zasadzie, nie ma sensu stosować eksploracji do poziomu (\*).

Jeśli organizacje zamierzają czerpać korzyści z eksploracji procesów, powinny zadbać o to, aby ich dzienniki zdarzeń były jak najwyższej jakości.

## PZ2. Pobieranie danych o zdarzeniach powinno wiązać się z wykonywaniem zapytań.

Jak przedstawiono na rysunku 5., czynności charakterystyczne dla eksploracji procesów, powinny być związane z wykonywaniem zapytań. Bez zdefiniowanych zapytań, trudno jest pobrać istotne dane o zdarzeniach. Jeśli mamy na przykład do czynienia z tysiącami tabel w bazie danych systemu ERP (np. SAP), wówczas bez zdefiniowania konkretnych zapytań niemożliwy jest wybór odpowiednich tabel zawierających właściwe dane.

Model procesu biznesowego pokazany na rysunku 1., opisuje cykl życia procesu dla przypadków (tj. instancji) procesu określonego typu. Stąd, zanim zastosujemy techniki

eksploracji procesów, musimy wybrać przypadki tego typu do analizy. Wybór powinien być rezultatem zapytań, które mogą okazać się nietrywialne. Rozważmy np. zarządzanie zamówieniami. Każde zamówienie może składać się z wielu pozycji, ponieważ klient może zamówić wiele produktów na jednym zamówieniu. Jedno zamówienie może zatem wiązać się wieloma dostawami. Jedna dostawa może też odpowiadać pozycjom z wielu zamówień. Stąd, istnieje relacja wiele-do-wielu między zamówieniami i dostawami oraz jeden-do-wielu między zamówieniami i pozycjami. Rozważając bazę danych o zamówieniach, pozycjach i dostawach, możemy utworzyć różne modele procesów. Jeden z nich może wykorzystywać dane do opisu cyklu życia zamówienia, inny - pojedynczej pozycji lub konkretnego dostawcy.

### PZ3. Model powinien obsługiwać podstawowe konstrukcje przepływu sterowania, m.in. współbieżność i wybór

Modelowanie procesów biznesowych cechuje duża różnorodność językowa (np. BPMN, EPC, sieci Petriego, BPEL, diagramy aktywności UML). Niektóre notacje mogą mieć wiele elementów, np. BPMN obejmuje ponad 50 elementów graficznych, a inne notacje zawierają tylko elementy podstawowe (np. sieci Petriego budowane są tylko z trzech różnych elementów: miejsc, tranzycji i łuków). Model procesu utożsamiany jest często z diagramem przepływu sterowania. Podstawowe elementy konstrukcyjne przepływu pracy (inaczej: wzorce przepływu), które występują we wszystkich głównych językach, obejmują następujące konstrukcje: sekwencja, działania równoległe (bramka AND), wybór (bramka XOR) i pętla. Techniki eksploracji procesów powinny oczywiście radzić sobie z tymi wzorcami. Jakkolwiek, niektóre z nich mają trudności z obsługą współbieżności i wspierają tylko łańcuchy Markowa/systemy tranzycji.

Rysunek 6. przedstawia, jaki jest rezultat użycia technik eksploracji procesów niezdolnych do obsługi współbieżności (brak bramek AND) do modelowania procesów

współbieżnych. Rozważmy następujący dziennik zdarzeń  $L = \{ \langle A, B, C, D, E \rangle, \langle A, B, D, C, E \rangle, \langle A, C, B, D, E \rangle, \langle A, C, D, B, E \rangle, \langle A, D, B, C, E \rangle, \langle A, D, C, B, E \rangle \}$ .  $L$  zawiera przypadki, które rozpoczynają się od  $A$  i kończą się na  $E$ . Czynności  $B, C, D$  pojawiają się między nimi w dowolnym porządku.

Rysunek 6a. przedstawia prostą reprezentację procesu, która wykorzystuje dwie bramki AND.

Przypuśćmy, że technika eksploracji procesów nie obsługuje bramek AND. W takiej sytuacji, można zaproponować następujące rozwiązania (6b., 6c.). Model BPMN z rysunku 6b., jest prosty, ale dopuszcza zbyt dużą wielość zachowań (przypadki, takie jak  $\langle A, B, B, B, E \rangle$  są zgodne z modelem, ale nie z dziennikiem zdarzeń).

Rysunek 6c. przedstawia model złożony, który uwzględni wszystkie przypadki z  $L$ , ale koduje każdą sekwencję w sposób bezpośredni. Przykład ten ilustruje niedopasowanie modelu do rzeczywistych procesów, które zawierają dziesiątki potencjalnie współbieżnych czynności (tj. modele pozwalają na zbyt wielką różnorodność zachowań) i/lub nadmierną złożoność, gdy modele nie obsługują współbieżności.

Z rysunku 6. wynika, że obsługa co najmniej podstawowych wzorców przepływu, jest bardzo ważna. Oprócz wspomnianych wzorców podstawowych, pożądana jest też wsparcie bramek OR, ponieważ umożliwiają one prostą reprezentację niewykluczających się decyzji i częściowej synchronizacji.

### PZ4. Powinny zostać ustanowione relacje pomiędzy zdarzeniami oraz innymi elementami modelu

Jak wskazano w punkcie 2. błędem jest założenie, że eksploracja procesów jest ograniczona do budowy modelu przepływu sterowania. Stosownie do rysunku 1., model procesu może obejmować wiele perspektyw (organizacyjną, czasową, danych). Co więcej, budowa modelu jest tylko jednym z trzech rodzajów eksploracji procesów, jak pokazano na rysunku 3. Dwa pozostałe rodzaje eksploracji (badanie zgodności i roz-

budowa) w dużym stopniu zależą od relacji między elementami modelu i zdarzeniami z dziennika zdarzeń. Ta zależność może być wykorzystywana do odtworzenia zdarzeń historycznych na modelu. Powtórne wywołanie zdarzeń może być wykorzystane do ujawnienia rozbieżności między dziennikiem zdarzeń i modelem. Np. niektóre zdarzenia z dziennika zdarzeń nie są możliwe do wykonania z punktu widzenia modelu. Techniki do sprawdzania zgodności kwantyfikują i diagnozują takie rozbieżności. Znaczniki czasu w dzienniku zdarzeń mogą być wykorzystane do analizy zachowania w trakcie odtwarzania zdarzenia. Rozbieżności czasowe między przypadkowo powiązаныmi czynnościami mogą być wykorzystywane, aby dodać do modelu spodziewany czas oczekiwania.

Przykłady te pokazują, że relacje między zapisami z dziennika zdarzeń i elementami w modelu służą jako punkt startu do różnych analiz. W niektórych przypadkach takie relacje mogą być skomplikowane. Np. zdarzenie może odwoływać się do dwóch różnych czynności lub nie jest jasne, do której czynności się odwołuje. Taka niejednoznaczność musi być zlikwidowana, aby można było właściwie zinterpretować rezultaty eksploracji. Oprócz problemu ustalenia zależności między zdarzeniami i czynnościami, występuje jeszcze problem relacji między zdarzeniami i instancjami procesu. Potocznie nazywa się to korelacją zdarzeń.

### PZ5. Modele powinny być traktowane jako celowa abstrakcja rzeczywistości.

Model opracowany jest na podstawie danych z dziennika zdarzeń, który reprezentuje określony obraz rzeczywistości. Taki obraz pozwala skonstruować model abstrakcyjny zachowań zapisanych w dzienniku zdarzeń. Obraz rzeczywistości z określonego dziennika zdarzeń pozwala na różnorodne spojrzenie na dane. Co więcej, różni użytkownicy mogą wymagać dostępu do różnych perspektyw danych. W zasadzie, model zbudowany na podstawie danych z dziennika zdarzeń można porównać do mapy geograficznej. W dalszej części niniejszej przewodniej zasady

rozwiemy to porównanie w dwóch wymiarach.

Po pierwsze, ważne jest, by uświadomić sobie, że dla określonego obszaru geograficznego nie można zbudować jednej uniwersalnej mapy. W zależności od rodzaju zastosowania, wyróżnia się mapy drogowe, turystyczne, ścieżek rowerowych, itp. Wszystkie one pokazują obraz tej samej rzeczywistości, ale z różnych perspektyw. To samo dotyczy modeli procesów. Model powinien wypuklać elementy odpowiednie dla poszczególnych kategorii użytkowników. Zbudowane modele mogą skupiać się na różnych perspektywach danych (przepływ sterowania, przepływ danych, czas, zasoby, koszty) i powinny zawierać informacje na różnych poziomach szczegółowości, np. menedżer zainteresowany jest ogólnym modelem procesu z uwzględnieniem informacji o kosztach, a analityk procesów - szczegółowym modelem procesu na potrzeby analizy odchyień od typowego przebiegu procesu. Zauważmy, że różne kategorie użytkowników zainteresowane są różnym poziomem szczegółowości modelu: poziom strategiczny (decyzje na tym poziomie mają efekt długoterminowy i są podejmowane w oparciu o zagregowane dane o długim horyzoncie czasowym), poziom taktyczny (decyzje podejmowane na tym poziomie mają efekt średnioterminowy i w większości dotyczą danych z ostatniego okresu) i poziom operacyjny (decyzje z tego poziomu mają skutki natychmiastowe i podejmowane są w oparciu o dane pochodzące z uruchomianego przypadku).

Po drugie, warto skorzystać ze znanych powszechnie zasad kartografii, kiedy przychodzi do opracowania czytelnego odwzorowania rzeczywistości. Przykładowo, mapy drogowe mogą nie zawierać dróg drugorzędnych i małych miejscowości. Są one pomijane lub przedstawiane w postaci zagregowanych kształtów (np. ulice lub przedmieścia zlewają się w miasto). Kartografowie nie tylko eliminują nieistotne detale, ale również używają kolorów do wyróżnienia ważnych elementów krajobrazu. Ponadto, elementy graficzne mają określony rozmiar podkreślający ich znaczenie (np. może występować różna grubość linii i rozmiar kropek). Mapy geograficzne mają również jasną interpretację osi X i Y, tj. układ mapy nie może być przed-

stawiony w dowolny sposób, ponieważ współrzędne punktów na mapie są jednoznacznie określone. Wszystko to jest pozornie w zupełnym kontraście z zasadą budowy modeli, gdzie nie używa się kolorów, rozmiarów i współrzędnych. Jednakże, idee zaczerpnięte z kartografii mogą być łatwo wdrożone do odwzorowania procesu.

Przykładowo, rozmiar symbolu czynności może być zastosowany do wypuklenia jej częstotliwości lub innej własności (tj. kosztu czy użytego zasobu). Szerokość łuku może oznaczać siłę zależności przyczynowej, a jego kolor - "wąskie gardło".

Powyższe obserwacje wskazują, jak ważny jest dobór właściwej reprezentacji modelu. Istotne jest to również z punktu widzenia prezentacji wyników dla użytkowników końcowych oraz doboru algorytmów dla odpowiednich modeli (W5).

## PZ6. Eksploracja procesów powinna być realizowana w sposób ciągły.

Eksploracja procesów może dostarczać istotnych "map" bezpośrednio powiązanych z danymi o zdarzeniach. Analiza predykcyjna może wykorzystywać zarówno dane historyczne, jak również może być realizowana w czasie rzeczywistym korzystając z danych bieżących. Co więcej, zmiany w procesach zachodzą także w trakcie ich analizy. Biorąc więc pod uwagę ich dynamiczną naturę, nierozsądnie byłoby postrzegać analizę procesów jako czynność jednorazową. Celem modelowania nie jest utworzenie modelu statycznego, ale tchnienie życia w model, aby zachęcić użytkowników i analityków do codziennej obserwacji zmian w procesie poprzez model.

Można to porównać do wykorzystania hybrydowych aplikacji sieciowych (mashap) wykorzystujących tagowanie geograficzne. Tysiące aplikacji sieciowych wykorzystuje Google Maps (np. aplikacje dostarczające informacje o warunkach drogowych, nieruchomościach, fast food'ach lub repertuarze kin na wybranym terenie). Można w nich płynnie powiększać lub zmniejszać wyświetlany obszar i wchodzić w interakcję z aplikacją (np. na mapie wyświetlane są wszystkie utrudnienia drogowe, a użytkownik może wybrać te szczegóły, którym chce się przyrzeć). Na podobnej zasadzie powinna być możliwa eksploracja procesów w

## Wyzwania:

W1: Wyszukiwanie, łączenie, czyszczenie danych o zdarzeniach

W2: Wykorzystanie złożonych dzienników zdarzeń o różnych charakterystykach

W3: Tworzenie reprezentatywnych testów wzorcowych (benchmarków)

W4: Zdolność do radzenia sobie z nieprzewidywalnymi zmianami

W5: Postęp w zakresie przezwyciężenia obciążenia reprezentacji w budowie modeli procesów

W6: Balansowanie między kryteriami jakościowymi takimi, jak dopasowanie, prostota, precyzja i generalizacja

W7: Eksploracja międzyorganizacyjna

W8: Wspomaganie operacyjne

W9: Połączenie eksploracji procesów z innymi rodzajami analiz

W10: Wzrost użyteczności dla użytkownika

W11: Poprawa czytelności dla użytkownika

czasie rzeczywistym wykorzystująca bieżące dane o zdarzeniach. Wykorzystując metaforę mapy, możemy mówić o zdarzeniach jako współrzędnych GPS, które mogą być wyświetlane na mapach w czasie rzeczywistym. Analogicznie do systemów nawigacji samochodowej, narzędzia eksploracji procesów mogą pomóc użytkownikom końcowym (a.) przejść przez proces (b.) śledzić dynamiczne informacje na mapach procesowych (np. śledzenie "utrudnień drogowych" w procesach biznesowych), (c.) w analizie predykcyjnej dotyczącej uruchomionych przypadków (np. czasu zakończenia opóźnionego procesu). Przykłady te pokazują, że szkoda



byłoby zaprzepaścić takie możliwości aktywnego wykorzystania modeli procesów. Z tego powodu eksploracja procesów powinna być postrzegana jako proces ciągły udostępniający informacje umożliwiające podjęcie odpowiedniej decyzji w czasie rzeczywistym w różnych przedziałach czasowych (minutach, godzinach, tygodniach, miesiącach).

## 4. Wyzwania

Eksploracja procesów jest ważnym narzędziem w nowoczesnych organizacjach, które muszą zarządzać skomplikowanymi procesami operacyjnymi. Z jednej strony, mamy do czynienia z nieprawdopodobnym wzrostem ilości danych o zdarzeniach, z drugiej, dane o zdarzeniach powinny dokładnie odpowiadać rzeczywistości procesów, aby sprostać wymaganiom związanym z weryfikacją zgodności, efektywnością i obsługą klientów. Pomimo stosowania w praktyce eksploracji procesów, ciągle stawiane są przed nią nowe wyzwania. Pokazuje to, że jest to ciągle rozwijająca się dziedzina wiedzy. W dalszej części wymienimy niektóre z tych wyzwań. Naszą intencją nie jest zamieszczenie pełnej ich listy. Z biegiem czasu mogą pojawić się nowe wyzwania, a dotychczasowe zostaną rozwiązane wraz z rozwojem tej dziedziny.

### W1. Wyszukiwanie, łączenie, czyszczenie danych o zdarzeniach.

Wyszukiwanie danych o zdarzeniach to wciąż duży problem. Zazwyczaj, należy pokonać kilka następujących przeciwności:

- Dane mogą być przechowywane w wielu fizycznych lokalizacjach. Należy wówczas je połączyć. Problem występuje, gdy dane te mają różne identyfikatory. Np. jeden system wykorzystuje do zidentyfikowania osoby nazwisko i datę urodzenia, a drugi numer PESEL.
- Dane o zdarzeniach często koncentrują się na obiektach, a nie na procesach. Np. zdarzenia mogą odnosić się do pojedynczych produktów, palet lub kontenerów poprzez kody RFID. Aby prześledzić konkretne

zamówienie Klienta, zdarzenia skoncentrowane na obiektach muszą zostać zagregowane i przetworzone.

- Dane o zdarzeniach mogą być niekompletne. Powszechnym problemem jest to, że zdarzenia wyraźnie odnoszą się do konkretnych instancji procesów. Możliwe jest wprawdzie uzyskanie pełnej informacji o procesie, ale wymaga to dużego nakładu pracy. Poza tym, informacje o czasie realizacji niektórych zdarzeń są niekiedy niedostępne, aczkolwiek można oczywiście interpolować znaczniki czasu.
- Dziennik zdarzeń może zawierać dane nietypowe, tzw. szum. Jak zdefiniować nietypowe zachowania? Jak takie zachowania wykryć? Aby oczyścić dane o zdarzeniach, trzeba umieć odpowiedzieć na te pytania.
- Dzienniki zdarzeń mogą zawierać informacje o różnym poziomie szczegółowości (różnej granulacji). W dzienniku zdarzeń systemu informatycznego szpitala zdarzenie może odpowiadać prostym testom krwi lub skomplikowanym procedurom chirurgicznym. Także znaczniki czasu mogą mieć różny poziom szczegółowości, począwszy od milisekund (28-09-2011:h11m 28s32ms342), a kończąc na zwykłej dacie (28-09-2011).
- Zdarzenia mogą pojawiać się w określonych kontekstach (pogoda, nakład pracy, dzień tygodnia, itp.). Kontekst ten może wyjaśniać określone zjawiska, np. czas odpowiedzi jest dłuższy, niż zwykle z powodu trwających nadal prac lub wakacji. Wprowadzenie kontekstu do analizy jest istotne. Oznacza to połączenie danych o zdarzeniach z danymi kontekstowymi. Wiąże się to z tzw. przekleństwem wymiarowości, gdy analizy napotykają na trudności obliczeniowe po dodaniu zbyt dużej liczby zmiennych.

Do poradzenia sobie z tymi trudnościami potrzebne są lepsze narzędzia i metodologie. Co więcej, jak wskazano wcześniej, organizacje muszą traktować dzienniki zdarzeń jak dane główne, a nie drugorzędne. Celem jest otrzymanie dziennika zdarzeń o dojrzałości z poziomu (\*\*\*\*\*) (patrz Tabela 1.). Tutaj przydadzą się informacje na temat zapewnienia jakości danych w hurtowniach danych.

Np. proste procedury weryfikacji danych wejściowych mogą przyczynić się do znaczącej redukcji błędów w dziennikach zdarzeń.

### W2. Wykorzystanie złożonych dzienników zdarzeń o różnych cechach

Dzienniki zdarzeń mogą mieć różne cechy. Niektóre mogą być ogromnych rozmiarów sprawiając kłopoty z ich obsługą, podczas gdy inne mają niewielkie rozmiary i nie zawierają wystarczającej liczby danych do przeprowadzenia wiarygodnego wniosku.

W niektórych dziedzinach, zapisywane są niewyobrażalne ilości danych o zdarzeniach. Dlatego też, wymaga się dodatkowego wysiłku aby osiągnąć właściwą wydajność i skalowalność danych. Np. ASML monitoruje w sposób ciągły wszystkie swoje sprzedane skanery litograficzne wykorzystywane w różnych organizacjach (np. Samsung, Texas Instruments) do produkcji chipów (ok. 70% chipów produkowanych na całym świecie, korzysta ze skanerów ASML). Istniejące narzędzia mają trudności z obsługą danych o takim rozmiarze (petabajty danych). Oprócz ilości zapisanych zdarzeń, występują inne cechy, jak średnia liczba zdarzeń/przypadek, podobieństwa między przypadkami, liczba unikalnych zdarzeń, liczba unikalnych ścieżek. Rozważmy dziennik zdarzeń L1 z następującymi cechami: 1000 przypadków, średnio 10 zdarzeń/przypadek, mała wariancja (np. kilka przypadków realizuje tę samą ścieżkę modelu). Dziennik zdarzeń L2 zawiera 100 przypadków, ale średnio ma 100 zdarzeń/przypadek i wszystkie przypadki realizują odmienne ścieżki w modelu. Oczywiście L2 jest przypadkiem dużo trudniejszym do analizy, niż L1, chociaż oba dzienniki mają podobny rozmiar (szacunkowo 10 000 zdarzeń).

Ponieważ dziennik zdarzeń zawiera tylko dane próbkowane o zdarzeniach, nie powinien być traktowany jako kompletny. Techniki eksploracji procesów powinny radzić sobie z danymi niekompletnymi i niepewnymi

poprzez wykorzystanie teorii Open World Assumption (OWA): fakt, że się coś nie wydarzyło nie oznacza, że nie może się zdarzyć. Nierozwiązanym problemem jest obsługa małych dzienników zdarzeń charakteryzujących się dużą zmiennością.

Jak wspomniano, niektóre dzienniki zdarzeń zawierają dane o zdarzeniach o bardzo niskim stopniu abstrakcji. Dzienniki te mają tendencję do przybierania dużych rozmiarów, a pojedyncze zdarzenia nie są istotne dla interesariuszy. Należy wówczas przejść na wyższy poziom agregacji danych o zdarzeniach. Np. podczas analizy procesu diagnozy i leczenia określonej grupy pacjentów, indywidualne wyniki testów z laboratorium szpitalnego nie są brane pod uwagę.

W tym momencie, należy zastosować metodę prób i błędów do oceny, czy dziennik zdarzeń ma właściwą postać. Odpowiednie narzędzia powinny pozwolić na przeprowadzenie szybkiego testu wykonalności w zakresie eksploracji dziennika zdarzeń. Taki test powinien wskazać na potencjalne problemy wydajności i ostrzegać, gdy dziennik zdarzeń jest zbyt skomplikowany lub szczegółowy.

### W3. Tworzenie reprezentatywnych testów wzorcowych (benchmarków)

Eksploracja procesów jest technologią przyszłościową. To wyjaśnia, dlaczego nie zdefiniowano jeszcze dobrych testów wzorcowych. Np. dostępnych jest wiele technik eksploracji procesów, a produkty różnych dostawców oferują różne techniki. Nie ma jednak konsensusu co do jakości tych technik. Chociaż znacząco różnią się funkcjonalnością i efektywnością, trudno jest porównać. Z tego powodu, należy zbudować dobre testy wzorcowe złożone z przykładowych zbiorów danych i reprezentatywnych kryteriów jakościowych.

Wiele zestawów dobrych testów wzorcowych jest dostępnych dla klasycznych technik eksploracji danych. Siłą napędową rozwoju testów wzorcowych są działania dostawców oprogramowania i badaczy mające na celu polepszenie efektywności ich technik. W przypadku eksploracji procesów to zadanie jest większym wyzwaniem. Np. model relacyjny zaproponowany przez Codd'a w roku 1969 jest prosty i szeroko wykorzystywany. W rezulta-

cie, na konwersję danych z jednej do drugiej bazy nie potrzeba dużego nakładu pracy i nie ma też problemów z ich interpretacją. W przypadku procesów, nie istnieje taki prosty model. Standardy w zakresie modelowania procesów są o wiele bardziej złożone i nie wszyscy dostawcy wykorzystują dokładnie te same definicje. Natura procesów jest dużo bardziej złożona, niż danych w postaci tabelarycznej.

Z tego powodu, ważne jest opracowanie reprezentatywnych testów wzorcowych dla eksploracji procesów. Np. istnieje wiele metryk wykorzystywanych do pomiaru jakości wyników eksploracji procesów (dopasowanie, prostota, precyzja, generalizacja). Co więcej, niektóre dzienniki zdarzeń są dostępne publicznie (por. [www.processmining.org](http://www.processmining.org)). Przykładem może być dziennik zdarzeń wykorzystany w trakcie pierwszego Business Process Intelligence Challenge (BPIC'11) zorganizowanego przez grupę roboczą (por. identyfikator cyfrowy: doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54).

Z jednej strony, powinny istnieć testy wzorcowe wykorzystujące rzeczywiste dane o zdarzeniach. Z drugiej strony, istnieje potrzeba stworzenia syntetycznych zbiorów danych przedstawiających określone charakterystyki. Takie syntetyczne zbiory danych umożliwiają rozwój technik eksploracji procesów, które muszą np. dostosowywać się do niekompletnych i zaszumionych danych o zdarzeniach lub specyficznych zbiorów procesów.

Oprócz budowy reprezentatywnych testów wzorcowych, istotne jest wypracowanie jednolitych kryteriów oceny wyników eksploracji procesów (por. W6). Ponadto, do oceny mogą zostać zaadaptowane techniki walidacji krzyżowej zaczerpnięte z eksploracji danych. Rozważmy przykładowo k-krotną walidację krzyżową. Można wówczas podzielić dziennik zdarzeń na k części. k-1 części wykorzystamy do uczenia modelu procesu, a techniki zgodności modelu z procesem można wykorzystać do oceny pozostałej jednej części dziennika. Możemy to powtarzać k razy, uzyskując ocenę jakości modelu.

### W4. Zdolność do radzenia sobie z nieprzewidywalnymi zmianami

Termin nieprzewidywalne zmiany odnosi się do sytuacji, w której proces zmienia się w trakcie jego analizy. Np. na początku dziennika zdarzeń dwie czynności mogą być wykonywane równolegle, natomiast później te same czynności w dzienniku zdarzeń wykonywane są sekwencyjnie. Procesy mogą się zmieniać ze względu na okresowe/sezonowe zmiany (np. w grudniu jest większy popyt lub w piątek wieczorem dostępna jest mniejsza liczba pracowników) lub ze względu na zmieniające się warunki (np. rynek staje się bardziej konkurencyjny). Tego typu zmiany wpływają na procesy i ważne jest, aby je zidentyfikować i przeanalizować. Nieoczekiwane zmiany w procesach mogą być odkrywane poprzez podział dziennika zdarzeń na mniejsze części i analizę występujących w nich prawidłowości. Taka analiza drugiego rzędu wymaga znacznie większej ilości danych o zdarzeniach. Niemniej jednak, niewiele procesów występuje w stanie równowagi i zrozumienie nieoczekiwanych zmian ma znaczenie podstawowe w zarządzaniu procesami. Z tego powodu, wymagane są dalsze badania i wsparcie narzędziowe w zakresie analizy nieoczekiwanych zmian.

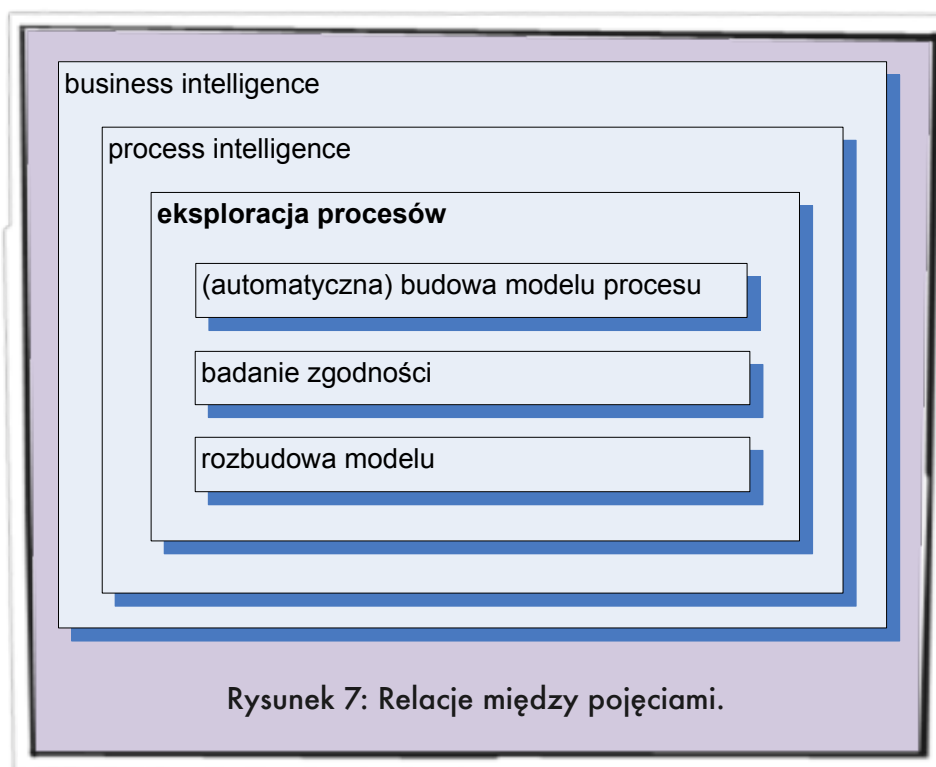
### W5. Postęp w zakresie przewyciężenia obciążenia reprezentacji w budowie modeli procesów

Techniki budowy modeli procesów wykorzystują określony język formalny (np. BPMN czy sieci Petriego). Jednakże, wizualizacja danych wynikowych ma zupełnie inny charakter, niż reprezentacja procesu wykorzystywana w budowie modelu na podstawie danych o zdarzeniach. Wybór docelowego języka obejmuje zazwyczaj kilka dodatkowych niejawnych założeń, co ogranicza przestrzeń poszukiwań. Nie można zbudować modelu procesów, które nie mogą być reprezentowane w wybranym języku. To "obciążenie reprezentacji" obecne w eksploracji procesów powinno być świadomie brane pod uwagę i nie powinno wiązać się z preferencjami w wyborze języka do wizualizacji danych wynikowych.

Rozważmy rys.6. Możliwość zapisu procesów współbieżnych w wybranym języku wpływa zarówno na wizualizację budowanego modelu, jak i na klasę modeli branych pod uwagę przez algorytm. Jeśli stosownie do obciążenia reprezentacji, nie jest możliwy zapis współbieżności (patrz: rys. 6a) i sytuacja, w której wiele czynności ma tą samą etykietę (patrz: rys. 6c), wówczas algorytm może zbudować tylko taki skomplikowany model, jak na rys. 6b. Ten przykład pokazuje, że potrzebny jest bardziej staranna i wyrafinowana analiza reprezentacji obciążenia.

## W6. Balansowanie między kryteriami jakościowymi takimi, jak dopasowanie, prostota, precyzja i generalizacja

Dzienniki zdarzeń są często niekompletne, tj. zawierają tylko wybrane zachowania procesu. Modele procesów dają wzrost wykładniczy liczby różnych ścieżek przebiegu procesu lub nawet nieskończoną ich ilość (w przypadku pętli). Co więcej, prawdopodobieństwa wystąpienia ścieżek są różne. Dlatego też nierealne jest założenie, że każda możliwa ścieżka jest reprezentowana w dzienniku zdarzeń. Przykładem tego, jak nierealne jest założenie o kompletności dziennika zdarzeń, jest proces złożony z 10 czynności, które mogą być realizowane równoległe i odpowiadające im zapisy z dziennika zdarzeń zawierające 10,000 przypadków. Ogólna liczba permutacji 10 współbieżnych czynności w modelu jest rzędu  $10! = 3,628,000$ . Stąd nie jest możliwe, aby każda permutacja była obecna w dzienniku zdarzeń, w którym występuje mniej przypadków (10,000), niż jest możliwych (3,628,000). Nawet, jeśli w dzienniku zdarzeń zapisane byłyby miliony ścieżek, jest mało prawdopodobne, aby były tam obecne wszystkie permutacje. Dodatkowa komplikacja wynika z faktu, że niektóre alternatywy są mniej częste i mogą być uznane za "szum". Nie jest możliwe wówczas zbudowanie poprawnego modelu dla zachowań procesu "z takim szumem". Zbudowany model musi od tego abstrahować; lepiej wówczas zbadać zachowania mniej częste wykorzystując techniki weryfikacji modelu.



Rysunek 7: Relacje między pojęciami.

Szum i niekompletność danych są dużym wyzwaniem dla eksploracji procesów. W rzeczywistości, występują cztery rywalizujące ze sobą miary jakościowe: (a) dopasowanie, (b) prostota, (c) precyzja i (d) generalizacja. Model dobrze dopasowany uwzględnia większość zachowań zaobserwowanych w dzienniku zdarzeń. Doskonale dopasowanie modelu oznacza, że wszystkie ścieżki przebiegu procesu z dziennika zdarzeń mogą zostać odtworzone w modelu od początku do końca. Model, który opisuje dane o zdarzeniach z dziennika zdarzeń wystarczająco dokładnie i jest najprostszym, jest najlepszym modelem. Ta zasada znana jest jako brzytwa Okhama. Dopasowanie i prostota nie są wystarczające do oceny jakości budowanego modelu procesu. Np. łatwo jest zbudować niezwykle prosty model sieci Petriego (tzw. model FLOWer), który odtwarza wszystkie ścieżki z analizowanego dziennika zdarzeń, tyle że dotyczy to również każdego innego dziennika odwołującego się do tego samego zbioru czynności. Nie ma również sensu budowanie modelu, który odtwarza wyłącznie określone zachowania zaobserwowane w dzienniku zdarzeń. Pamiętajmy, że dziennik zdarzeń zawiera tylko przykładowe zachowania i że wiele dopuszczalnych ścieżek przebiegu procesu, nie zostało jeszcze zaobserwowanych i zapisanych w dzienniku. Precyzja modelu oznacza, że nie dopuszcza on zbyt

dużej różnorodności zachowań. Oczywiście w modelu FLOWer brak jest precyzji. Model, który nie jest precyzyjny, jest "niedopasowany". Niedopasowanie występuje wówczas, gdy model dokonuje zbyt dużej generalizacji przykładów zachowań pochodzących z dziennika zdarzeń (tj. model dopuszcza zachowania dalece odbiegające od zapisanych w dzienniku). Model powinien posiadać zdolność generalizacji, tak by nie ograniczać się tylko do przykładów zapisanych w dzienniku. Model bez zdolności do generalizacji jest zbyt dokładnie dopasowany do danych, co oznacza, że budowany jest bardzo specyficzny model niekoniecznie odzwierciedlający badaną regularność. Wiadomo bowiem, że dziennik zdarzeń przechowuje dane odnoszące się tylko do przykładowych zachowań (tj. specyficzny model ogranicza się do przykładowego dziennika zdarzeń, następny dziennik zdarzeń dotyczący tego samego procesu może wygenerować zupełnie inny specyficzny model).

Balansowanie między wymaganiami miar jakościowych: dopasowania, prostoty, precyzji i generalizacji jest sporym wyzwaniem. Z tego powodu większość skutecznych technik eksploracji procesów jest w różny sposób sparametryzowana. Występuje potrzeba udoskonalenia algorytmów w celu zrównoważenia tych rywalizujących ze sobą miar jakościowych. Co więcej, każdy



wykorzystany parametr powinien być zrozumiały dla użytkownika końcowego.

## W7. Eksploracja międzyorganizacyjna (eksploracja procesów między współpracującymi organizacjami)

Tradycyjnie, eksploracja procesów wykorzystywana jest wewnątrz pojedynczej organizacji. Jednakże, w miarę rozwoju technologii udostępniania aplikacji, integracji łańcuchów dostaw i obliczeń w chmurze pojawiają się scenariusze, zgodnie z którymi udostępniane są do analizy dzienniki zdarzeń współpracujących ze sobą organizacji. W szczególności występują dwa warianty *międzyorganizacyjnej eksploracji procesów*.

Po pierwsze, możemy rozważyć sytuację, w której różne organizacje współpracują ze sobą w realizacji tego samego przebiegu procesu. Można to porównać do układania puzzli, w których każdy element jest podprocesem realizowanym przez współpracujące organizacje. Z tych podprocesów można ułożyć całość (puzzle/przypadek). Analiza dzienników zdarzeń z pojedynczej organizacji jest niewystarczająca. Aby zbudować model całego procesu, dzienniki zdarzeń różnych organizacji powinny zostać połączone (analizowane łącznie). Nie jest to proste zadanie, ponieważ zdarzenia wykraczające poza granice organizacyjne muszą być skorelowane.

Po drugie, możemy rozważyć sytuację, w której różne organizacje realizują te same procesy dzieląc się doświadczeniem, wiedzą lub wspólną infrastrukturą. Np. Salesforce.com zarządza procesami sprzedaży wielu swoich klientów. Z jednej strony, klienci współdzielą infrastrukturę (procesy, bazy danych, itd.), z drugiej - nie są oni zmuszeni do wyboru konkretnego modelu procesu, ponieważ można tak skonfigurować system, aby ten sam proces był obsługiwany przez różne modele. Inny przykład to procesy obsługi klienta w urzędzie miasta (np. procedury związane z pozwoleniami). Pomimo tego, że każdy magistrat w dowolnym kraju realizuje podobne procesy, mogą występować pomiędzy

nimi różnice. Oczywiście, nas interesuje analiza tych różnic. Organizacje mogą uczyć się od siebie nawzajem, a dostawcy usług mogą doskonalić swoje usługi i oferować wartość dodaną wypracowaną na podstawie wyników międzyorganizacyjnej eksploracji procesów.

Istnieje potrzeba opracowania nowych technik analizy obu wariantów międzyorganizacyjnej eksploracji procesów. Techniki te powinny uwzględniać aspekty prywatności i bezpieczeństwa. Organizacje mogą nie chcieć dzielić informacji z powodów konkurencji lub braku zaufania. Dlatego ważne jest opracowanie technik eksploracji procesów zapewniających prywatność.

## W8. Wspomaganie operacyjne

Początkowo, eksploracja procesów koncentrowała się na analizie danych historycznych. Jednakże obecnie, wiele źródeł danych jest aktualizowana w czasie (prawie) rzeczywistym. Dostępna jest też odpowiednia moc obliczeniowa do analizy zdarzeń w momencie ich pojawienia się. Z tego powodu, eksploracja procesów nie powinna ograniczać się do analizy off-line, a powinna być wykorzystywana do wspomaganie operacyjnego w czasie rzeczywistym. Można zdefiniować trzy następujące kierunki działań operacyjnych: rozpoznanie, predykcja, rekomendacja. Rozpoznanie polega na zidentyfikowaniu momentu, w którym przebieg procesu odbiega od jego modelu, generując stosowny alert. Często chcemy generować takie powiadomienia natychmiast (aby móc na nie reagować), a nie w trybie off-line.

Dane historyczne mogą być wykorzystywane do budowy modeli predykcyjnych. Mogą one być wykorzystywane do sterowania uruchomionymi instancjami procesów. Np. możliwa jest predykcja czasu pozostałego do wykonania procesu dla danego przypadku. Na podstawie tej predykcji, można również zbudować system rekomendacji, który zaproponuje określone działania w celu zredukowania kosztów lub skrócenia czasu przepływu. Zastosowanie technik eksploracji procesów w czasie rzeczywistym jest dodatkowym wyzwaniem w

Manifest został po raz pierwszy opublikowany w Business Process Management Workshop 2011, Lecture Notes in Business Information Proceedings, vol. 99, Springer Verlag, 2011 i został przetłumaczony na różne języki (patrz: strona internetowa grupy roboczej IEEE dotyczącej eksploracji procesów - IEEE Task Force on Process Mining;: <http://www.win.tue.nl/ieeetfpm/>.

sensie mocy obliczeniowej i jakości danych.

## W9. Połączenie eksploracji procesów z innymi rodzajami analiz

Zarządzanie operacyjne, a w szczególności badania operacyjne to gałęzie nauk o zarządzaniu, które kładą duży nacisk na wykorzystanie modelowania. Dostępna jest duża liczba modeli matematycznych, począwszy od programowania liniowego i planowania projektów, do modeli kolejek, łańcuchów Markowa i symulacji. Eksplorację danych można zdefiniować jako "analizę (często dużych) zbiorów danych w celu odnalezienia nieoczekiwanych powiązań i agregacji danych w nowy sposób, który jest zarówno zrozumiały, jak i użyteczny dla właściciela danych". Zostało opracowane szerokie spektrum technik: klasyfikacja (np. drzewa klasyfikacyjne), regresja, grupowanie (np. k-średnich), rozpoznawanie wzorców (np. reguły asocjacyjne).

Obie dziedziny (zarządzanie operacyjne i eksploracja danych) dostarczają wartościowych technik analitycznych. Sporym wyzwaniem jest połączenie tych technik z eksploracją procesów. Rozważmy np. symulację. Techniki eksploracji procesów mogą być wykorzystane do uczenia modelu symulacji na podstawie danych historycznych. Następnie, model symulacji może być wykorzystany do "wspomaganie operacyjnego". Ze względu na ścisłe powiązanie dziennika zdarzeń z modelem, model może zarówno odtwarzać zdarzenia historyczne, jak i przeprowadzić symulację od bieżącego stanu procesu poprzez element sterowania symulacją na podstawie

danych bieżących w postaci przycisku "szybkiego przewijania".

Pożądaną jest również wykorzystanie eksploracji procesów łącznie z graficzną analizą danych. Analiza wizualna ma za zadanie wyodrębnić i graficznie przedstawić powiązań pomiędzy obiektami i zdarzeniami, ukrytych w danych dostarczanych przez dzienniki zdarzeń, łącząc automatyczną analizę danych z interaktywną wizualizacją dla lepszego zrozumienia dużych i złożonych zbiorów danych. Analizy wizualne wykorzystują fenomen ludzkiego wzroku do rozpoznawania wzorców w nieustrukturyzowanych danych prezentowanych we wzorach graficznych. Poprzez połączenie technik automatycznej eksploracji procesów z interaktywną analityką wizualną możliwe jest uzyskanie większej liczby spostrzeżeń dotyczących danych o zdarzeniach.

## W10. Wzrost użyteczności dla użytkownika

Jednym z celów eksploracji procesów jest utworzenie "dynamicznego modelu procesu", tj. modelu, który może pracować w sposób nieprzerwany, a nie modelu statycznego, który nie będzie przydatny i zostanie odłożony na półkę. Nowe dane o zdarzeniach mogą być wykorzystane do odkrywania zachowań nie uwzględnionych w dotychczasowym modelu. Powiązanie między danymi o zdarzeniach i modelami procesów powoduje dostosowanie aktualnego modelu do bieżącego stanu i ostatnich aktywności. Z tego powodu, interakcja użytkownika z modelem powinna być uwzględniana w jego codziennej pracy. Taka interakcja jest bardzo cenna, ale wymaga intuicyjnego interfejsu użytkownika. Wyzwaniem jest ukrycie wyrafinowanych algorytmów eksploracji procesów za przyjaznym dla użytkownika interfejsem, który automatycznie ustala parametry i sugeruje odpowiedni rodzaj analizy.

## W11. Poprawa czytelności dla użytkownika

Nawet jeśli uzyskanie danych wyjściowych z modelu eksploracji procesów jest łatwe, nie oznacza to, że wyniki są rzeczywiście przydatne. Użytkownik może mieć problemy ze

zrozumieniem wyników lub ich interpretacją. Wyniki powinny być prezentowane w odpowiedni sposób, tak aby uniknąć tego typu problemów (por. PZ5). Co więcej, wiarygodność uzyskanych wyników powinna być jasno określona. W niektórych przypadkach może brakować danych do wyciągnięcia określonych wniosków. W rzeczywistości, dostępne obecnie techniki eksploracji procesów nie ostrzegają przed słabym lub nadmiernym dopasowaniem do danych. Zawsze pokazują w danych wyników model, nawet jeśli w sposób oczywisty występowała niewystarczająca liczba danych wejściowych do jego zbudowania.

## Epilog

Grupa robocza do spraw eksploracji procesów przy IEEE ma na celu (a) promowanie zastosowań eksploracji procesów, (b) wskazywanie twórcom oprogramowania, konsultantom, menadżerom biznesowym i użytkownikom, kiedy stosować dostępne obecnie techniki i (c) stymulować badania dotyczące eksploracji procesów. Manifest określa podstawowe zasady i intencje grupy roboczej. Po wprowadzeniu do eksploracji procesów, manifest kataloguje niektóre przewodnie zasady (sekcja 3) i wyzwania (sekcja 4). Przewodnie zasady mogą być wykorzystane w celu uniknięcia oczywistych błędów. Lista wyzwań ma na celu wskazanie kierunków badań i opracowań rozwiązań komercyjnych. Oba zagadnienia mają na celu zwiększenie poziomu dojrzałości eksploracji procesów.

W ramach podsumowania, kilka słów na temat terminologii. Następujące terminy są wykorzystywane w obszarze eksploracji procesów: eksploracja przepływu pracy, eksploracja procesów biznesowych, automatyczne odkrywanie wiedzy z procesów biznesowych, process intelligence. Różne organizacje wykorzystują różną terminologię dla nakładających się koncepcji. Np. Gartner promuje termin automatyczne odkrywanie wiedzy z procesów biznesowych (Automated Business Process Discovery - ABPD), Software AG używa terminu Process Intelligence w odniesieniu do swojej platformy sterowania. Termin eksploracja przepływu pracy wydaje się

mniej odpowiedni, ponieważ budowa modeli przepływu pracy jest tylko jednym z możliwych zastosowań eksploracji procesów. Podobnie, dodanie określenia biznesowy ogranicza zakres do określonych zastosowań eksploracji procesów. Występuje duża liczba aplikacji eksploracji procesów (np. analiza użytkownika systemów high-tech czy analizowania stron webowych), w których to określenie wydaje się nieodpowiednie. Chociaż odkrywanie wiedzy w procesach jest istotną częścią spektrum eksploracji procesów, jest tylko jedną z możliwych zastosowań. Analiza zgodności, predykcja, eksploracja organizacyjna, analiza sieci społecznych, itd. są innymi możliwymi zastosowaniami, które wykraczają poza odkrywanie wiedzy w procesach.

Rys. 7 pokazuje powiązania omówionych terminów. Wszystkie technologie i metody, których celem jest dostarczenie użytecznej informacji, która może być wykorzystana do wspomagania podejmowania decyzji może być umieszczona w ramach terminu Business Intelligence. (Business) Process Intelligence może być postrzegane jako kombinacja BI i BPM, tj. techniki BI są wykorzystywane do analizy i udoskonalenia procesów i ich zarządzania. Eksploracja procesów może być postrzegana jako konkretyzacja inteligentnego procesu biorącego dane o zdarzeniach jako punkt startowy. (Automatyczne) odkrywanie wiedzy z procesów (biznesowych) jest tylko jednym z trzech podstawowych rodzajów eksploracji procesów. Rys. 7 może być trochę mylący w tym sensie, że większość narzędzi BI nie dostarcza funkcjonalności eksploracji procesów jak opisano w niniejszym dokumencie. Termin BI jest często zwyczajowo odnoszony do określonych narzędzi lub metod pokrywających tylko małą część szerszego spektrum BI.

Mogą występować komercyjne powody do używania alternatywnych terminów. Niektórzy dostawcy mogą także podkreślać szczególny aspekt (tj. odkrywanie, czy inteligencja). Jednakże, aby uniknąć zamieszania, lepiej jest wykorzystywać termin eksploracja procesów w odniesieniu do dyscypliny opisanej w manifestie.

# Słownik

## **Automatyczna budowa modelu**

**procesu biznesowego:** zobacz: budowa modelu procesu.

**Badanie zgodności:** sprawdzenie, czy przebiegi procesu zapisane w dzienniku zdarzeń odpowiadają modelowi procesu i na odwrót; celem jest wykrycie rozbieżności i ich oszacowanie; badanie zgodności jest jednym z trzech podstawowych rodzajów eksploracji procesów.

**Budowa modelu procesu:** jeden z trzech podstawowych rodzajów eksploracji procesów; uczenie modelu realizowane jest na danych z dziennika zdarzeń, np. algorytm alfa na podstawie danych z dziennika zdarzeń jest w stanie zidentyfikować wzorce przebiegu procesu i przedstawić model przepływu sterowania w procesie jako sieć Petriego.

**Business Intelligence (BI):** szeroki wachlarz narzędzi i metod wykorzystujących dane do wspomagania procesów decyzyjnych.

**Business Process Intelligence:** zobacz Process Intelligence.

**Czynność:** wyraźnie określony krok procesu; zdarzenia mogą dotyczyć rozpoczęcia, realizacji, anulowania, itp. czynności dla konkretnej instancji procesu.

**Dopasowanie:** miara określająca, w jakim stopniu model odwzorowuje proces zapisany w dzienniku zdarzeń; model ma idealne dopasowanie, jeśli każde zdarzenie zapisane w dzienniku zdarzeń jest reprezentowane w modelu (model odwzorowuje wówczas cały przebieg procesu).

**Dziennik zdarzeń:** zbiór zdarzeń wykorzystywany jako dane wejściowe do eksploracji procesu; zdarzenia nie muszą pochodzić z jednego dziennika zdarzeń (tj. mogą być przechowywane w oddzielnych zbiorach zdarzeń, zdarzenia mogą być rozproszone w różnych tabelach bazy danych).

**Eksploracja danych:** analiza (często dużych) zbiorów danych w celu odkrycia nieoczekiwanych zależności i obliczenia statystyk w sposób

umożliwiający uzyskanie nowego spojrzenia na dane.

**Eksploracja procesów:** techniki, narzędzia i metody wykorzystywane do budowy modelu procesu, jego weryfikacji i rozbudowy na podstawie danych o przebiegu rzeczywistych procesów (tj. nie hipotetycznych) pochodzących z dzienników zdarzeń powszechnie dostępnych we współczesnych systemach informatycznych.

**Instancja procesu:** pojedynczy przebieg analizowanego procesu; dziennik zdarzeń zawiera dane tworzone przez instancje procesu; przykładami instancji procesu są zamówienia klientów, zgłoszenie szkody, wnioski kredytowy, itd.

**Międzyorganizacyjna eksploracja procesów:** eksploracja procesów zastosowana do dzienników zdarzeń pochodzących z różnych organizacji.

**MXML:** format danych, o składni wzorowanej na XML, wykorzystywany do zapisywania zdarzeń z przebiegu procesu w dzienniku zdarzeń; obecnie w eksploracji procesów format MXML jest zastępowany przez nowy niezależny sprzętowo format danych - XES.

**Obciążenie języka reprezentacji:** wybór notacji dla prezentacji wyników eksploracji procesów.

**Precyzja:** miara określająca dopuszczalne odchylenie przebiegu procesu od jego modelu, model o niskiej precyzji jest "niedoszacowany".

**Process Intelligence:** dział business intelligence zajmujący się zarządzaniem procesami biznesowymi.

**Prostota:** miara odwołująca się do zasady Brzytwy Okhama, tj. najprostszemu modelowi, który może wyjaśnić zachowanie procesu zaobserwowane w dzienniku zdarzeń, jest najlepszy; prostota może być kwantyfikowana w różny sposób, np. liczbą węzłów i krawędzi w modelu procesu.

**Przypadek:** zobacz instancja procesu.

**Rozbudowa modelu:** jeden z trzech podstawowych rodzajów eksploracji procesów; model procesu jest rozszerzany i doskonalony na podstawie danych z dziennika zdarzeń; przykładowo, można zidentyfikować wąskie

gardła w procesie uruchamiając przebieg procesu na modelu i analizując znaczniki czasu dla zdarzeń.

**Uogólnienie:** miara określająca, w jakim stopniu model odwzorowuje niezaobserwowane (niezapisane w dzienniku zdarzeń) zachowania procesu; nadmierne dopasowanie do dziennika zdarzeń model, nie jest w stanie uogólnić w stopniu wystarczającym zachowania procesu.

**Wsparcie operacyjne:** analiza danych z dziennika zdarzeń w czasie rzeczywistym w celu monitorowania przebiegu procesu oraz wpływania na jego wykonanie; w ramach wsparcia operacyjnego, wyróżnia się trzy rodzaje działań: wykrywanie (generuje powiadomienie, jeśli obserwowany przebieg procesu odbiega od jego modelu), predykcja (przewidywanie dalszego przebiegu procesu na podstawie przebiegu dotychczasowego, np. określenie czasu, jaki pozostał do zakończenia procesu), rekomendacja (proponowanie właściwych działań prowadzących do wyznaczonego celu, np. minimalizacji kosztów).

**XES:** standard formatu danych z dziennika zdarzeń odwołujący się do składni XML; uznany przez IEEE Task Force jako domyślny format wymiany danych o zdarzeniach między narzędziami do monitorowania zdarzeń i aplikacjami do eksploracji procesów (tj. [www.xes-standard.org](http://www.xes-standard.org)).

**Zarządzanie procesami biznesowymi:** dziedzina wiedzy, która stanowi połączenie techniki informacyjnej z naukami o zarządzaniu w zakresie ich wykorzystania w operacyjnych procesach biznesowych.

**Zdarzenie:** działanie zapisane w dzienniku zdarzeń; np. rozpoczęcie, zakończenie, anulowanie czynności dla instancji procesu.

**Zmienne środowisko procesu:** zjawisko polegające na częstych zmianach procesów w czasie; obserwowany proces może stopniowo (lub nagle) zmienić się z powodu zmian sezonowych lub wzrastającej konkurencji, co komplikuje analizę.



## Autorzy

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 Don-  
gen  
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 Nez-  
had  
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