

MANAGEMENTUL PROCESELOR DE AFACERI - PROVOCĂRI ȘI PERFORMANȚE ACTUALE

Teză destinată obținerii
titlului științific de doctor
la
Universitatea Politehnica Timișoara
în domeniul INGINERIE ȘI MANAGEMENT

ing. Cristian-Ilie OLARIU

Conducător științific: prof.univ.dr.ing. Anca DRĂGHICI
Referenți științifici: prof.univ.dr.ing. Claudiu-Vasile KIFOR
prof.univ.dr.ing. Călin-Florin BĂBAN
prof.univ.dr.ing. George-Gustav SAVII

Ziua susținerii tezei: 12.09.2015

Seriile Teze de doctorat ale UPT sunt:

- | | |
|---|--|
| 1. Automatică | 10. Știința Calculatoarelor |
| 2. Chimie | 11. Știința și Ingineria Materialelor |
| 3. Energetică | 12. Ingineria sistemelor |
| 4. Ingineria Chimică | 13. Inginerie energetică |
| 5. Inginerie Civilă | 14. Calculatoare și tehnologia informației |
| 6. Inginerie Electrică | 15. Ingineria materialelor |
| 7. Inginerie Electronică și Telecomunicații | 16. Inginerie și Management |
| 8. Inginerie Industrială | 17. Arhitectură |
| 9. Inginerie Mecanică | 18. Inginerie civilă și instalații |

Universitatea Politehnica din Timișoara a inițiat seriile de mai sus în scopul diseminării expertizei, cunoștințelor și rezultatelor cercetărilor întreprinse în cadrul scolii doctorale a universității. Seriile conțin, potrivit H.B.Ex.S Nr. 14 / 14.07.2006, tezele de doctorat susținute în universitate începând cu 1 octombrie 2006.

Copyright © Editura Politehnica – Timișoara, 2015

Această publicație este supusă prevederilor legii dreptului de autor. Multiplicarea acestei publicații, în mod integral sau în parte, traducerea, tipărirea, reutilizarea ilustrațiilor, expunerea, radiodifuzarea, reproducerea pe microfilme sau în orice altă formă este permisă numai cu respectarea prevederilor Legii române a dreptului de autor în vigoare și permisiunea pentru utilizare obținută în scris din partea Universității Politehnica din Timișoara. Toate încălcările acestor drepturi vor fi penalizate potrivit Legii române a drepturilor de autor.

România, 300159 Timișoara, Bd. Republicii 9,
tel. 0256 403823, fax. 0256 403221
e-mail: editura@edipol.upt.ro

CUVÂNT ÎNAINTE

La finalizarea și susținerea publică a prezentei lucrări, cu sfială și mult respect, îi mulțumesc doamnei conducător științific de doctorat, *prof.univ. dr. ing. Anca DRĂGHICI*, care mi-a îndrumat pașii în hățișul cercetării și fără de care nu aș fi putut finaliza și susține în bune condiții teza. Sprijinul necondiționat pe care l-am avut în acești 5 ani din partea Doamnei Profesor a venit totdeauna când mă aflam la o răscruce a cercetării, dar și a vieții mele. Modul dinamic, dar natural cu care și-a făcut simțită prezența în modelarea și desăvârșirea mea profesională, a fost de fiecare dată extrem de inspirat și încurajator. Spațiul și timpul nu au fost obstacole în calea colaborării noastre, iar dilemele și întrebările mele au avut soluții simple, fiind percepute în etape ale evoluției demersului de cercetare și ale devenirii mele ca specialist.

Vă mulțumesc pentru oferirea ocaziei de a face parte din echipa proiectului Leonardo da Vinci (Transfer of Innovation): *Certified Business Process Manager – CertiBPM* (contr. nr. LLP-LdV/TOI/10/RO/010), care m-a ajutat în realizarea activității de cercetare și publicistice.

Se cuvin aduse mulțumiri Comisiei de referenți științifici:

- În primul rând, Domnului *prof. univ. dr. ec. ing. Marian-Liviu MOCAN*, decanul Facultății de Management în Producție și Transporturi, *președinte al Comisiei de Susținere Publică a tezei*, care a acceptat cu amabilitate să prezideze ședința și să-mi fie alături în cea mai importantă zi a vieții mele profesionale din ultimii ani;
- Domnului *prof. univ. dr. ing. Claudiu-Vasile KIFOR* (Universitatea „Lucian Blaga” din Sibiu), care prin pedanta analiză a tezei și observațiile pertinente realizate a contribuit la rafinarea unor detalii ale lucrării și evidențierea mai bună a rezultatelor de cercetare obținute în cadrul programului doctoral. Mulțumesc pentru aprecierile făcute și pentru recenzarea tezei;
- Domnului *prof. univ. dr. ing. Călin-Florin BĂBAN* (Universitatea din Oradea) pentru dialogul purtat în ideea clarificării unor aspecte privind demersul abordat, precum și pentru recenzia profesionistă a tezei de doctorat și pentru observațiile făcute care au contribuit la structurarea aspectelor de ordin metodologic prezente în lucrare;
- Domnului *prof. univ. dr. ing. George-Gustav SAVII* (Universitatea Politehnica Timișoara) pentru acceptul de a analiza și a rafina teza de doctorat în ceea ce privește abordările din perspectiva științei calculatoarelor. Totodată, țin să-i mulțumesc pentru ideile oferite în momente critice, pentru explicarea unor aspecte relative la concluziile și rezultatele ce au decurs din aplicarea scenariului de cercetare. Încurajarea permanentă din partea domniei sale m-a motivat spre finalizarea programului doctoral și susținerea tezei în bune condiții.

De-a lungul celor cinci ani de studii doctorale au fost alături de mine membrii Comisiei de Îndrumare: doamna *prof. univ. dr. ing. Gabriela PROȘTEAN*, domnul *prof. univ. dr. ing. Constantin - Dan DUMITRESCU* și domnul *prof. dr. ing. Marian MOCAN*, cărora vreau să le mulțumesc pentru observațiile profesioniste făcute la susținerea referatelor mele, observații ce au contribuit la îmbunătățirea demersului de cercetare și la finalizarea în bune condiții a tezei de doctorat.

Mulțumesc pentru înțelegere și sprijin tuturor colegilor de la SC Wirtek SRL Cluj-Napoca și de la Innomate A/S Danemarca și care m-au inspirat în momentele decizionale critice ale cercetărilor, au înțeles să mă degrevez de unele sarcini atunci când activitatea de la programul doctoral a fost mai intensă sau când am participat la conferințe naționale sau internaționale.

Mulțumesc domnului dr. Joos Buijs de la Universitatea Tehnică din Eindhoven, Olanda pentru punerea la dispoziția a datelor ce au ajutat la validarea cercetării.

În final, mulțumesc familiei mele pentru răbdarea, înțelegerea și ajutorul continuu oferit pe parcursul activității mele din cadrul doctoratului (2010 - 2015) și nu numai. Mulțumesc tuturor prietenilor care m-au înțeles, m-au ajutat și au fost lângă mine în toți acești ani.

Timișoara, septembrie 2015

ing. Cristian-Ilie OLARIU

Olariu, Cristian-Ilie

**MANAGEMENTUL PROCESELOR DE AFACERI - PROVOCĂRI ȘI
PERFORMANȚE ACTUALE**

Teze de doctorat ale UPT, Seria 16, Nr. 19, Editura Politehnica, 2015, 201 pagini, 113 figuri, 20 tabele.

ISSN:2343-7928

ISSN-L:2343-7929

ISBN:978-606-554-987-6

Cuvinte cheie: Business Process Management (BPM), Business Intelligence, Process Mining, Data Mining, Modelarea proceselor, Mijloace software inteligente, Proces de afaceri

Rezumat,

Prezenta teză de doctorat este o lucrare de cercetare interdisciplinară, care se situează ca problematică la intersecția domeniilor de știința calculatoarelor și management (având reverberații în zona sistemelor informatice de management, a managementului performanței și a managementului operațional și strategic), recurgând la utilizarea, exploatarea și valorizarea unor cunoștințe diverse aferente spațiului complex al ingineriei și managementului. Obiectivul general a fost acela de a concepe o metodologie de implementare a proiectelor BPM în cadrul organizațiilor, care să permită o mai bună susținere a deciziei manageriale strategice și să confere informații pentru afaceri (testarea și validarea sa). Metodologia elaborată poartă numele *CertiBPM-PM* și este o metodologie care îmbină elemente clasice de implementare a unui proiect BPM cu tehnici specifice Process Mining.

CUPRINS

CUVÂNT ÎNAINTE	3
CUPRINS	5
NOTAȚII, ABREVIERI, ACRONIME	8
LISTA DE FIGURI	11
LISTA DE TABELE	14
INTRODUCERE	15
a. Motivarea și importanța temei de cercetare alese.....	15
b. Obiectivele cercetării și modul de atingere a acestora (panoramarea structurii tezei de doctorat).....	17
1. CADRUL CONTEXTUAL ȘI CONCEPTUAL AL CERCETĂRII	23
1.1. Managementul proceselor de afaceri (BPM) - istoric și necesitate.....	23
1.2. Definiții și noțiuni cheie ale abordării BPM.....	29
1.3. BPM și sistemul de management al organizației.....	36
1.4. Nevoia de certificare BPM pe piața din România.....	41
1.5. Domeniul Process Mining în contextul BPM.....	43
1.6. Concluzii.....	45
2. CUNOȘTINȚE PRELIMINARE PENTRU REALIZAREA DEMERSULUI DE PROCESS MINING	47
2.1. Modelarea proceselor de afaceri.....	47
2.1.1. Rețele Petri.....	48
2.1.2. Rețele Flux de Lucru.....	51
2.1.3. Rețele cauzale tip <i>C-nets</i>	52
2.1.4. Lanțul Procesului Conducut de Evenimente (EPCs).....	54
2.1.5. Diagrame YAWL.....	55
2.1.6. Diagrame BPMN.....	57
2.2. Data Mining.....	60
2.2.1. Definiție. Caracteristici. Beneficii.....	60
2.2.2. Tehnici pentru Data Mining.....	66
2.2.2.1. Clasificarea, regresia și clustering-ul.....	68
2.2.2.2. Analiza secvențială sau a succesiunilor.....	69
2.2.2.3. Tehnica rețelelor neuronale.....	69
2.2.2.4. Tehnica arborilor de decizie.....	71
2.2.2.5. Tehnica Naive Bayes.....	73
2.2.2.6. Tehnica k-NN.....	73
2.2.2.7. Comparație între tehnicile de Data Mining.....	74
2.3. Concluzii.....	76
3. CERCETĂRI TEORETICE PRIVIND PROBLEMATICA REALIZĂRII PROIECTELOR DE PROCESS MINING	78
3.1. Descrierea generală a proiectelor de Process Mining.....	78
3.1.1. Process Mining.....	78
3.1.2. Relația dintre modelul de proces și jurnalul de evenimente.....	79
3.2. Imaginea de ansamblu a unui proiect de Process Mining.....	79

3.3. Pregătirea datelor	80
3.3.1. Surse de date.....	80
3.3.2. Definirea termenilor folosiți	82
3.3.3. Formate suportate	83
3.3.4. Calitatea datelor extrase.....	86
3.3.5. Directive pentru înregistrarea evenimentelor	88
3.4. Descoperirea procesului	89
3.4.1. Algoritmul α	89
3.4.2. Algoritmi avansați	93
3.4.2.1. Algoritmul HeuristicsMiner	93
3.4.2.2. Algoritmul Fuzzy Miner	96
3.4.2.3. Algoritmul Genetic Process Mining	96
3.4.2.4. Algoritmi bazați pe teoria regiunilor	97
3.4.2.5. Algoritmul Inductive Miner	98
3.4.2.6. Alți algoritmi	98
3.5. Verificarea conformității	99
3.5.1. Verificarea conformității folosind amprente de cauzalitate	100
3.5.2. Verificarea conformității folosind reluarea bazată pe jetoane	101
3.5.3. Verificarea conformității bazată pe aliniere	103
3.6. Suportul operațional detecția, predicția și recomandarea	105
3.7. Alte perspective ale Process Mining	106
3.7.1. Mining-ul punctelor de decizie	108
3.7.2. Mining-ul operațional.....	109
3.7.3. Mining-ul rețelelor sociale	111
3.7.4. Mining-ul organizațional.....	113
3.7.5. Combinarea diferitelor perspective	114
3.8. Cadru de lucru pentru Process Mining rafinat.....	116
3.9. Clasificarea modelelor obținute	118
3.9.1. Procese Lasagna	118
3.9.2. Procese Spaghetti	119
3.9.3. Clasificarea funcțiilor organizației după tipul de procese.....	119
3.10. Concluzii	120
4. MIJLOACE SOFTWARE INTELIGENTE CARE SUSȚIN CREȘTEREA	
PERFORMANȚELOR PROCESELOR DE AFACERI	121
4.1. Poziționarea mijloacelor software inteligente pentru analiza proceselor	121
4.2. Mijloace software pentru Process Mining	122
4.2.1. Platforma ProM.....	123
4.2.1.1. ProM 6	123
4.2.1.2. ProM 6 Package Manager	126
4.2.1.3. XESame	126
4.2.2. Disco.....	129
4.2.3. Alte aplicații software pentru Process Mining	130
4.3. Programe software pentru Data Mining	130
4.3.1. RapidMiner.....	131
4.3.2. Extensia RapidProM.....	131
4.4. Concluzii	133
5. METODOLOGIA DE IMPLEMENTARE A UNUI PROIECT BPM CU AJUTORUL	
PROCESS MINING	134
5.1. Implementarea unui proiect BPM în viziunea <i>CertiBPM</i>	134

5.2. Metodologii de implementare Process Mining existente	137
5.2.1. Metodologia PDM	137
5.2.2. Metodologia PDM în domeniul medical	138
5.2.3. Modelul L*	139
5.2.4. Metodologia PM ²	143
5.3. Metodologia CertiBPM îmbunătățită cu elemente de Process Mining	144
5.4. Concluzii	147
6. CERCETĂRI APLICATIVE PRIVIND APLICAREA DEMERSULUI METODOLOGIC CERTIBPM-PM ÎN VEDEREA IDENTIFICĂRII UNOR CĂI DE CREȘTERE A PERFORMANȚELOR PROCESELOR DE AFACERI	149
6.1. Descrierea contextului de cercetare	149
6.2. Considerente privind modul de aplicare a metodologiei de cercetare	150
6.2.1. Obținerea jurnalului	150
6.2.2. Inspecția jurnalului	151
6.2.3. Descoperirea procesului	161
6.2.3.1. Algoritmul α	161
6.2.3.2. Algoritmul HeuristicsMiner	163
6.2.3.3. Algoritmul Fuzzy Miner	165
6.2.3.4. Algoritmul Inductive Miner	167
6.2.4. Compararea și alegerea modelului de proces	168
6.2.5. Analiza performanței	170
6.2.6. Analiza organizațională	174
6.2.7. Raportul final al analizei	176
6.2.8. Transferul rezultatelor	180
6.3. Concluzii	180
7. CONCLUZII GENERALE, CONTRIBUȚII PERSONALE ȘI PERSPECTIVELE CERCETĂRII	182
7.1. Concluzii generale asupra cercetării realizate	182
7.2. Contribuții personale	185
7.2.1. Contribuții în planul cercetărilor asupra referențialului bibliografic	185
7.2.2. Contribuții în planul cercetărilor teoretice	186
7.2.3. Contribuții în planul cercetărilor aplicative și experimentale	186
7.3. Direcții viitoare de cercetare	186
BIBLIOGRAFIE	188
ANEXA 1 – Poster conținând notații pentru modelarea proceselor de afaceri (BPMN)	199
ANEXA 2 – Jurnal de evenimente în format XES	200
ANEXA 3 – Activitățile jurnalului de evenimente considerate în cadrul cercetărilor aplicative	201

NOTAȚII, ABREVIERI, ACRONIME

N,A,A	Descriere (română)	N,A,A	Descriere (engleză)
APA	Arhitectura Proceselor de Afaceri	BPA	Business Process Architecture
ASII	Arhitectura Sistemelor Informatice Integrate	ARIS	Architecture of Integrated Information Systems
DCD	Descoperirea Cunoașterii din bazele de Date	KDD	Knowledge Discovery in Databases
DD	Depozite de Date	DW	Data Warehouses
DM	Data Mining	DM	Data Mining
ETI	Extragere, Transformare, Încărcare	ETL	Extract, Transform, Load
FEE	Flux Extensibil de Evenimente	XES	eXtensible Event Stream
GMO	Grupul Managementului de Obiect	OMG	Object Management Group
IALF	Încă Alt Limbaj pentru Flux de Lucru	YAWL	Yet Another Workflow Language
IA	Informații pentru Afaceri	BI	Business Intelligence
ÎA	Învățare automată	ML	Machine Learning
IAE	Integrarea Aplicațiilor Întreprinderii	EAI	Enterprise Application Integration
ICP	Indicator Cheie de Performanță	KPI	Key Performance Indicator
IE	Indicatori de Echilibrare	BS	Balanced Scorecards
IMMC	Integrarea Modelului de Maturitate a Capabilității	CMMI	Capability Maturity Model Integration
IMPA	Inițiativa Managementului Procesului de Afaceri	BPMI	Business Process Management Initiative
IPA	Informații pentru Procesele de Afaceri	BPI	Business Process Intelligence
IPAp	Interfață pentru Programarea de Aplicații	API	Application Programming Interface
LEPA	Limbaj de Execuție a Proceselor de Afaceri	BPEL	Business Process Execution Language
LME	Limbaj de Marcare Extensibil	XML	eXtensible Markup Language
LMEM	Limbaj de Marcare Extensibil pentru Mining	MXML	Mining eXtensible Markup Language
LMEM-NS	Limbaj de Marcare Extensibil pentru Mining Notat Semantic	SA-MXML	Semantically Annotated Mining eXtensible Markup Language
LMPA	Limbaj de Modelare Proces de Afaceri	BPML	Business Process Modeling Language

LPE	Lanțul Procesului condus de Evenimente	EPC	Event-driven Process Chain
MAA	Monitorizarea Activității Afacerii	BAM	Business Activity Monitoring
MoPA	Modelare a Proceselor Afacerii	BPMo	Business Process Modeling
MC	Managementul Cazurilor	CM	Case Management
MMC	Modelul de Maturitate a Capabilității	CMM	Capability Maturity Model
MCT	Managementul Calității Totale	TQM	Total Quality Management
MDP	Metodologia de Diagnosticare a Procesului	PDM	Process Diagnostics Methodology
MFL	Managementul Fluxului de Lucru	WFM	Workflow Management
MPA	Managementul Proceselor Afacerii	BPM	Business Process Management
MPPM	Metodologia de Proiect pentru <i>Process Mining</i>	PM2	<i>Process Mining</i> Project Methodology
MRA	Managementul Regulilor Afacerii	BRM	Business Rules Management
NMPA	Notația Modelării Proceselor de Afaceri	BPMN	Business Process Modeling Notation
NMPA 2.0	Notația și Modelul Proceselor de Afaceri	BPMN 2.0	Business Process Model and Notation
PA	Procesele Afacerii	BP	Business Processes
PAA	Platformă Analitică Avansată	AAP	Advanced Analytics Platform
PAOL	Procesarea Analitică OnLine	OLAP	OnLine Analytical Processing
PD	Piețe de Date	DM	Data Marts
PLNÎ	Programare Liniară În Numere Întregi	ILP	Integer Linear Programming
PM	Process Mining	PM	Process Mining
PRI	Planificarea Resurselor Întreprinderii	ERP	Enterprise Resource Planning
RBL	Regiuni bazate pe Limbaj	LBR	Language-Based Regions
RBS	Regiuni Bazate pe Stări	SBR	State-Based Regions
RFL	Rețele Flux de Lucru	WF-nets	Workflow Nets
RPA	Re-ingineria Proceselor Afacerii	BPR	Business Process Re-engineering
RPC	Rețele Petri Colorate	CPN	Colored Petri Nets
SAP	Sisteme, Aplicații & Produse în procesarea datelor	SAP	Systems, Applications & Products in data Processing
SFC	Schimbare Fără Cauză	CWC	Change Without Cause
SI	Sisteme Informatice	IS	Information Systems

SICP	Sisteme Informatice de Conștientizare a Proceselor	PAIS	Process-Aware Information Systems
SIE	Sisteme Informatice Executive	EIS	Executive Information Systems
SIMPA	Software Inteligent pentru Managementul Proceselor de Afaceri	iBPMS	Intelligent Business Process Management Software
SMFL	Sisteme pentru Managementul Fluxului de Lucru	WfMS	Workflow Management System
SMPA	Software pentru Managementul Proceselor Afacerii	BPMS	Business Process Management Software
RN	Rețele Neuronale	NN	Neural Networks
Rețele-C	Rețele Cauzale tip C-nets	C-Nets	Causal Nets
TI	Tehnologia Informației	IT	Information Technology
TIC	Tehnologia Informației și a Comunicării	ICT	Information and Communication Technology

LISTA DE FIGURI

Fig. I.1.	Modelul de tip „cutie neagră” asociat cercetării științifice prin doctorat .	18
Fig. I.2.	Logica adoptată în cercetare aferentă programului doctoral	19
Fig. I.3.	Structura tezei de doctorat	20
Fig. 1.1.	Tradiția BPM - continuare a tradiției manageriale, a controlului calității și a tehnologiei informației.....	25
Fig. 1.2.	Sinteza perspectivelor și abordărilor relevante care au contribuit la apariția și dezvoltarea BPM	30
Fig. 1.3.	Diferențe fundamentale dintre BPM și alte domenii de management.....	33
Fig. 1.4.	Ciclul de viață al procesului de afaceri în manieră BPM [22]	34
Fig. 1.5.	Structura cadru a documentației unui sistem de management.....	38
Fig. 1.6.	Tipologia proceselor	40
Fig. 1.7.	Exemplu privind tipologia proceselor într-o organizație.....	40
Fig. 1.8.	Contextul alegerii Process Mining ca soluție pentru îmbunătățirea proceselor de afaceri.....	44
Fig. 1.9.	Piața de software BPM conform BPTrends [81]	45
Fig. 2.1.	Rețea Petri și elemente de topologie specifice unui proces	48
Fig. 2.2.	Graful de accesibilitate dezvoltat din rețeaua Petri din Fig. 2.1	50
Fig. 2.3.	Graful de acoperire dezvoltat din rețeaua Petri din Fig. 2.1	50
Fig. 2.4.	Rețea cauzală tip C-nets dezvoltată din rețeaua Petri din Fig. 2.1	53
Fig. 2.5.	Notăția standard EPC	54
Fig. 2.6.	Diagramă EPC dezvoltată din rețeaua Petri din Fig. 2.1	55
Fig. 2.7.	Notăția standard YAWL.....	56
Fig. 2.8.	Diagramă YAWL dezvoltată din rețeaua Petri din Fig. 2.1	56
Fig. 2.9.	Notăția standard BPMN.....	58
Fig. 2.10.	Notății frecvent folosite în reprezentările modelelor de procese, conform BPMN (preluare din materialele de formare CertiBPM).....	58
Fig. 2.11.	Diagramă BPMN dezvoltată din rețeaua Petri din Fig. 2.1	59
Fig. 2.12.	Faza de transformare către aplicația executabilă (preluare din materialele de formare CertiBPM)	60
Fig. 2.13.	Acumularea de cunoaștere ce constituie suportul deciziei de la nivelul managementului unei organizații.....	62
Fig. 2.14.	Etapele KDD	63
Fig. 2.15.	Etapele unui proces de Data Mining	64
Fig. 2.16.	Clase de activități aplicabile în demersul de Data Mining.....	66
Fig. 2.17.	Activități posibile de Data Mining în organizație	67
Fig. 2.18.	Tehnici de Data Mining	68
Fig. 2.19.	Exemplul unei rețele neuronale cu 6 noduri [132].....	70
Fig. 2.20.	Exemplul unui arbore de decizie [132]	72
Fig. 2.21.	Luarea deciziilor cu ajutorul Process Mining și Data Mining.....	76
Fig. 3.1.	Relația dintre modelul de proces și jurnalul de evenimente	79
Fig. 3.2.	Etapele unui proiect de Process Mining	81
Fig. 3.3.	Pregătirea datelor folosite pentru Process Mining	82
Fig. 3.4.	Meta-modelul XES explicat schematic prin diagrame UML [74]	84
Fig. 3.5.	Probleme de calitate ale datelor din jurnalele de evenimente	87
Fig. 3.6.	Rețea Petri descoperită cu algoritmul α din jurnalul de evenimente L	92

Fig. 3.7.	Cele patru dimensiuni calitative care caracterizează un model de proces descoperit	93
Fig. 3.8.	Graful de dependență rezultat din jurnalul de evenimente L	95
Fig. 3.9.	Învățarea disocierilor și a unificărilor într-o rețea causală tip C-Nets adnotată cu frecvențe	95
Fig. 3.10.	Altă rețea Petri descoperită din jurnalul de evenimente $L (P2)$	100
Fig. 3.11.	Verificarea conformității la nivel de traseu prin metoda reluării bazate pe jetoane pentru rețeaua Petri $P2$	103
Fig. 3.12.	Verificarea conformității prin alinierea unui traseu cu o rulare a modelului de proces.....	104
Fig. 3.13.	Perspective ale Process Mining	107
Fig. 3.14.	Legătura dintre atributele de eveniment și proces.....	107
Fig. 3.15.	Puncte de decizie în BPMN	108
Fig. 3.16.	Tipare temporale utile Mining-ului operațional	110
Fig. 3.17.	Rețea socială cu noduri și arce ponderate	111
Fig. 3.18.	Rețea socială reprezentând transferul de muncă.....	113
Fig. 3.19.	Cum percep oamenii orbi un elefant.....	114
Fig. 3.20.	Provocările unui proiect de simulare și riscurile asociate fiecărei etape	115
Fig. 3.21.	Comparație între simularea tradițională și cea pe bază de Process Mining.....	116
Fig. 3.22.	Alte direcții aplicate în Process Mining	117
Fig. 3.23.	Cadrul de lucru pentru Process Mining rafinat.....	118
Fig. 3.24.	Clasificarea funcțiilor organizației după tipul de procese	119
Fig. 4.1.	Spectru de mijloace software pentru Process Mining și Data Mining....	121
Fig. 4.2.	Poziționarea mijloacelor software pentru Process Mining în funcție de tipul de utilizator și beneficii	122
Fig. 4.3.	ProM – interfața inițială de utilizare (captură ecran)	124
Fig. 4.4.	ProM – Spațiul de lucru (captură ecran).....	124
Fig. 4.5.	ProM – Secțiunea Acțiuni (captură ecran)	125
Fig. 4.6.	ProM – Secțiunea Observare (captură ecran)	125
Fig. 4.7.	ProM 6 Package Manager (captură ecran)	126
Fig. 4.8.	XESame – interfața inițială de utilizare (captură ecran)	127
Fig. 4.9.	XESame – Secțiunea Mapare (captură ecran)	128
Fig. 4.10.	XESame – Secțiunea Acțiune (captură ecran).....	128
Fig. 4.11.	Disco – capturi ecran ale diferitelor interfețe	129
Fig. 4.12.	RapidMiner (captură ecran).....	132
Fig. 4.13.	RapidProM – extensia de legătură dintre platformele ProM și RapidMiner (captură ecran).....	132
Fig. 5.1.	Metodologia de implementarea a unui proiect BPM în viziunea CertiBPM	135
Fig. 5.2.	Metodologia PDM	138
Fig. 5.3.	Metodologia PDM-Healthcare.....	139
Fig. 5.4.	Sub-procesul „Analiza secvențelor de clustering” din Metodologia PDM-Healthcare	139
Fig. 5.5.	Obiective și acțiuni în cazurile de folosire a Process Mining în organizație	140
Fig. 5.6.	Ciclul de viață L^* pentru managementul proiectelor de Process Mining	141
Fig. 5.7.	Cicluri de viață reprezentative: CRISP-DM și SEMMA	142
Fig. 5.8.	Relația dintre ciclul de viață L^* și cazurile de folosire	142
Fig. 5.9.	Metodologia PM^2 [68]	143
Fig. 5.10.	Metodologia CertiBPM-PM	146

Fig. 5.11. Factorii determinanți care au contribuit la apariția metodologiei CertiBPM-PM	147
Fig. 6.1 Etapa „Obținerea jurnalului” explicată în detaliu (Metodologia CertiBPM-PM).....	150
Fig. 6.2 Etapa „Inspekția jurnalului” explicată în detaliu (Metodologia CertiBPM-PM).....	151
Fig. 6.3 Detalii despre proces din ProM > Log Visualizer > Dashboard (captură ecran)	152
Fig. 6.4 Detalii despre proces din ProM > Log Visualizer > Summary (captură ecran)	153
Fig. 6.5 Investigarea ratei de sosire a cazurilor cu ajutorul XDottedChart	154
Fig. 6.6 Investigarea zonelor cu cazuri puține cu ajutorul XDottedChart	155
Fig. 6.7 Investigarea prelucrărilor activităților în loturi cu ajutorul XDottedChart	155
Fig. 6.8 Investigarea zonelor cu cazuri având durate diferite cu ajutorul XDottedChart	156
Fig. 6.9 Investigarea rolurilor resurselor cu ajutorul XDottedChart.....	157
Fig. 6.10 Investigarea momentelor de începere a lucrului cu ajutorul XDottedChart	157
Fig. 6.11 Investigarea rolurilor grupurilor cu ajutorul XDottedChart	158
Fig. 6.12 Filtrarea jurnalului inițial cu plug-in-ul FLuSH (captură ecran)	159
Fig. 6.13 Distribuția cazurilor în funcție de canalele de aplicare.....	160
Fig. 6.14 Filtrarea jurnalului inițial cu plug-in-ul FLA (captură ecran)	161
Fig. 6.15 Modele create din jurnalul inițial cu algoritmul α (rețea Petri și diagramă BPMN)	162
Fig. 6.16 Modele create din jurnalul filtrat cu algoritmul α (rețea Petri și diagramă BPMN)	163
Fig. 6.17 Modele descoperite din jurnalul filtrat cu algoritmul HeuristicsMiner ...	164
Fig. 6.18 Model descoperit din jurnalul inițial cu algoritmul HeuristicsMiner (ProM 5.2) – diagramă BPMN	165
Fig. 6.19 Model descoperit din jurnalul inițial cu algoritmul Fuzzy Miner	166
Fig. 6.20 Model Fuzzy descoperit din jurnalul inițial/filtrat cu cele mai frecvente activități și arce.....	166
Fig. 6.21 Model creat din jurnalul inițial cu algoritmul Inductive Miner	167
Fig. 6.22 Model creat din jurnalul filtrat cu algoritmul Inductive Miner și conversiile lui.....	168
Fig. 6.23 Devieri de la modelul creat cu algoritmul Inductive Miner	169
Fig. 6.24 Analiza variantelor de proces în urma verificării conformității	170
Fig. 6.25 Model fuzzy reprezentat cu frecvențe și durate medii între activități... ..	172
Fig. 6.26 Frecvența în timp a evenimentelor și a cazurilor active din jurnalul inițial	173
Fig. 6.27 Cele mai frecvente variante de cazuri.....	173
Fig. 6.28 Rețea socială reprezentând transferul de lucru între resurse	175
Fig. 6.29 Rețea socială reprezentând subcontractarea lucrului între resurse	175
Fig. 6.30 Rețea socială reprezentând resursele ce lucrează împreună la aceleași cazuri	176
Fig. 6.31 Modelul fuzzy filtrat ce conține secvența A10→A02	178
Fig. 6.32 Compararea proceselor a căror solicitare este via Internet cu cele de la Ghișeu	179

LISTA DE TABELE

Tabelul 1.1. Partenerii din proiectul CertiBPM	41
Tabelul 1.2. Modulele și elementele de studiu necesare certificării CertiBPM	42
Tabelul 1.3. Distribuția participanților la training-ul CertiBPM pe tip de companie ..	42
Tabelul 2.1. Analiza comparativă a celor patru tehnici de Data Mining	75
Tabelul 3.1. Directive pentru înregistrarea evenimentelor [11]	88
Tabelul 3.2. Matricea amprentei jurnalului de evenimente L	90
Tabelul 3.3. Matricea valorilor succesiunii directe pentru jurnalul de evenimente L	94
Tabelul 3.4. Matricea valorilor de dependență pentru jurnalul de evenimente L	94
Tabelul 3.5. Matricea amprentelor rețelelor Petri $P1$ și $P2$	101
Tabelul 3.6. Matricea amprentei cu diferențele dintre rețelele Petri $P1$ și $P2$	101
Tabelul 3.7. Exemplu de matrice resurse-activități	112
Tabelul 3.8. Exemplu de matrice de transfer al lucrului	112
Tabelul 6.1. Grupuri specializate pe anumite activități	158
Tabelul 6.2. Caracteristicile plug-in-urilor de filtrare (ProM).....	159
Tabelul 6.3. Compararea modelelor obținute cu ajutorul coeficientului de potrivire	169
Tabelul 6.4. Durata medie și frecvența dintre activități pentru identificarea gâtuirilor	171
Tabelul 6.5. Varianta de caz cea mai frecventă	174
Tabelul 6.6. Resursele ce execută secvența de activități $A05 \rightarrow A06$	177
Tabelul 6.7. Resursele ce execută secvența de activități $A10 \rightarrow A02$	178
Tabelul 6.8. Compararea cazurilor a căror solicitare este via Internet cu cele de la Ghișeu	179

INTRODUCERE

a. Motivarea și importanța temei de cercetare alese

Teza de doctorat, „**MANAGEMENTUL PROCESELOR DE AFACERI - PROVOCĂRI ȘI PERFORMANȚE ACTUALE**”, este o lucrare de cercetare interdisciplinară, care se situează ca problematică la intersecția domeniilor de știința calculatoarelor și a managementului, având reverberații în zona sistemelor informatice de management, a managementului performanței și a managementului operațional și strategic, prezenta lucrare apelând la utilizarea, exploatarea și valorizarea unor cunoștințe diverse aferente spațiului complex al ingineriei și managementului. Se impune a fi făcută precizarea că nivelul ridicat de interdisciplinaritate al cercetărilor este determinat de însăși domeniul științei managementului proceselor de afaceri (*Business Process Management, BPM*), acesta aflându-se la rândul său în zona de interferență a mai multor domenii de științe, inginerști și manageriale:

- Managementul strategic, managementul operațional și managementul performanței, prin natura implicațiilor sale la nivelul creșterii competitivității organizaționale;
- Managementul sistemului informatic, prin natura bazei de cunoștințe și de cunoaștere pe care le utilizează exploatează mai eficient la nivel operațional.

Ca urmare, relațiile dintre domeniile de știință precizate, definesc rolul și scopul BPM în cadrul organizațiilor, care au adoptat aceste abordări și practici ale domeniului, obiectivul final fiind acela de a alinia, până la coincidență, strategia de afaceri cu sistemul informatic de management, astfel încât să se obțină efecte sinergice la nivelul managementului performanței în organizații.

BPM reprezintă o abordare care asigură îmbunătățirea, optimizarea continuă a performanței unei organizații. Pentru maximizarea efectelor (în termeni de beneficii) oricărei abordări de tip BPM, orice inițiativă tehnologică asociată implementării și exploatarei acestuia ce implică re-concepția arhitecturii sistemului informatic trebuie să pornească de la identificarea, documentarea și definirea componentei de business, adică a proceselor afacerii. În final, abordarea BPM în organizație se soldează cu implementarea unei soluții de tip sistem de management al proceselor de afaceri (*Business Process Management System, BPMS*), materializată într-un software specializat, care găzduiește modelele proceselor de lucru interne și este vizibil, accesibil pe web. *BPMS* reprezintă componenta sistemului informatic menită să susțină departamentul specializat (denumit cel mai frecvent BPM Office), care are ca obiectiv definirea, modelarea și automatizarea proceselor de lucru interne.

Practica organizațiilor a demonstrat că adoptarea soluțiilor de tip *BPMS* a condus la înregistrarea unor progrese evidente la nivelul reducerii birocrăției, creșterii productivității muncii, dar mai ales în planul creșterii adaptabilității și agilității organizațiilor pe piață și, nu în ultimul rând, au facilitat identificarea unor soluții inovative de satisfacere mai bună a clienților.

Domeniul de știință BPM beneficiază de două viziuni: una a reprezentanților comunității de administrare a afacerilor și cealaltă a comunității software, viziunile fiind diferite. Profesioniștii în administrarea afacerilor se focalizează pe descrierea și

definirea proceselor care să conducă la optimizări în utilizarea resurselor, reducerea costurilor, creșterea productivității, considerând tehnologia informației drept furnizor de mijloace în implementarea și susținerea BPM. Pe de altă parte, profesioniștii din domeniul tehnologiei informației consideră, de cele mai multe ori, că obiectivele și indicatorii de performanță asociați derulării proceselor de afaceri, precum și reglementările organizaționale sunt doar nevoi sau restricții care trebuie considerate în demersul de abstractizare.

Prezenta teză de doctorat susține că trebuie adoptată o înțelegere comună a celor două viziuni (administrarea afacerii și tehnologia informației, software). Totodată, realizarea corectă și robustă a proceselor de afaceri din perspectiva tehnologiei informației sau dezvoltării de aplicații software va conduce la creșterea satisfacției clientului și va contribui la crearea și menținerea avantajului competitiv al unei organizații, rezultat obținut prin colaborarea productivă între cele două categorii de specialiști. Astfel, prin conținutul său și prin rezultatele obținute, teza de doctorat face referire directă la modul în care metodele și mijloacele domeniului de știință BPM pot fi implementate în cadrul organizațiilor, astfel încât să se realizeze o mai bună racordare a acestora la dinamica și tendințele actuale ale mediului economic.

Scopul programului doctoral, derulat și încununat cu finalizarea și susținerea tezei de doctorat, a fost acela de **realizare a unor cercetări teoretice și aplicative pentru caracterizarea modului de implementare a BPM în organizații, în vederea creșterii informațiilor pentru afaceri (Business Intelligence, BI)**. Astfel, cercetările efectuate au condus la investigații pentru identificarea acelor metode și mijloace care pot contribui substanțial la îmbunătățirea calității procesului decizional în afaceri. În acest context, un rol important a revenit cercetărilor privind:

- **Data Mining** sau procesul de extragere, explorare a șabloanelor model dintr-un volum mare de date (baze, depozite sau sisteme de date), prin combinarea metodelor din statistică și inteligență artificială cu cele din managementul bazelor de date, și
- **Process Mining**, demers ce presupune extragerea cunoștințelor din evenimentele înregistrate de sistemul informatic.

Toate acestea au avut rolul de a susține funcția de prognoză și previziune managerială, dar și actul de decizie prin:

- reducerea semnificativă a influenței distructive a jocurilor de putere sau a grupurilor de presiune din cadrul organizațiilor (prin eliminarea argumentelor nefondate ale acestora și considerarea unor fapte și evenimente evidente ale organizației);
- eliminarea entuziasmului decizional nefundamentat, subiectiv și supradimensionat;
- reducerea timpilor morți petrecuți cu activitățile de raportare periodică (colectarea de rapoarte, consolidări și ajustări diverse), prin operaționalizarea modului de realizare a acestora;
- reducerea timpului necesar adoptării unei decizii, deoarece *Data Mining* și *Process Mining* furnizează rapoarte rapid și în formă grafică adecvată, susținând astfel procesul decizional;
- evitarea blocajelor decizionale prin lipsa unor informații sau cunoștințe clare, veridice, științific fondate asupra fenomenelor sau proceselor;
- reducerea rolului departamentului de IT în generarea rapoartelor propriu-zise în favoarea utilizatorului final, concomitent cu reducerea sau eliminarea timpului destinat acestora;
- limitarea supra-exprimării ego-ului (sau a argumentării unor lideri de opinie) în comunitatea managerială.

Rezultatele și concluziile obținute s-au constituit ca fundament pentru crearea unei **metodologii de implementare a unui proiect BPM cu ajutorul tehnicilor specifice Process Mining**. Ca urmare, **obiectivul principal al tezei de doctorat a fost acela de a elabora și valida această metodologie**. Acest demers a fost posibil datorită nivelului de cunoștințe și de cunoaștere dobândit prin programul doctoral (dovadă fiind CV-ul din Anexa 6 a tezei), a experienței profesionale în domeniul BPM și al managementului de la nivel operațional dovedite de doctorand.

Cercetările efectuate în cadrul programului doctoral au fost valorificate prin publicarea unui număr de 12 lucrări științifice (8 articole publicate în străinătate și 4 articole publicate în țară), din care: 6 articole indexate în baza de date ISI Thomson; 2 articole aflate în curs de indexare în baza de date ISI Thomson; 4 articole indexate în alte baze de date cum sunt Springer Link, Elsevier, EconPapers, Ulrich.

Totodată, cercetările realizate sunt racordate la preocupările colectivului de cadre didactice din cadrul Facultății de Management în Producție și Transporturi (Centrul de Cercetare Inginerie și Management) al Universității Politehnica Timișoara. Cercetarea științifică a beneficiat de sprijinul logistic și financiar al proiectului **Leonardo da Vinci, Transfer of Innovation: Certified Business Process Manager – CertiBPM (contr. nr. LLP-LdV/TOI/10/RO/010)**, din cadrul Lifelong Learning Programme, proiect coordonat de d-na prof. dr. ing. Anca DRĂGHICI în perioada 2011-2013.

Prin concluziile formulate și contribuțiile originale asociate cercetărilor, considerăm că prezenta teză de doctorat are un potențial apreciabil de interes și aplicabilitate, respectiv extinderea și generalizarea demersului de cercetare adoptat în cazul organizațiilor.

b. Obiectivele cercetării și modul de atingere a acestora (panoramarea structurii tezei de doctorat)

Obiectivul general al programului doctoral a fost acela de a concepe o metodologie de implementare a proiectelor de BPM, care să permită o mai bună susținere a deciziei manageriale strategice și să confere informații pentru afaceri (inclusiv testarea și validarea sa). Analiza preliminară a modului de atingere a acestui obiectiv s-a realizat prin intermediul modelului „cutiei negre”, după cum este prezentat în Fig. I.1.

După realizarea unui inventar realist al resurselor și rezultatelor așteptate în urma realizării cercetărilor, au fost definite **obiectivele operaționale** derivate și asociate fiecărui capitol al tezei de doctorat:

OP1 Cercetări asupra referențialului bibliografic pentru caracterizarea managementului proceselor de afaceri (*Business Process Management, BPM*) și a unor concepte conexe acestuia, relevante pentru cercetarea de față.

OP2 Cercetări asupra referențialului bibliografic pentru caracterizarea demersului de *Process Mining* și a unor metode și mijloace aferente acestuia (modele ale proceselor de afaceri și *Data Mining*)

OP3 Cercetări teoretice pentru identificarea elementelor cheie care fac parte dintr-un proiect de *Process Mining*.

OP4 Cercetări teoretico-aplicative pentru identificarea și caracterizarea mijloacelor software asociate metodologiei pentru derularea proiectelor de *Process Mining* în organizații.

OP5 Cercetări teoretice pentru conceperea metodologiei *CertiBPM-PM*, propuse pentru derularea proiectelor de *Process Mining* în organizații.

OP6 Cercetări aplicative pentru demonstrarea modului de susținere a deciziei manageriale strategice prin demersul de *Process Mining – Data Mining* în cazul implementărilor de BPM în organizații (testarea și validarea metodologiei *CertiBPM-PM* propuse).

Modul de atingere a obiectivelor, general și operaționale, este demonstrat prin structura logică adoptată a cercetării și reflectată prin succesiunea capitolelor tezei, precum și a conținutului acestora. În Fig. I.2 se prezintă logica adoptată a cercetării (relația capitole – obiective de cercetare).

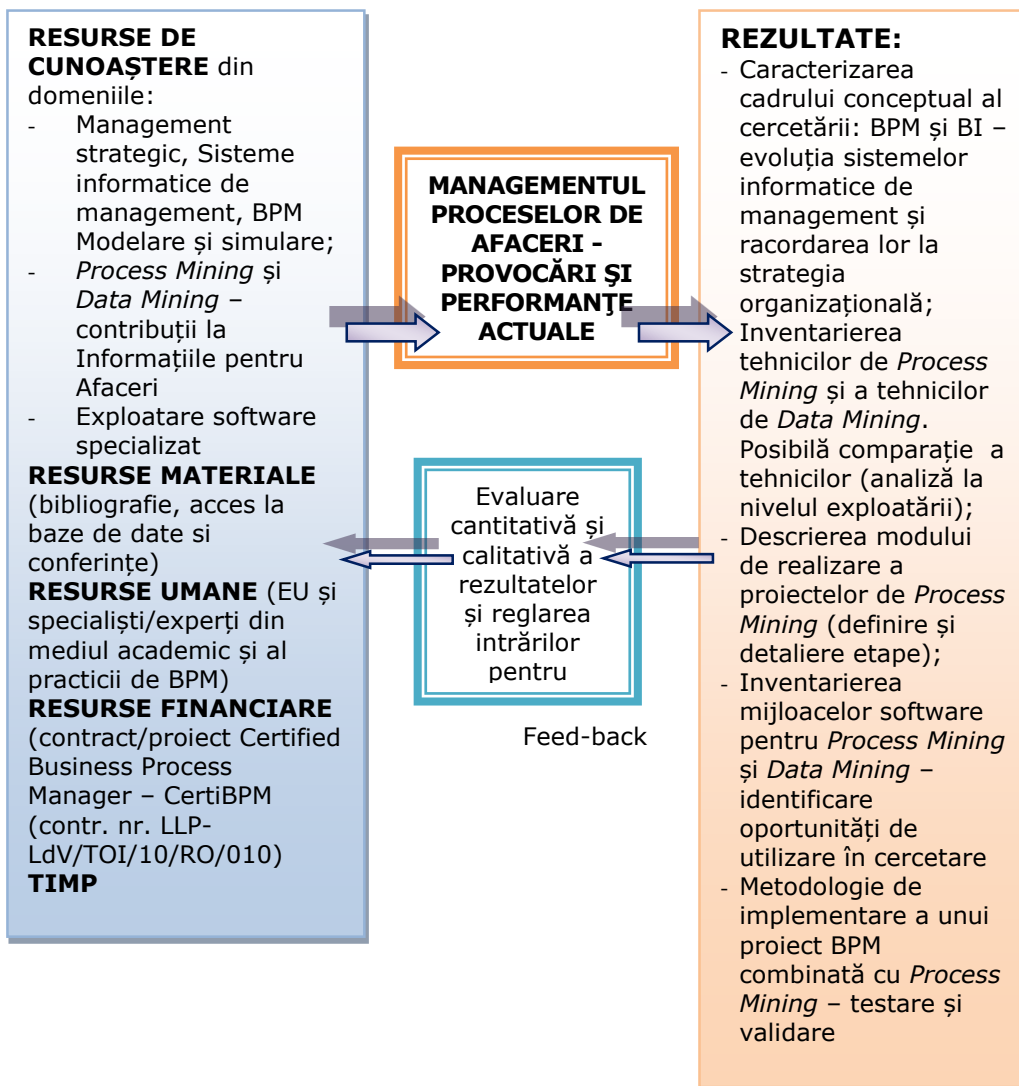


Fig. I.1. Modelul de tip „cutie neagră” asociat cercetării științifice prin doctorat

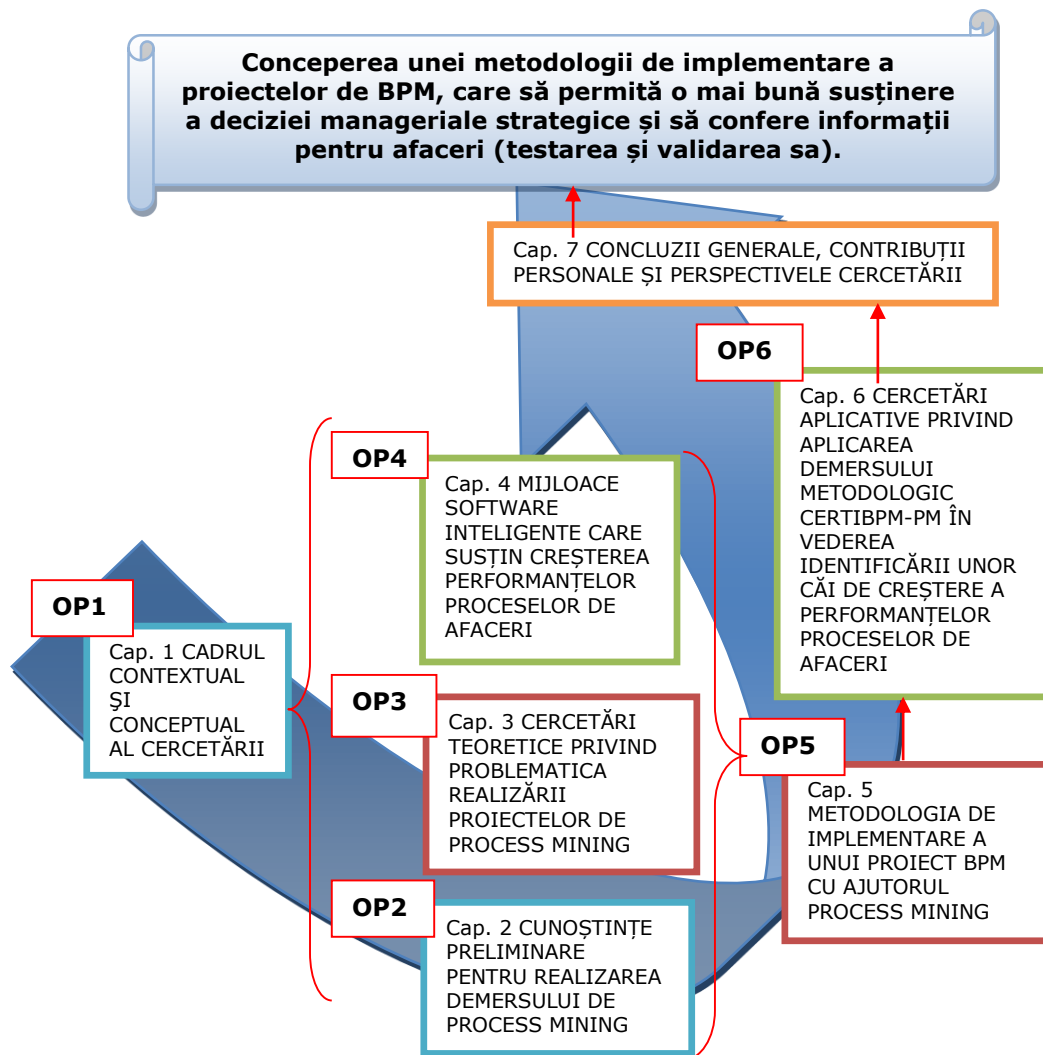


Fig. I.2. Logica adoptată în cercetare aferentă programului doctoral

Tratarea temei este logică, graduală și bine structurată: pornind de la o abordare de tip puzzle, prin care sunt identificate și caracterizate „piese de cunoaștere”, prin descrierea conceptelor și abordărilor în domeniul BPM și BI, în contextul managementului strategic actual (prin sinteze asupra referențialului bibliografic), pentru ca state-of-the-art să constituie baza și motivația cercetărilor teoretice și aplicative în domeniul *Process Mining* (ca demers metodologic asociat modelării proceselor de afaceri), *Data Mining* (prin caracterizarea tehnicilor specifice) care au condus la definirea cadrului procesual detaliat de realizare a proiectelor de *Process Mining* în organizații (din perspectiva managementului acestora și prin detalierea mijloacelor software disponibile în prezent, care pot susține informațiile pentru afaceri).

Toate acestea au condus la conceperea **metodologiei CertiBPM-PM**, care a fost testată și validată prin cercetări aplicative. În final, dezvoltării creative i se succed o serie de dovezi privind impactul pozitiv pe care abordarea BPM combinată cu *Process Mining* o aduce în îmbunătățirea procesului decizional strategic și în ceea ce privește sprijinul derulării, pe baze științifice, a funcțiilor de prognoză și control managerial, în cadrul organizațiilor.

Prin concluziile și contribuțiile originale formulate în finalul cercetării, care au fost incluse în capitolul 7 al tezei, se poate afirma că tema abordată a fost soluționată și toate obiectivele de cercetare propuse au fost atinse.

Teza de doctorat este o lucrare vastă, totalizând **201 pagini (inclusiv 3 anexe)**, demarând cu un capitol introductiv, 6 capitole aferente problematicii de cercetare și un capitol final ce înglobează: concluzii, contribuții personale și direcții viitoare de cercetare. Teza cuprinde **113 de figuri, 20 de tabele, 62 de relații matematice, precum și o listă cu 174 de titluri bibliografice citate** în lucrare. Succesiunea și conținutul fiecărui capitol de lucrare sunt convergente spre atingerea obiectivului operațional asociat. Ca urmare, în Fig. I.3 se prezintă structura tezei de doctorat, care demonstrează modalitatea planificată de atingere a obiectivelor de cercetare propuse.

Primul capitol al tezei, **CADRUL CONTEXTUAL ȘI CONCEPTUAL AL CERCETĂRII (24 pag.)**, cuprinde lămuriri de natură conceptuală referitoare la disciplina BPM. Sunt prezentate și analizate originile acestei discipline, definiții și noțiuni cheie specifice BPM, necesitatea și rolul ei în sistemul de management al organizației. Din studiul referințelor bibliografice a rezultat că disciplina este o îmbinare de abordări din perspectiva controlului calității, din perspectiva managementului și din perspectiva tehnologiei informației. De-a lungul existenței disciplinei, tehnologia informației a avut un rol important, fiind factorul cu influența cea mai mare. Contextul actual al fenomenului „Big Data” [39] influențează BPM și apar noi tendințe de prelucrare inteligente a datelor unei companii, *Data Mining* și *Process Mining*.

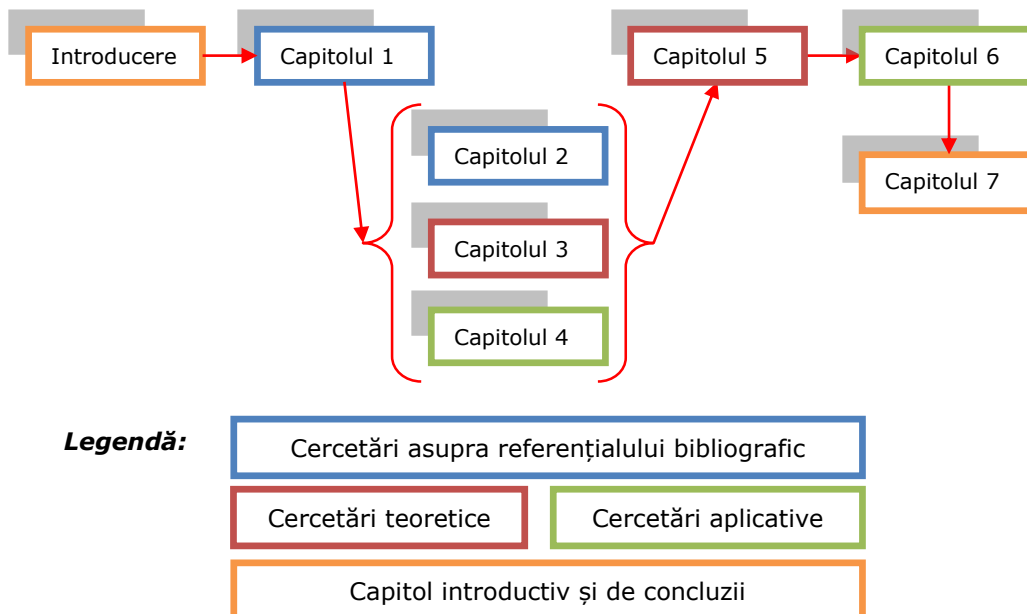


Fig. I.3. Structura tezei de doctorat

Process Mining cuprinde tehnici care permit analiza proceselor de afaceri cu ajutorul datelor existente în jurnalele de evenimente. *Process Mining* este o îmbinare între modelarea de proces și *Data Mining*. Tot în capitolul 1 s-a prezentat proiectul *CertiBPM*, care reprezintă o importantă punte de transfer a cunoștințelor BPM către piața românească. Pe baza analizei referințelor bibliografice s-a concluzionat că *Process Mining* are impactul cel mai puternic din punct de vedere tehnologic asupra BPM și s-a formulat ca obiectiv de cercetare studierea amănunțită a acestui domeniu.

Capitolul 2, CUNOȘTINȚE PRELIMINARE PENTRU REALIZAREA DEMERSULUI DE PROCESS MINING (**31 pag.**) conține elemente teoretice de bază necesare înțelegerii ulterioare a tehnicilor *Process Mining*: modelarea proceselor de afaceri și tehnicile specifice *Data Mining*. S-a întocmit referențialul bibliografic referitor la modelarea proceselor de afaceri, fiind analizate și sintetizate reprezentări de modele de procese, atât modele matematice (rețele Petri, rețele cauzale tip C-nets), cât și modele sub formă de diagrame folosite în industrie (BPMN, YAWL și EPC). Cunoașterea acestor modele este la fel de importantă precum cunoașterea tipurilor de date folosite în programarea calculatoarelor. Referențialul bibliografic referitor la *Data Mining* sintetizează cele mai folosite tehnici, cu accent pe acele tehnici care aduc valoare în *Process Mining*, fiind realizată o comparație a acestora. De asemenea, sunt sintetizate etapele unui proces de *Data Mining* și beneficiile pe care această metodologie le aduce în luarea de decizii în cadrul companiilor.

Capitolul 3, CERCETĂRI TEORETICE PRIVIND PROBLEMATICA REALIZĂRII PROIECTELOR DE PROCESS MINING (**43 pag.**), urmează etapele unui proiect *Process Mining*, iar în fiecare etapă sunt prezentate tehnicile *Process Mining* folosite acesteia. Prima etapă o reprezintă *extragerea datelor* și are ca rezultat crearea unui *jurnal de evenimente*. Sistemele informatice lasă urme (*logs*) în funcție de ce au executat. Aceste urme sunt datele care ne interesează pentru a crea jurnalul de evenimente. Au fost cercetate formatele de jurnale de evenimente, directivele pentru înregistrarea evenimentelor și aspectele referitoare la calitatea datelor extrase. Urmează etapa *descoperirii procesului*, în care au fost prezentați cei mai folosiți algoritmi. Acești algoritmi au ca date de intrare jurnalul de evenimente, iar ca date de ieșire modele de proces, care au fost documentate în capitolul 2. *Descoperirea procesului* reprezintă analiza lui din perspectiva fluxului de control. În etapa următoare are loc *verificarea conformității*, care ajută la conectarea evenimentelor din jurnal cu activitățile aferente modelului, având loc alinieri între jurnalul de evenimente și modelul descoperit. Modelul aliniat va fi extins prin *adăugarea de perspective* (organizaționale, temporale, cazuistice). Toate aceste etape conduc la *obținerea unui model final* care va fi investigat pentru identificarea diferitelor probleme (gâtuiuri, devieri, alocare proastă de resurse ș.a.). Cercetările întreprinse în cadrul acestui capitol prezintă fidel stadiul actual al cunoașterii din domeniul *Process Mining*, realizându-se un referențial bibliografic structurat, ce poate fi folosit pentru o rapidă inițiere în domeniul *Process Mining*. Pentru o mai bună înțelegere a acestui domeniu de către alți cercetători, tehnicile și algoritmi prezentați conțin exemple practice.

Capitolul 4, MIJLOACE SOFTWARE INTELIGENTE CARE SUSȚIN CREȘTEREA PERFORMANȚELOR PROCESOR DE AFACERI (**13 pag.**) conține cercetări teoretice și aplicative privind alegerea mijloacelor software potrivite care pot constitui suportul implementării unui proiect BPM în manieră *Process Mining*. Criteriile după care s-au ales mijloacele software au fost:

- Să fie capabile să ofere suport pentru toate etapele unui proiect de *Process Mining*, prezentate în capitolul 3.

- Să fie accesibile din punct de vedere al costului, astfel fiind alese mijloace software open-source.
- Să se poată interconecta cu soluții software de *Data Mining*
- Să fie ușor de utilizat

Mijloacele software care au fost alese pot constitui un pachet de aplicații software util pentru investigații atât de *Process Mining*, cât și de *Data Mining*.

Capitolul 5, METODOLOGIA DE IMPLEMENTARE A UNUI PROIECT BPM CU AJUTORUL PROCESS MINING (**15 pag.**), prezintă cercetări teoretice privind metodologiile existente pentru implementarea de proiecte BPM și pentru implementarea de proiecte *Process Mining*. Este descrisă metodologia de implementare a unui proiect BPM din cadrul cursului de certificare *CertiBPM*. De asemenea, au fost documentate metodologiile existente de *Process Mining*. **Îmbinarea acestor două tipuri de metodologii a contribuit la crearea unei noi metodologii denumită *CertiBPM-PM (CertiBPM for Process Mining)*, ce prezintă elementul de noutate al prezentei cercetări.** Literatura de specialitate pentru *Process Mining* abundă în foarte multe aspecte teoretice derivate din știința calculatoarelor, dar prezintă puține referințe privind aplicarea acestor tehnici în industrie. În prezent, numărul de metodologii pentru proiecte *Process Mining* este foarte scăzut (au fost identificate doar 4, ultima fiind apărută foarte recent, în luna iunie 2015). **Această nișă de cercetare identificată poate fi exploatată prin propuneri de metodologii novatoare.**

Capitolul 6, CERCETĂRI APLICATIVE PRIVIND APLICAREA DEMERSULUI METODOLOGIC CERTIBPM-PM ÎN VEDEREA IDENTIFICĂRII UNOR CĂI DE CREȘTERE A PERFORMANTELOR PROCESELOR DE AFACERI (**33 pag.**), prezintă validarea metodologiei *CertiBPM-PM* propuse. Etapele specifice *Process Mining*, definite în cadrul metodologiei *CertiBPM-PM*, au fost parcurse succesiv, iar tehnicile prezentate în capitolul 3 (de *descoperire* a modelului, de *verificare a conformității* acestuia, de *analiză organizațională* și de *analiză a performanței*) au fost utilizate. Ca suport software, au fost folosite acele mijloace prezentate detaliat în capitolul 4. Analiza a avut două iterații. În prima iterație investigarea a pornit de la jurnalul de evenimente inițial. A fost nevoie de o a doua iterație, pentru că s-au aplicat filtre pe jurnalul inițial deoarece acesta conținea activități rare care adăugau complexitate procesului. Prin analiza procesului, s-au identificat două gâturi, pentru care au fost identificate posibilele cauze, iar apoi s-au făcut recomandări pentru îmbunătățirea procesului. Metodologia *CertiBPM-PM* a arătat astfel, că modelarea manuală a proceselor se poate elimina din implementarea clasică a unui proiect BPM, această activitate fiind înlocuită cu o metodă de investigare rapidă a proceselor, recurgându-se la folosirea unor mijloace și tehnici specifice *Process Mining*.

Capitolul 7, intitulat CONCLUZII. CONTRIBUȚII PERSONALE. PERSPECTIVE VIITOARE ALE CERCETĂRII (**6 pag.**), prezintă concluziile cercetării bibliografice, teoretice și aplicative realizate și evidențiază aportul original, personal al autorului în domeniul managementului proceselor de afaceri, cu precădere al racordării acestuia la dinamica actuală privind creșterea performanțelor organizațiilor. De asemenea, sunt prezentate direcții de cercetare cu potențial de dezvoltare ulterioară.

În finalul lucrării este prezentată o listă a referințelor bibliografice și trei anexe conținând detalii ale cercetării aplicative realizate.

1. CADRUL CONTEXTUAL ȘI CONCEPTUAL AL CERCETĂRII

Capitolul de față are ca scop prezentarea rezultatelor unei analize și sinteze asupra referențialului bibliografic, în vederea descrierii cadrului contextual și conceptual al tezei de doctorat definit de noțiunile: proces de afaceri, managementul proceselor de afaceri, informații pentru afaceri (*Business Intelligence, BI*).

Obiectivul operațional aferent cercetărilor din acest capitol a fost: **OP1** Cercetări asupra referențialului bibliografic pentru caracterizarea managementului proceselor de afaceri (Business Process Management, BPM) și a unor concepte conexe acestuia, relevante pentru cercetarea de față.

1.1. Managementul proceselor de afaceri (BPM) - istoric și necesitate

Într-o lume a afacerilor aflată într-o continuă schimbare, apare frecvent necesitatea de redefinire a paradigmei de management (fie cele care abordează managementul strategic, fie cele care abordează managementul operațional), precum și nevoia de aliniere a acestor paradigme la tendințele globale din domeniul administrării afacerilor.

În mediul economic actual caracterizat de piețe globalizate, cu clienți mereu informați asupra ultimelor tendințe în inovarea produselor și proceselor (atât tehnologice, cât și de imagine, design) și cu progrese accelerate în domeniul tehnologiei informației și a comunicării (TIC) [106], **nevoia de schimbare** nu este doar un exercițiu de domeniul teoriei managementului, ci se dovedește a fi un demers necesar și obligatoriu în practica organizațiilor. Ca urmare, managementul schimbării sau abordarea sa pozitivă sub forma îmbunătățirii continue a proceselor sunt abordări obligatorii (beneficiind de metodologii, metode și mijloace specifice) în cadrul organizațiilor moderne preocupate de creșterea performanțelor lor. Acest tip de comportament organizațional și-a dovedit efectul pozitiv în susținerea și menținerea avantajului competitiv sustenabil.

Companiile care au definit un model de afaceri bazat pe procese complexe și pe relații funcționale stufoase au dezvoltat și implementat gradual practici de analiză și redefinire (îmbunătățire, re-inginerie etc.) a acestor procese, cu scopul de a eficientiza activitatea de ansamblu a organizației (eficientizare ce a impus definirea și actualizarea explicită a proceselor) [157]. Acest tip de comportament organizațional este rezultatul, nu doar al nevoii de documentare și trasabilitate a proceselor de afaceri, cât mai ales al dorinței de a crea o bază solidă de cunoaștere pentru fundamentarea deciziilor manageriale.

În acest context, practicile și abordările legate de managementul proceselor de afaceri (BPM) au fost generalizate, rezultatul fiind conceperea și dezvoltarea de metode și mijloace noi de management de proces, precum și de metodologii complementare adecvate noilor condiții de operare a companiilor în noul mediu concurențial.

Însă, susținerea unor asemenea metodologii manageriale în cazul companiilor mari sau a unităților de producție industriale nu a mai putut fi posibilă prin folosirea unor instrumente clasice sau tradiționale, ci a fost necesară dezvoltarea de platforme informatice care să centralizeze rapid, fidel și coerent toate aceste informații legate de procesele organizației. La începutul anilor '80, aceste nevoi organizaționale au condus la dezvoltarea conceptului de *Sisteme de Informatice Executiv* – *SIE* (*Executive Information Systems – EIS*) având rolul de suport al procesului decizional de la nivel executiv [71].

Datorită schimbărilor rapide survenite în mediul extern puternic concurențial, nu doar organizațiile au fost nevoite să se schimbe structural, ci și managerii și întreg personalul acestora. Comportamentul organizațional materializat în acțiunea sau reacțiunea promptă pe piețe puternic concurențiale, cu o dinamică accentuată, urmând scenariile diverse (și de multe ori necunoscute ca direcție, sens sau intensitate a fenomenelor) a devenit o caracteristică critică a tuturor actorilor noului mediu de afaceri [85] și care nu mai poate fi neglijată. Aceste fenomene au influențat atât pregătirea noilor strategii, cât și abordarea operațională, de zi cu zi. Mai mult, trecerea la noi paradigme care să fie utile mediului de afaceri, determinând un comportament organizațional adecvat, dezirabil atingerii unor niveluri de competitivitate ridicate, trebuie să se realizeze în mod adaptiv-participativ, conform conceptului *Schimbării Fără Cauză* – *SFC* (*Change Without Cause – CWC*) [45]. Ca urmare, managementul organizațiilor moderne este chemat să uzeze tot mai mult de funcția de prognoză, dar folosind mijloace moderne, operative și eficiente, prin intermediul cărora să poată fi simulate diferite scenarii de orientare a afacerilor care să surprindă concurența (să fii cu un pas înaintea celorlalți), să satisfacă mai bine nevoile viitoare ale clienților, totul în favoarea avantajului competitiv sustenabil.

Astfel, în cazul acestui mediu de afaceri extrem de dinamic și competitiv, lipsa unei schimbări pro active de strategie poate duce chiar la eliminarea organizației de pe piață, chiar dacă aceasta beneficiază de notorietate sau tehnologii de ultimă oră [97]. În prezent, lupta dintre companii are loc în domeniul abordărilor particularizate pentru client, precum și a capacității organizației de reducere a costurilor [172].

După cum se poate observa, pentru a rămâne competitivă, organizația trebuie să se reinventeze, să își îmbunătățească constant procesele prin crearea, asimilarea și implementarea de concepte și metode inovative, novatoare, simultan cu asigurarea suportului provenit din partea tehnologiei informației și comunicării. Astfel, se creează premisele apariției și implementării conceptului de managementul proceselor de afaceri (*Business Process Management, BPM*).

O simplă căutare pe motorul de căutare Google a termenului *Business Process Management*¹ (*BPM*), realizată în luna martie 2015, a returnat 12,9 milioane de rezultate, ceea ce demonstrează că acest concept este destul de cunoscut și exploatat în întreaga lume. De asemenea, interesant este faptul că prima subcategorie a acestui termen, *Business Process Management Software*², a returnat 10,3 milioane de rezultate, ceea ce reprezintă aproximativ 80% din paginile web identificate pentru termenul *BPM*. De aici rezultă că, la nivel internațional, se manifestă un interes foarte mare pentru aspectele relative privind implicațiile tehnologiei informației pentru operaționalizarea *BPM* în cadrul organizațiilor.

¹ Expresia căutării a fost: "*Business Process Management*" (frază exactă). Căutarea după expresia *Business Process Management* returnează 293 milioane rezultate, dar multe pagini nu sunt relevante, pentru aceasta s-a apelat la o frază exactă

² Expresia căutării a fost: "*Business Process Management*" *AND Software*

Pentru ca o companie să rămână competitivă pe piață, trebuie să fie productivă, adică să producă cât mai mult și la un nivel calitativ cât mai ridicat, într-un timp cât mai scurt. Aceste deziderate au atras după sine nevoia de a gestiona un volum mare de date (fenomen cunoscut în prezent ca *Big Data*), fapt care a condus la dezvoltarea și implementarea unor metode și mijloace de realizare operativă.

În acest context au apărut mijloacele de tip sistem informatic, iar tehnicile exploatate cu succes de acestea au la bază cunoașterea și cunoștințele din domeniul științific al managementului proceselor de afaceri. Între sistemele informatice și procesele de afaceri există o relație directă:

- Procesele de afaceri pot fi executate atât manual, cât și automat, cu ajutorul sistemelor informatice [22].
- Procesele se pot gestiona cu ajutorul sistemelor informatice, dând naștere așa-numitelor Sisteme Informatice de Conștientizare a Proceselor – SICP (Process-Aware Information Systems – PAIS). Acestea includ sisteme clasice de Management al Fluxului de Lucru – MFL (Workflow Management – WFM), dar și alte sisteme care asigură o flexibilitate ridicată sau un suport pentru sarcini specifice [8].

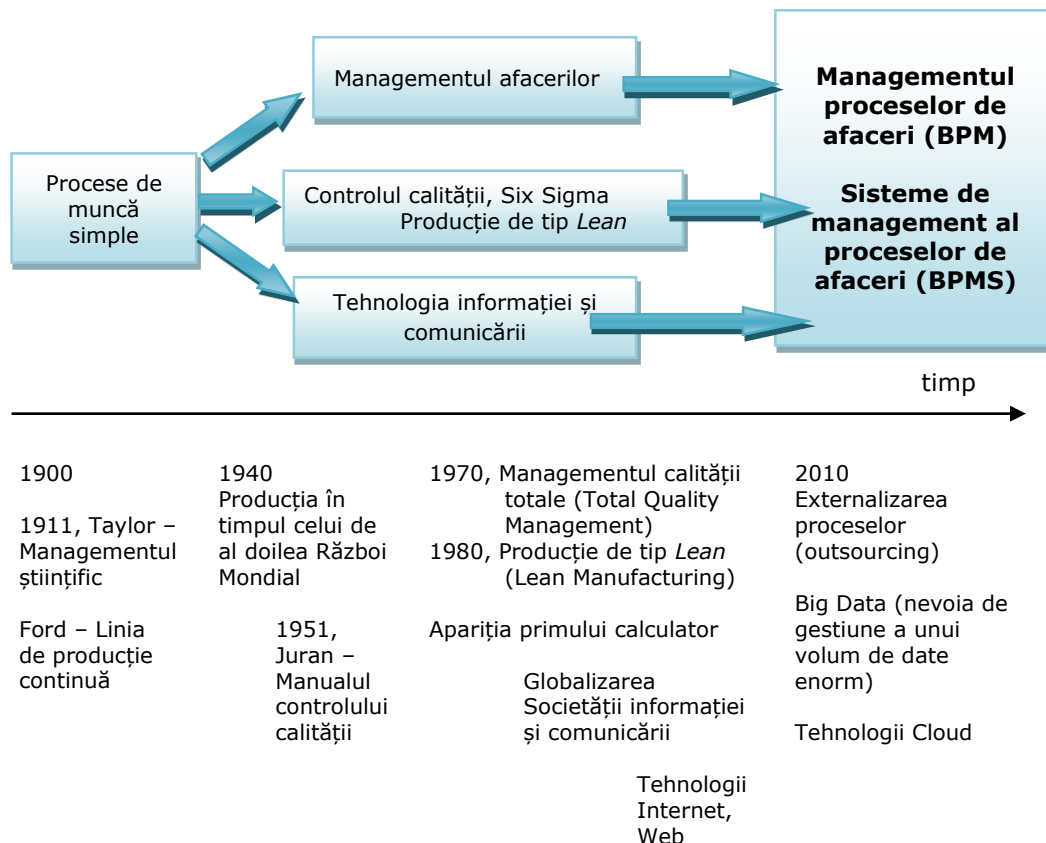


Fig. 1.1. Tradiția BPM - continuare a tradiției manageriale, a controlului calității și a tehnologiei informației

BPM poate fi considerat un termen-umbrelă pentru mai multe abordări și metodologii orientate asupra automatizării, controlului sau operaționalizării proceselor de afaceri. Modul în care știința BPM a apărut și a evoluat a fost marcat de trei abordări majore ale proceselor: perspectiva controlului calității, perspectiva managementului și perspectiva tehnologiei informației [82].

BPM este câteodată un termen confuz, tocmai pentru că înțelegerea lui este direct influențată de către subiectivismul adeptului tradiției, de exemplu un specialist format într-o școală de management va da o importanță mai scăzută controlului calității, sistemului informatic sau tehnologiei informației. De aceea, se impune exemplificarea modului de evoluție al orientării spre/pe procese a organizațiilor, din trei perspective sau tradiții.

a. Perspectiva controlului calității

În 1911, Frederick Winslow Taylor publică celebra carte „*Principiile managementului științific*”, punând bazele controlului calității și ale îmbunătățirii proceselor. Dintre principiile sale enumerăm: simplificarea muncii, experimentarea sistematică pentru identificarea celui mai bun mod de realizare a unei sarcini de muncă, realizarea de sisteme de control ce măsoară modul de realizare a producției, divizarea proceselor de muncă. În contextul acestor baze științifice ale managementului și organizării muncii, operatorii umani/muncitorii au devenit specialiști, din generaliști, aceștia cunoscând și realizând foarte bine doar o mică parte a unui proces, dar necunoscând procesele din aval sau amonte. Astfel, apare pentru prima dată termenul de manager, adică persoana care cunoaște, la un anumit nivel ierarhic, întregul proces [65].

În aceeași perioadă, Henry Ford a conceptualizat dezvoltarea producției de automobile, ca un singur proces; a proiectat și modularizat fiecare activitate a acestui proces, pentru a se asigura că acesta decurge ușor și eficient (în 1913 a început să funcționeze linia de producție, iar în perioada 1908 – 1927, Ford Motor Company a vândut 15 milioane de mașini). Această simplificare a muncii a continuat prin apariția unor noi metode și tehnici în domeniul asigurării calității proceselor.

În 1951, Joseph Juran, evreu de origine română, tipărește prima ediție a renumitei cărți „*Manualul controlului calității*”, astfel stabilindu-se sursa enciclopedică de informații și cunoaștere în mișcarea de control al calității. Este recunoscut ca întemeietor al metodei de control statistic al calității proceselor (*Statistical Process Control*) și a unui program de instruire conex, alături de marele specialist Edwards Deming. Datorită implementării ideologiilor lor în domeniul calității, Japonia devine prima țară (după distrugerile majore suferite în timpul celui de al Doilea Război Mondial) care abordează cu interes major problema controlului calității. Această mișcare a anilor 1950-1960 este cunoscută drept „miracolul japonez”, pentru că industria japoneză a devenit în acest scurt timp a doua mare putere economică a lumii.

Tot din perspectiva controlului calității fac parte și mișcările *Managementului Calității Totale (Total Quality Management, TQM)* apărută în anii '70, mișcarea sau modul de gândire *Lean* (adesea tradus în limba română ca *suplu*), dezvoltată de Toyota la începutul anilor '80 și *6 Sigma (Six Sigma)* dezvoltat de Motorola la sfârșitul anilor '80. TQM este o abordare de management care vizează succesul pe termen lung, prin preocuparea constantă pentru creșterea satisfacției clienților; TQM are ca scop integrarea calității în toate procesele care au loc în organizație. Unele concepte TQM, alături de Re-ingineria Proceselor Afacerii (*Business Process Reengineering, BPR*), sunt transferate și în cadrul abordărilor actuale din domeniul BPM [133]. Six Sigma și Lean fac parte din familiile compatibile de tehnici și mijloace care au ca scop

îmbunătățirea proceselor. Filosofia metodologiei Lean este centrată pe eliminarea pierderilor (termenul japonez folosit în acest caz este 無駄 = muda, acesta având sensul de cheltuieli, rămășițe, resturi, deșeuri), fiind considerat ca un subset de tehnici al BPM [93], iar metodologia Six Sigma își propune îmbunătățirea calității proceselor prin identificarea și eliminarea defectelor, a erorilor sale (prin minimizarea variabilității). Deoarece demersul Six Sigma este considerat un precursor al abordărilor de BPM, multe din tehnicile de îmbunătățire folosite își au originea în demersul Six Sigma [48].

Un alt concept aferent perspectivei controlului calității este Modelul de Maturitate a Capabilității – MMC (*Capability Maturity Model, CMM*), dezvoltat de Software Engineer Institute (SEI) de la Universitatea Carnegie-Mellon. Acest concept apare în anul 1995, la inițiativa Departamentului de Apărare a Statelor Unite, care avea nevoie de o metodă de a evalua furnizorii de soluții software. CMM se focalizează pe procesele unei companii de software, pe care le evaluează și le încadrează în niveluri de maturitate [134]. Pe baza rezultatelor obținute în aplicarea CMM în domeniul dezvoltării de mijloace software, acest model a fost generalizat și extins la nivelul altor industrii, sub forma Integrării Modelelor de Maturitate a Capabilității - IMMC (*Capability Maturity Model Integration, CMMI*) [46].

b. Perspectiva managementului

Această abordare este îmbrățișată mai ales de profesioniștii din domeniul economico-financiar, spre deosebire de perspectiva controlului calității, în care se regăsesc cu preponderență ingineri și specialiști în asigurarea calității. În viziunea perspectivei de management, performanța holistică a organizației devine prioritară, esențială.

Reprezentanți de seamă ai acestei perspective sunt Thomas Gilbert și Geary Rummler, pionieri ai dezvoltării abordărilor tehnologiei pentru performanța umană (*Human performance technology, HPT* sau conceptul similar *human performance improvement, HPI* sau *human performance assessment, HPA*), care considerau organizațiile ca sisteme în cadrul cărora procesele-cheie pentru creșterea performanței se concentrează pe motivarea, formarea și managementul angajaților. Ei susțin că îmbunătățirea performanței managerilor și a angajaților este cheia către îmbunătățirea proceselor [147]. Încă din 1962, membrii acestei mișcări, reuniți în Societatea Internațională pentru Îmbunătățirea Performanței (*International Society for Performance Improvement, ISPI*³), pun accentul pe definirea clară a proceselor de afaceri ale organizației și a rolurilor și activităților operatorilor umani în cadrul acestora, astfel fiind susținută creșterea sustenabilă, cuantificată prin rezultate măsurabile și cu valoare adăugată semnificativă ca parametri pentru performanță.

O altă personalitate a viziunii managementului în definirea BPM este Michael Porter, o autoritate în domeniul managementului strategic și competitivității (companiilor, regiunilor și statelor). Potrivit acestuia, strategia este strâns legată de modul în care companiile își organizează activitățile în cadrul lanțurilor de valoare (în sensul generării de valoare, termenul englez folosit fiind *value-chains*), și care sunt baza creării, dezvoltării avantajului competitiv al unei companii. În esență, lanțurile de valoare sunt procesele decisive prin intermediul cărora o companie generează valoare pentru clienții săi. Toate celelalte procese sunt definite prin raportarea lor la lanțul de valoare. De aici, Porter face pentru prima dată distincție între procesele de bază și procesele de suport ale unei organizații, dar folosește deseori denumirea de activități pentru ceea ce numim în prezent procese. Totodată, Porter susține că

³ <http://www.ispi.org/pl/about/what-is-performance-improvement.pdf>

poziționarea strategică în piață și implementarea aceluși lanț de valoare prin procese clar definite, conduc către obținerea avantajului competitiv, într-un mod unic, particular unei anume organizații [137].

Pornind de la lanțul de valoare ce contribuie la realizarea și susținerea avantajului competitiv, multe companii au implementat acest demers, care, de fapt nu reprezintă decât o **schimbare de paradigmă în organizații, de la viziunea sistemică** (focalizată pe susținerea funcțiilor întreprinderii, așa cum au fost definite de Henri Fayol și ținând seama de evoluția lor) **spre orientarea pe procesele de afaceri** (focalizate pe crearea de valoare, plus-valoare pentru toate categoriile de stakeholderi (părți interesate), pe creșterea continuă a performanței care determină avantajul competitiv sustenabil al organizației).

Metodologia Tabloul de bord Echilibrat (*Balanced Scorecard – BS*) lansată de Kaplan și Norton în 1996, a apărut în contextul inconsistenței indicatorilor financiari pentru caracterizarea performanței, a schimbărilor survenite în metodele și factorii de producție, creșterea ponderii activelor intangibile sau imateriale (pe fondul dezvoltării conceptului de capital intelectual) și a evoluției strategiilor de creare a valorii, de la gestionarea activelor tangibile la strategii bazate pe cunoștințe (deci pe valori intangibile) [96]. Caracteristică metodologiei *Balanced Scorecard* este prezentarea unui mix de rezultate referitor la diferitele aspecte ale performanței, generate în urma unor măsurări financiare și non-financiare comparate cu o valoare „țintă” (indicatori cheie de performanță, *Key Performance Indicators, KPIs*), totul în cadrul unui singur raport concis. Acesta nu este menit să fie un înlocuitor pentru rapoartele financiare sau operaționale tradiționale din cadrul companiilor, ci un rezumat succint care surprinde cele mai relevante informații pentru a susține decizia managerială. Noua viziune asupra performanței organizaționale creează presiune asupra angajaților, care sunt constrânși în realizarea proceselor funcționale, într-un timp foarte scurt și la un nivel de calitate tot mai ridicat, astfel fiind generată valoarea și satisfacția pentru toți clienții.

c. Perspectiva tehnologiei informației

Tehnologia a fost permanent un catalizator pentru îmbunătățirea (creșterea) vitezei și a volumului unei afaceri. Această perspectivă este legată de utilizarea calculatoarelor (a mijloacelor hardware, în general) și a mijloacelor software, pentru a automatiza procesele afacerii. Tehnologia informației a apărut relativ recent, dar a revoluționat complet modul în care operează organizațiile astăzi. Începând cu anii '50, calculatoarele și infrastructurile de comunicare digitale sunt factori dominanți, care au influențat procesele de afaceri și managementul acestora. Câteva exemple sunt elocvente, în acest sens:

- Un caz interesant este cel al companiei Ford, care, după ce a achiziționat Mazda, a realizat că japonezii aveau doar 5 oameni în departamentul de achiziții, pe când departamentul similar de la Ford avea 400 de oameni [76]. Implementarea unui sistem informatic cu o bază de date comună și o abordare pentru schimbarea radicală a procesului de achiziții au condus la îmbunătățirea sa, personalul reducându-se cu 75%.
- Într-un alt articol cu impact mare, Davenport și Short [50] încurajează managerii să urmărească toate procesele atunci când încearcă să îmbunătățească operațional o afacere, în loc să se limiteze la îmbunătățirea unei singure funcții a afacerii. Astfel, se accentuează și rolul tehnologiei informației ca stimulent pentru re proiectarea proceselor de afaceri.

În contextul perspectivei tehnologiei informației, abordarea oferită prin Reingineria Proceselor Afacerii (*Business Process Reengineering, BPR*) a fost lansată ca

fiind o metodologie de schimbare radicală a proceselor, pornind de la premisa că tehnologia informației nu mai este doar un factor suport pentru gestionarea datelor, ci componenta principală de suport pentru toate procesele organizației. Mai târziu, BPR s-a dovedit a avea o abordare relativ greșită din cauza:

- supra-radicalismului aplicat de manageri, care cereau ca procesele să fie schimbate din rădăcină, iar uneori unele procese funcționale au fost eliminate;
- folosirii necorespunzătoare a conceptului, astfel că orice îmbunătățire a proceselor era catalogată și abordată ca un proiect BPR;
- imaturității și a necorespunderii suportului tehnic, întrucât majoritatea proceselor erau definite prin intermediul unor aplicații ale tehnologiei informației, dar care se pot schimba în urma unor eforturi considerabile în direcția dezvoltării software.

Cu toate acestea, este recunoscut faptul că metodologia BPR a fost precursorul apariției BPM. S-a observat din analiza a peste 100 de companii [22] că organizațiile orientate pe procese, acelea ale căror modele de afaceri aveau la bază conceptul *îmbunătățirii continue a proceselor* pentru satisfacerea clienților, au dovedit o performanță totală mult mai mare decât cele care nu aveau această orientare. De asemenea, curentul BPR a avut un mare răsunset în industria IT, influențând apariția sistemelor pentru Planificarea Resurselor Întreprinderii (*Enterprise Resource Planning, ERP*) și a Sistemelor pentru Managementul Fluxului de Lucru (*Workflow Management System, WfMS*).

- Sistemele ERP sunt mai mult decât o bază de date partajată, ele având incorporate numeroase module ce contribuie la operaționalizarea, automatizarea funcțiilor organizaționale (de exemplu, contabilitate, aprovizionare, planificarea producției, logistică etc.).
- WfMS sunt sisteme ce distribuie activitățile între diferiți actori, având ca bază modele de proces, și care susțin eficiența implementării ideii muncii orientate pe proces (*work oriented Processes*).
- Ulterior au apărut serviciile Web care au făcut posibilă interconectarea diferitelor sisteme WfMS, care la rândul lor au evoluat datorită integrării cu alte aplicații de business, astfel devenind mult mai complexe și dând naștere Sistemelor de Management al Proceselor de Afaceri (*Business Process Management Systems, BPMS*).

În sinteză, Fig. 1.2 prezintă cele trei perspective și abordările relevante care au motivat și fundamentat apariția științei BPM.

1.2. Definiții și noțiuni cheie ale abordării BPM

Pentru o înțelegere mai bună a conceptului Managementul Proceselor de afaceri (în limba engleză *Business Process Management, BPM*) este utilă explicarea fiecărui termen al expresiei, asimilabilă în cazul de față unui concept complex, de natură interdisciplinară.

Termenul **afacere** (în limba engleză, *Business*) face referire la indivizi, care interacționează, pentru a executa un set de activități care aduc valoare clienților și un câștig celor care au inițiat, derulat și realizat activitățile (părți interesate). O altă definiție face precizarea că o afacere reprezintă o activitate legală derulată în domeniile agricol, industrial, comerț, financiar etc., cu scopul obținerii de profit [24]. Deci, *termenul afacere este asociat, în mod evident, celui de activitate derulată în vederea obținerii de profit.*

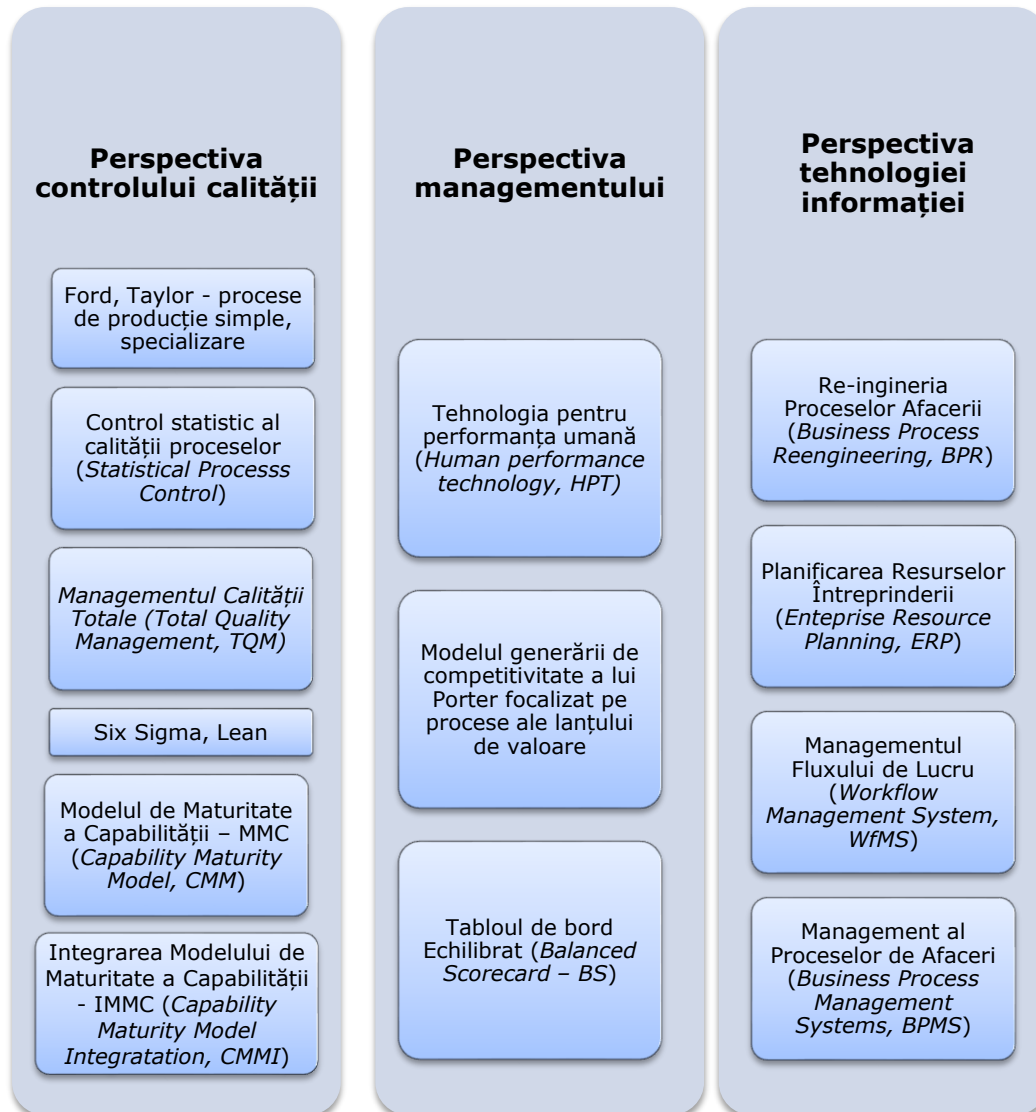


Fig. 1.2. Sinteza perspectivelor și abordărilor relevante care au contribuit la apariția și dezvoltarea BPM

Din punct de vedere terminologic, Dicționarul explicativ al limbii române definește **procesul** ca pe o succesiune de stări, etape, stadii prin care trec, în desfășurarea lor temporală, în schimbarea lor, diverse obiecte, fenomene etc.. Termenul este asociat cu verbele: evoluție, dezvoltare, desfășurare; acțiune [24]. O altă accepțiune este cea prin care procesul este definit ca o succesiune a unor operații, stări sau fenomene prin care se efectuează o lucrare, se produce o transformare (sau are loc o evoluție, dezvoltare, desfășurare; acțiune) [23]. Dicționarul Oxford definește

procesul ca reprezentând o serie de acțiuni sau măsuri luate pentru a realiza un anumit scop [173].

În contextul BPM, procesul este strâns legat de termenul „business” și, deci, se impune explicarea conceptului: proces de afaceri (în limba engleză, *Business Process, BP*), precum și a semnificației sale de-a lungul timpului (evoluția sa).

În anul 1993, Thomas Davenport face afirmația că „un proces de afaceri este pur și simplu un ansamblu structurat de activități menite să producă un anumit rezultat pentru un anumit client sau o anumită piață”. Această definiție asociază procesul de afaceri cu cel de realizare a produsului sau a unui serviciu în cadrul unei organizații de tip industrial. „Un proces este, așadar, o ordonare specifică de activități de muncă, derulate în timp și spațiu, cu un început, un sfârșit, iar intrările și ieșirile (procesului) sunt în mod clar identificate, rezultând un cadru pentru măsurile ce trebuie luate în vederea îmbunătățirii sau optimizării sale” [49]. În același an, Hammer și Champy defineau procesul de afaceri ca „o colecție de activități care preiau una sau mai multe tipuri de intrări și creează o ieșire, astfel contribuind la furnizarea de valoare pentru client” [77]. Johansson definește în mod asemănător procesul de afaceri ca pe un „set de activități înrudite care aplică o transformare pe o intrare, astfel rezultând o ieșire, iar în mod ideal acea transformare ar trebui să adauge valoare intrării și să creeze o ieșire care este mai utilă și mai eficientă beneficiarului fie din amonte (utilizatorul final), fie din aval (furnizorul)” [94].

În anul 1995, Rummler și Brache oferă o altă perspectivă proceselor de afaceri, propunând o primă clasificare a acestora, valabilă și astăzi. „Procesul de afaceri reprezintă o serie de măsuri menite a produce un produs sau serviciu. Cele mai multe procese sunt trans-funcționale, adică se întind pe „spațiul alb” dintre elementele unei organigrame. Unele procese conduc la realizarea unui produs sau serviciu care este destinat unui client extern unei organizații, acestea fiind denumite **proces primare**. Alte procese produc produse care sunt invizibile pentru clientul extern, dar sunt esențiale pentru gestionarea eficientă a activității, acestea purtând denumirea de procese de sprijin” [147].

Mai târziu, în anul 1999, Warboys definește „procesul de afaceri ca o schimbare structurată, adică existența unui **tipar de evenimente succesive**, pe care un observator îl poate recunoaște în diferite exemple concrete (sau apariții) și care poate avea diferite instanțe” [162]. Tot în anul 1999, apare o definiție larg acceptată a procesului de afaceri, elaborată de Coaliția pentru Managementul Fluxurilor de Lucru (Workflow Management Coalition, WfMC) și prin care acesta „este determinat de un set de una sau mai multe proceduri inter-relaționate sau activități, care împreună realizează un obiectiv organizațional, în cadrul unei structuri organizatorice ce are definite roluri funcționale și relații” [170].

Pornind de la definiția lui Davenport, în anul 2003, Smith și Fingar critică lipsa unor aspecte esențiale pentru aplicații practice, cum sunt: coordonarea și definirea vagă a naturii colaborării, a activităților de tip colaborativ și tranzacționale aferente proceselor de afaceri. În viziunea lor „procesul de afaceri este setul complet și dinamic coordonat al activităților de colaborare și tranzacționale care oferă valoare clienților” [152].

În anul 2005, Ould definește procesul de afaceri ca fiind „un set coerent de activități desfășurate de către un ansamblu colaborativ de roluri pentru a atinge un scop, iar gruparea activităților organizaționale în procese trebuie să fie condusă de înțelegerea modului în care organizația operează” [131]. Ould extinde conținutul conceptului prin includerea unor aspecte cheie asociate realizării sale: roluri, actori, acțiuni, suport, interacțiune, obiective și rezultate.

Definiția evoluează, iar în anul 2007 conceptul procesului de afaceri era perceput ca „un set de activități care se efectuează în mod coordonat într-un mediu organizațional și prin intermediul unui suport tehnic adecvat, în vederea atingerii unui obiectiv”. De asemenea, este recunoscut faptul că procesele de afaceri pot exista atât într-o singură organizație, cât și inter-organizațional, situație în care procesele inter-conectează, inter-relaționează organizații [169].

O definiție mai recentă (2013) precizează că: „procesul de afaceri reprezintă o colecție de evenimente inter-relaționate, activități și puncte de decizie ce presupun existența unui număr de actori și obiecte; toate acestea împreună conduc la obținerea unui rezultat, ce aduce valoare cel puțin unui client” [65].

Pornind de la definițiile de mai sus, vom încerca să dăm o definiție proprie a procesului de afaceri, considerată mai adecvată domeniului și temei de cercetare, și care pornește de la premisa potrivit căreia totul în această lume este supus schimbării. Ca urmare, o **definiție a procesului de afaceri trebuie dată în contextul continuei schimbări a sistemului organizațional, determinată de dinamica mediului intern în corelație cu cea a mediului extern.**

Procesul de afaceri poate fi definit printr-un set de activități, evenimente și puncte de decizie succesive și/sau paralele, relaționate între ele, care pot avea o întindere atât intra-organizațională, cât și inter-organizațională, set care aduce beneficii continue tuturor stakeholderilor (părților interesate): actori ce intervin sau interacționează în cadrul realizării procesului de afaceri și clienți/consumatori/utilizatori asupra cărora se răsfrâng rezultatele proceselor.

Practicile actuale în domeniul BPM recunosc că pentru a furniza beneficii continue tuturor stakeholderilor, procesul trebuie adaptat și schimbat continuu, printr-o analiză constantă a rezultatelor (out-put) sale. **Fiecare execuție a procesului reprezintă o instanță care poate urma o cale unică.** Dar realitatea economică a demonstrat că apar *excepții sau abateri de la orice model*, iar dacă frecvența acestora crește, ele vor determina necesitatea declanșării unor schimbări (sunt surse primare de schimbare).

Într-o accepțiune simplistă, *managementul proceselor de afaceri (BPM)* se referă la gestionarea, administrarea proceselor de afaceri, dar literatura de specialitate evidențiază mai multe definiții ce demonstrează complexitatea conceptului:

- „BPM reprezintă disciplina sau știința ce combină cunoștințe din domeniul tehnologiei informației, precum și din științele managementului, pe care le aplică ulterior în cadrul proceselor operaționale ale afacerii” [8].
- „BPM include concepte, metode și tehnici care asigură suportul proceselor de afaceri în faza de concepție (design), administrare, configurare, realizare și analiză” [169].

BPM este o știință care utilizează metode de management, dar într-o altă manieră. Analiza și comparația BPM cu alte domenii ale managementului evidențiază diferențe, așa cum este prezentat în Fig. 1.3.

După panoramarea unor perspective și elemente definitorii ale proceselor de afaceri, se vor trece în revistă conceptele cheie de operare în BPM. Pentru a înțelege modul de realizare a unui proces de afaceri din punct de vedere al științei BPM, este necesară cunoașterea **ciclului de viață** al acestuia. În Fig. 1.4 sunt prezente, și succint descrise, aceste faze: evaluare, analiză și (re)design, configurare și realizare [22].

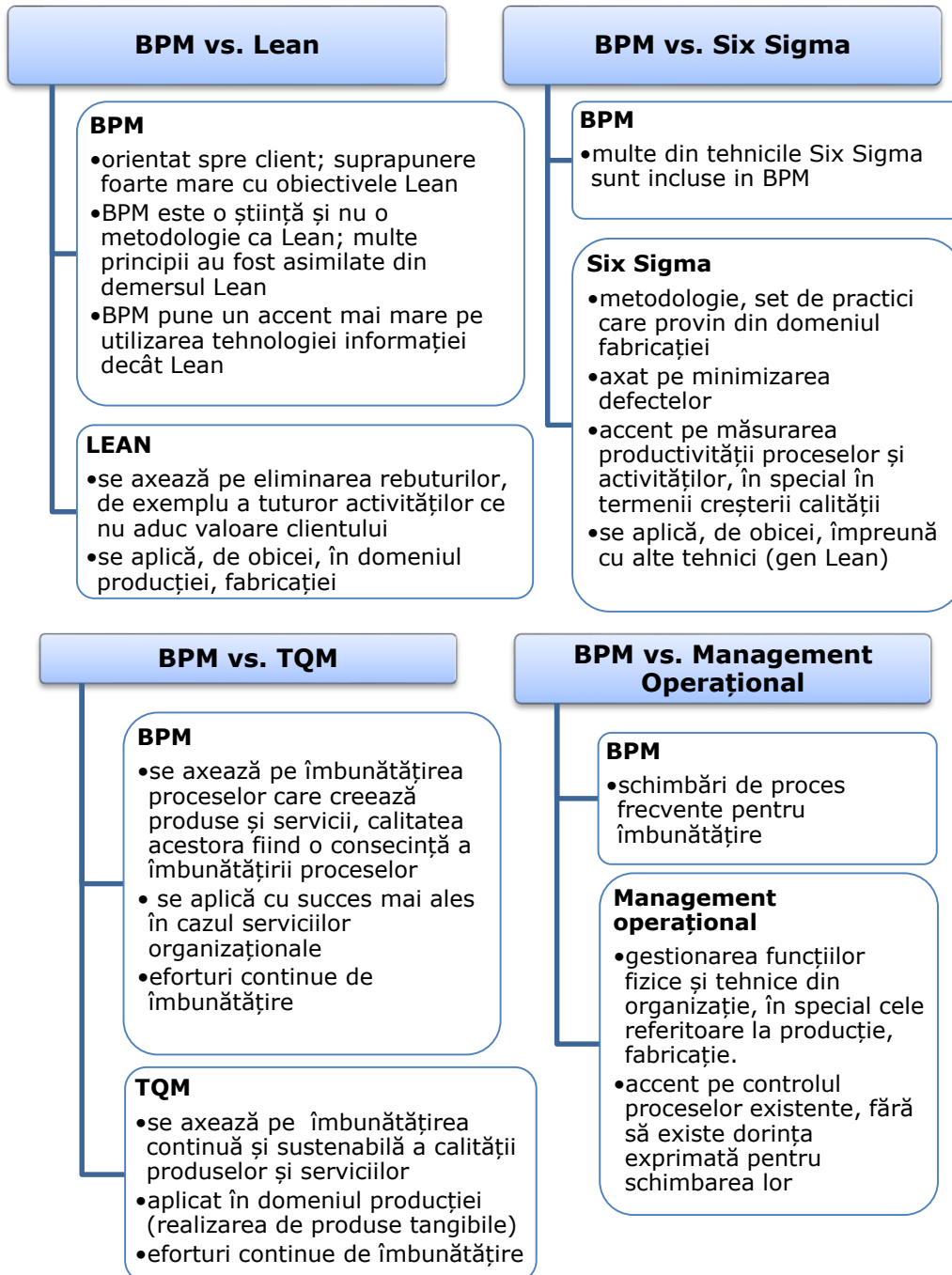


Fig. 1.3. Diferențe fundamentale dintre BPM și alte domenii de management

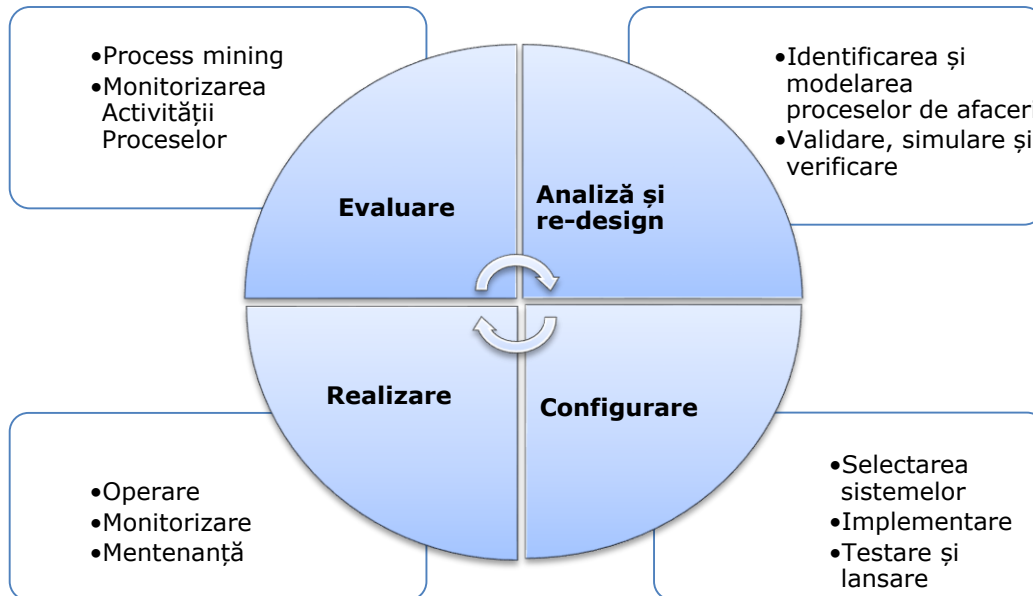


Fig. 1.4. Ciclul de viață al procesului de afaceri în manieră BPM [22]

Procesul de afaceri demarează cu faza de *analiză și (re)design (reconcepție)*, în care au loc anchete (interviuri și observații) privind: modul, contextul organizațional și suportul tehnic de realizare. Pe baza acestor anchete, procesele de afaceri sunt identificate, revizuite și validate, rezultând modele ale acestora. Reprezentarea proceselor de afaceri se face adesea prin modelarea lor grafică, ceea ce facilitează comunicarea și înțelegerea între diferitele părți interesate asupra subiectului BPM. În cazul companiilor mari, având numeroase procese în derulare, sunt necesare modele ale tuturor proceselor pentru un bun management al acestora, dar și pentru a facilita implementarea, realizarea lor printr-o prezentare sugestivă, orientată spre utilizator [91]; modelele de proces permit obținerea unei imagini clare asupra acestora, precum și a relațiilor dintre ele. În acest context, modelarea proceselor de afaceri, ca și metodă de reprezentare grafică a proceselor [27], precum și a distribuției și a relaționării dintre acestea, a devenit o etapă esențială în cadrul metodologiilor de management, dezvoltate la nivel organizațional. **Modelarea Proceselor Afacerii (Business Process Modeling, BPMo) reprezintă o metodă structurată pentru descrierea și analiza oportunităților de îmbunătățire a obiectivelor de business [166], stabilite și urmărite în cadrul organizației.**

După ce modelul procesului a fost identificat, vor urma activități de rafinare prin revizii ale acestuia, cu scopul validării sale. De obicei, reviziile sunt realizate prin intermediul unor workshop-uri (ateliere), în care sunt analizate toate scenariile de utilizare identificate, valide. Scenariile sunt simulate și în funcție de rezultate: se aleg scenariile reprezentative sau se corectează ori se renunță la unele scenarii neconcludente. Din punct de vedere tehnologic, cele mai multe sisteme de BPM dețin medii de simulare, care facilitează analiza proceselor. În cele mai multe cazuri practice, analiza proceselor de afaceri se realizează prin mai multe simulări și iterații, activitatea purtând denumirea de re-design.

Faza de re-design (reconcepție) ar trebui redenumită ca *faza de descoperire a procesului*, pentru că obiectivul acesteia este de a stabili dacă descrierea efectivă a

procesului, realizată în principal de conceptori din domeniul IT, corespunde și este fidelă modului de realizare a procesului în realitatea organizației (având în vedere toate variabilele, certitudinile sau incertitudinile activității sale), ceea ce presupune consultarea unei varietăți mari de „actori” care au anumite interese față de proces: beneficiari, implementatori și inițiatori, manageri, controlori etc. Ca urmare, ceea ce se întâmplă în mod real în cadrul fazei de (re)design (reconcepție) este de fapt o descoperire și detaliere pe care conceptorii din domeniul IT o realizează asupra tuturor activităților și elementelor relevante, de interes pentru un proces dat [65]. Altfel spus, procesul există deja în realitatea organizației, doar că modelul exact al acestuia trebuie descoperit, iar un demers de re-concepție nu se poate aplica în cazul unui *artefact* inexistent.

Următoarea fază a ciclului de viață este *configurarea* (sau *implementarea*) în care modelul procesului de afaceri este transformat în formă executabilă. Această fază poate fi relativ scurtă, dacă sistemul de BPM suportă execuția imediată a procesului [130] sau poate fi mai lungă, dacă este necesară dezvoltarea de aplicații software de către specialiști IT. Există cazuri în care nu există niciun suport informatic pentru realizarea acestei faze și atunci aceasta poate fi realizată printr-un set de politici și proceduri pe care angajații întreprinderii trebuie să se le respecte. Pentru a respecta modelul procesului de afaceri, uneori sunt necesare anumite integrări de sisteme software diferite și de configurare a interfețelor acestora. Un aspect important care stă la baza integrării este tranzacționabilitatea sistemului final, ceea ce presupune realizarea tuturor activităților sau a niciuneia. După ce a fost realizată integrarea tuturor sistemelor software necesare, se verifică și validează anumite activități precum testarea integrării, migrarea datelor și se realizează instruirea utilizatorilor. După cum se poate observa, această etapă este preponderent tehnică.

A treia fază a ciclului de viață este cea a *realizării* sau „punerii în scenă” a procesului și reprezintă momentul în care procesele de afaceri sunt executate și monitorizate. În această fază, există posibilitatea realizării unor ajustări [8], adică în urma rezultatelor de feedback se pot opera modificări (în sensul ajustării) ale datelor de proces, dar nu în ceea ce privește reconfigurarea modelului. În această fază, sistemul de BPM supraveghează execuția instanțelor proceselor afacerii așa cum au fost modelate în fazele anterioare. De asemenea, se realizează monitorizarea, iar sistemul poate afișa (informații de tip vizual) starea proceselor, putându-se interveni asupra acesteia, după caz.

În cadrul fazei de realizare sunt colectate mai multe date, printre cele mai importante fiind jurnalele de execuție, adică marcasele de timp ale activităților. Aceste date sunt utile în analize ulterioare de verificare a conformității și îmbunătățire a proceselor, ele constituind feed-back-ul pentru re-design (reconcepție), în cele mai multe cazuri.

Ultima fază, *evaluarea* (sau *diagnoza*) încheie ciclul de viață al unui proces de afaceri. În cadrul acesteia se stabilesc cerințele de schimbare a procesului, pe baza unui inventar de cauze externe (provenite din mediul extern organizației, precum reglementări legislative noi sau presiuni ale competiției) sau cauze interne (nevoia de creștere a performanței procesului). În funcție de rezultatele evaluării, un nou ciclu de viață poate fi demarat.

Dacă asupra datelor, colectate în faza de realizare, se aplică anumite tehnici precum *Process Mining* (PM, tradus în limba română ca *Mineritul procesului*, dar neutilizat ca atare în literatura noastră!) sau *Monitorizarea Activității Afacerii* (*Business Activity Monitoring, BAM*) [93], se pot obține informații prețioase în legătură cu „starea de sănătate” a proceselor rulate:

- *Process Mining* este o disciplină relativ nouă din domeniul științei calculatoarelor, a tehnologiei informației, și care se situează, din punct de vedere conceptual, între Învățarea Automată (*Machine Learning, ML*) și *Data Mining (DM, tradus în limba română ca Minerit de date, dar neutilizat sub această formă în literatura noastră sau în limbajul de specialitate!)*, pe de o parte și este asociată analizei și modelării proceselor, pe de altă parte [8]. Scopul demersului de *Process Mining* este acela de a descoperi, monitoriza și îmbunătăți procesele reale din cadrul organizațiilor, prin extragerea de cunoștințe din jurnale de evenimente și din alte surse de culegere a datelor, cum sunt sistemele informatice.
- Prin intermediul BAM se pot colecta și examina informații legate de performanța procesului, informații ce reprezintă o condiție esențială pentru implementarea cu succes și evaluarea măsurilor de optimizare continuă a proceselor de afaceri. Astfel, BAM furnizează măsuri de performanță actuale, stringent necesare care pot fi comparate cu obiectivele stabilite.

După cum se poate observa, în primele două faze ale ciclului de viață, rolul principal îl au modelele proceselor, iar în ultimele două faze, rolul principal îl constituie datele. Rolul modelelor în cadrul BPM sau WfM este de a raționaliza procesele (în termeni de re-design) și de a furniza suport în luarea deciziilor, în interiorul proceselor (planificare și control). Modelele proceselor pot fi utilizate pentru: stabilirea responsabilităților, analiza conformității, prognoza (estimarea) performanței folosind simularea și configurarea unui WfMS sau BPMS.

Practica organizațiilor care au implementat sisteme de tip BPM sau WfM a scos în evidență și anumite limite ale analizei bazate pe modelele proceselor:

- Verificarea și analiza performanței se bazează pe disponibilitatea unor modele de înaltă calitate;
- Când modelele și realitatea au foarte puțin în comun, nu este justificată analiza bazată pe modele;
- De obicei, s-a constatat o aliniere deficitară între modelele realizate manual, cu tehnici de modelare simple, și realitate.

Ca urmare, demersul de *Process Mining* poate contribui substanțial la îmbunătățirea analizei bazată pe modele, prin stabilirea unei conexiuni directe între acestea și datele actuale ale evenimentelor unui proces.

1.3. BPM și sistemul de management al organizației

Multe organizații/companii utilizează modelarea proceselor lor ca parte integrantă a politicilor de management. Astfel, se încearcă implementarea celei mai bune structuri pentru a avea un sistem de management al calității (SMC) eficient. În acest context, un sistem de management este un ansamblu de procese și proceduri utilizate pentru a se asigura că o organizație poate îndeplini toate sarcinile necesare pentru atingerea obiectivelor sale. Altfel spus, sistemul de management al calității reprezintă documentul de stabilire sistematică și formală a proceselor, el fiind un punct de plecare pentru o muncă organizată într-o companie.

Demersul de stabilire a unui sistem de management este standardizat. Exemple de standarde sunt: ISO 9001 Managementul calității, ISO 14001 Managementul de mediu, ISO/IEC 27001 Managementul Securității Informației, SA8000 Responsabilitatea Socială.

De ce? În fiecare organizație, atât managementul, cât și forța de muncă în general au tendința de a opune rezistență la schimbare, fie că ea se referă la sisteme, cultura organizațională sau procedurile aferente mediului de lucru. Acesta este un

comportament firesc; oamenii sunt fericiți să continue să facă ceea ce au făcut dintotdeauna. Schimbările aduse de introducerea unui sistem de management al calității (SMC), în special în sectoarele de servicii, nu sunt o excepție. Cu toate acestea companiile care rezistă la schimbări vor deveni mai puțin eficiente și astfel, mai puțin competitive pe piețele în care clienții cer produse și servicii impecabile.

Definirea și implementarea unui sistem de management al calității nu este un demers necunoscut practicii organizațiilor. Scopul acestui demers este de a oferi un sistem pentru dezvoltarea și îmbunătățirea proceselor printr-o abordare structurată, o implementare eficientă și un control mai bun. Răspunsurile la câteva întrebări pot facilita și explica modul de realizare a SMC și pot aduce clarificări privind necesitatea unui astfel de sistem.

„De ce avem nevoie de documentarea unui proces?” Unii ar putea argumenta că „toată lumea știe procesul și a fost instruită cum să-și facă munca, așa că de ce să se documenteze procesul?” Răspunsul rezidă în faptul că, în industriile de servicii, mai ales într-un mediu cu grad ridicat de uzură și rutină cum este cazul externalizării procesului operațional (*Business Process Outsourcing, BPO*), este tot mai imperativă documentarea proceselor. Manualul Calității sau Manualul de Management al Sistemului este documentul cel mai important pentru sistemul de management al calității. El este principalul document realizat cu scopul de a oferi o imagine de ansamblu a sistemului de management al calității implementat. Mulți clienți pot/ar dori să aibă acces la acest document, deoarece ei doresc să se asigure că organizația a avut în vedere toate sistemele critice de management al calității. În cazul în care există reclamații privind produsul, clienții se pot referi la manualul sistemului de management al calității pentru a înțelege sistemele de calitate.

Cine, Când și Ce? Răspunsurile sunt date prin descrierile procesului.

Cum? Răspunsurile ar trebui identificate în instrucțiunile detaliate, precum și în formulare care au menirea de a detalia modul de realizare a proceselor și activităților componente.

„De ce o evidență a înregistrărilor?” Evidența înregistrărilor este una dintre cele mai meticuloase și importante cerințe ale standardul ISO 9001. Meticuloasă, deoarece înregistrările trebuie identificate, îndosariate, protejate și controlate pe parcursul ciclului lor de viață; importantă, deoarece ele conțin istoria modului în care sistemul de management al calității funcționează. Energia, efortul și costul depuse pentru realizarea evidenței înregistrărilor de calitate sunt investiții în curs de desfășurare în vederea construirii unei baze de referință pentru analiză, conformitate și îmbunătățire. Pe baza unor înregistrări de calitate, se poate răspunde la întrebări cum ar fi: „de ce s-a întâmplat acest lucru?”, „când a apărut această problemă prima dată?” și „s-a rezolvat problema?” Cu astfel de informații valoroase la îndemână, înregistrările devin extrem de importante pentru organizație, în preocuparea acesteia pentru îmbunătățire continuă. În plus, înregistrările de calitate sunt o referință primară pentru auditorii interni și externi pentru a evalua conformitatea cu cerințele (evaluarea eficacității sistemului de management al calității). În Fig. 1.5 se prezintă structura cadru a documentației pentru realizarea unui sistem de management.

ISO 10013, Linii directoare pentru dezvoltarea de manuale de calitate, oferă un exemplu de structură a documentației pentru implementarea ISO 9001, Sisteme de management al calității. În timp ce acest document sugerează folosirea unei structuri pe trei niveluri, cele mai multe companii implementează structuri pe patru sau cinci niveluri pentru a-și include înregistrările proprii într-o manieră structurată.

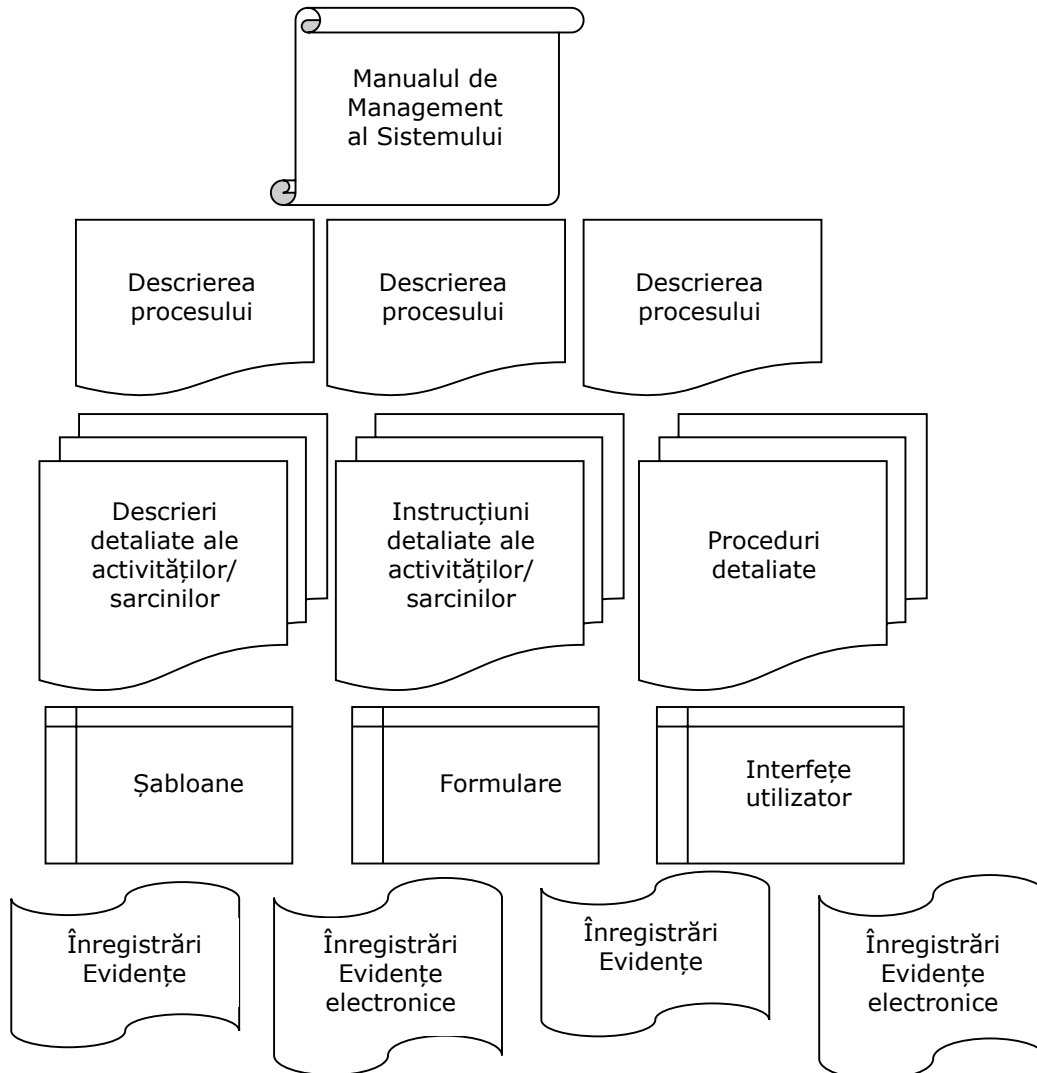


Fig. 1.5. Structura cadru a documentației unui sistem de management

O structură tipică a documentației pe cinci niveluri cuprinde:

- Primul nivel: Manualul Sistemului de Management;
 - Al doilea nivel: descrieri ale procesului;
 - Al treilea nivel: instrucțiuni, proceduri;
 - Al patrulea nivel: șabloane, formulare;
 - Al cincilea nivel: înregistrări (pe suport de hârtie, înregistrări electronice).
- În acest cadru sistematic de lucru apare necesitatea managementului riguros al documentelor organizației ca suport indispensabil demersului.

Manualul Calității împreună cu descrierea procesului pot fi considerate constituția internă și legea cadru a organizației. Pe baza acestora se creează premisele trecerii de la viziunea sistemică (fără ca

aceasta să fie ignorată sau negată în continuare!) a organizației spre viziunea orientată spre proces.

Comparativ cu abordarea sistemică a organizațiilor (având drept referință organigrama acestora, structura pe servicii, departamente, birou etc.), schimbarea de viziune și focalizarea asupra proceselor poate crea animozități, conflicte între structurile funcționale, formale create. În cazul în care diferite activități din cadrul unui proces necesită competențe diferite, aceasta va atrage după sine implicarea unor specialiști din departamente diferite sau chiar a unor departamente diferite din cadrul organizației (și chiar colaborarea dintre mai multe organizații).

Adoptarea viziunii orientată spre proces implică în mod necesar schimbare inter-funcțională și inter-organizațională. În cazul viziunii sistemice, nimeni nu este responsabil de procese, deoarece organizația este împărțită în departamente funcționale sau specializate după unicitate tehnologică (de exemplu, divizia turnate – forjate, divizia prelucrări mecanice etc.) sau de produs (de exemplu, divizii de asamblare produs final) în cadrul cărora activitățile de muncă nu sunt organizate în jurul proceselor, ci în jurul sarcinilor. Pe de altă parte, viziunea orientată spre proces este focalizată pe satisfacerea cât mai bună a clientului și presupune abordarea pe orizontală a afacerii, trans-departamental, ceea ce poate afecta climatul acestora (relațiile de muncă). În cadrul acestei viziuni:

- **Managerii de proces** gestionează sistemele pentru toate procesele, modificările de desfășurare etc.;
- **Proprietarii de proces** sunt pe deplin responsabili pentru derularea efectivă și eficientă a unui proces, de la început până la sfârșit. Aceștia conduc echipa de proces, care poate funcționa în mare măsură pe cont propriu. Prin urmare, rolurile de management trebuie să se schimbe drastic de la planificarea bugetului și control, la orientare și sprijin pentru departamentele operaționale;
- **Proiectanții/designerii de proces** adună informații, modelează și documentează procesul (de multe ori aceștia sunt și proprietarii de proces);
- **Executanții procesului** utilizează descrierile de proces și îndeplinesc sarcinile așa cum sunt descrise. Toți angajații pot fi executanți ai procesului;
- **Analistul de proces** conduce modelarea procesului și este, de obicei, un manager de proces sau un consultant extern.

Modul de aplicare a viziunii orientată pe procese, în cadrul organizațiilor, a condus la delimitări privind tipologia acestora (Fig. 1.6, cu exemplificare în Fig. 1.7):

- **Procesele de management** implică în principal, administrarea, conducerea organizației, menținerea sistemului de management al calității, strategia de dezvoltare etc. Aceste procese sunt efectuate de către conducerea superioară a organizațiilor;
- **Procesele principale** sau de bază sunt procesele ce contribuie direct la crearea valorii produselor și serviciilor. În această categorie se pot nominaliza: dezvoltarea produsului, ingineria produsului, achiziții, producție, logistică etc.

Procesele de sprijin care sunt de tip administrativ și contribuie la performanța proceselor principale, performanța însemnând neîntreruperea producției, produse sau servicii de calitate, personal competent, mediu sigur pentru lucrători, cash-flow sau existența disponibilității de resurse financiare, management informațional etc.

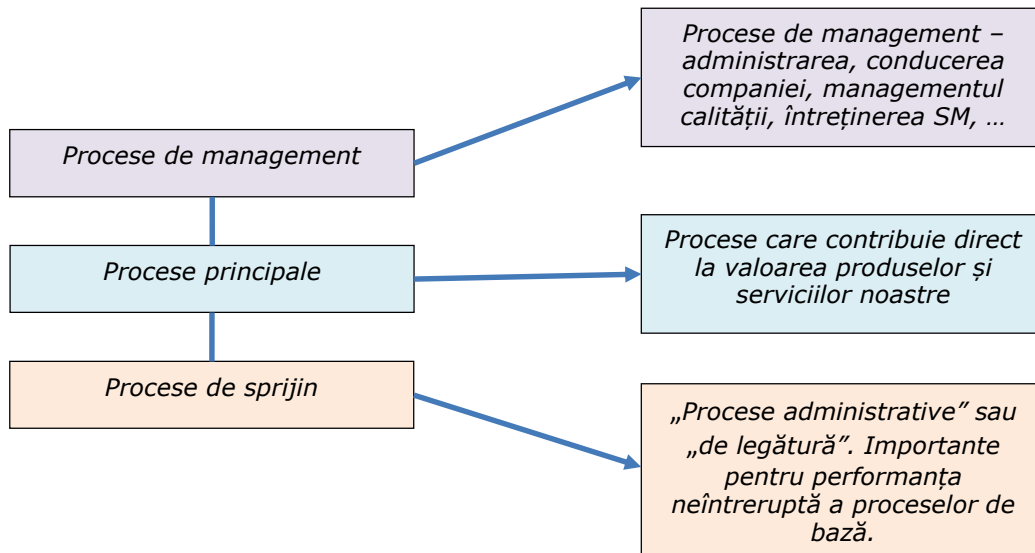


Fig. 1.6. Tipologia proceselor

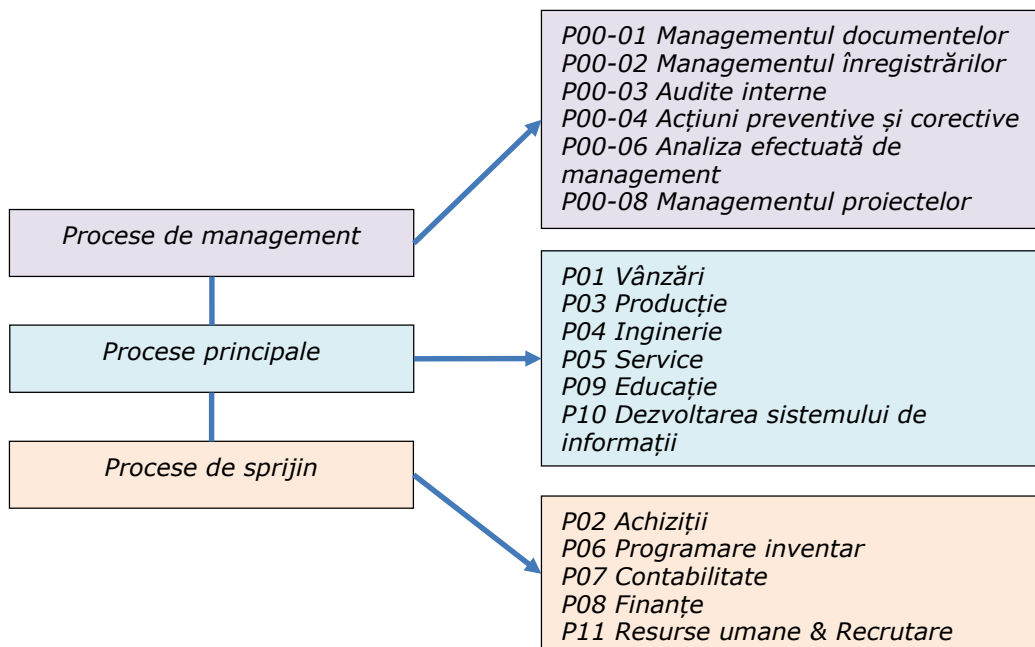


Fig. 1.7. Exemplu privind tipologia proceselor într-o organizație

1.4. Nevoia de certificare BPM pe piața din România

Proiectul *CertiBPM - Certified Business Process Manager* (LLP-LdV/TOI/10/RO/010) s-a desfășurat între anii 2011 și 2013. Motivația proiectului *CertiBPM* o reprezintă satisfacerea nevoilor de training BPM pentru piața românească [59, 64]. Competențele de îmbunătățire a proceselor reprezintă un avantaj pentru angajații companiilor românești, având un impact pozitiv asupra rezultatelor lor. Unul dintre pașii pentru această îmbunătățire este acela de a educa angajații, reprezentanții industriei, viitorii angajați (masteranzi[60] și doctoranzi) în domeniul BPM. Există cursuri BPM, cărți și alte materiale de la diferiți furnizori, care prezintă diferite aspecte BPM, dar nu există niciun curs coerent pentru piața românească. Abordarea propusă este legată de transferul procesului de inovație din Slovenia și Austria către România [63]. Se compune din: transferul de educație, conceptul ECQA de formare profesională și certificare pentru universități și pentru domeniul manufacturier în România [58]. Aspecte ale impactului proiectului sunt:

1. *CertiBPM* se va baza pe cunoștințe și competențe BPM în diferite sectoare industriale, care sunt reprezentate de membrii consorțiului;
2. Transferul conceptului ECQA și a platformei de învățare către România;
3. Punerea în aplicare a trainingului *CertiBPM* prin stabilirea organismelor de examinare pentru certificare și prin stabilirea mijloacelor de evaluare în mediu multilingv;
4. *CertiBPM* și sistemul de certificare european ECQA va aduce un beneficiu industriei manufacturiere din România, în special pentru industria de automobile și de telecomunicații [61, 62].

Subiectele propuse în cadrul certificării *CertiBPM* se potrivesc cu competențele necesare pentru a deveni un specialist BPM (nivel de bază și nivel avansat). Toate materialele de instruire au fost furnizate în mai multe limbi (engleză, slovenă, română și germană) și au fost încărcate pe un sistem de e-learning. Un set de întrebări de testare au fost definite, care oferă baza pentru procesul de certificare a cursanților. În Tabelul 1.1 sunt prezentați partenerii implicați în proiectul *CertiBPM*.

Programul inițial de instruire și materialele au fost dezvoltate interactiv și creativ de către membrii proiectului (în timpul întâlnirilor de proiect față-în-față sau virtual folosind facilități de conferință Skype, în decursul perioadei decembrie 2010 - decembrie 2011). Lucrul în spațiul virtual a generat interes printre membrii proiectului prin încercarea de îmbunătățire a procesului în manieră BPM [127, 128]. Materialele de formare versiunea 2012 au fost dezvoltate în cadrul consorțiului internațional al proiectului. Modulele de studiu *CertiBPM* [59, 129] sunt prezentate în Tabelul 1.2.

Tabelul 1.1. Partenerii din proiectul *CertiBPM*

Nr.	Nume	Rol
1	Universitatea Politehnica Timișoara, Romania	Aplicant (coordonator)
2	BICERO Business Informatics Center Rozman. Maribor, Slovenia	Partener
3	proHUMAN Cooperation and Business Management Ltd. Maribor, Slovenia	Partener
4	International Software Consulting Network Ltd. Graz, Austria	Partener
5	Denkstatt, Romania	Partener

Tabelul 1.2. Modulele și elementele de studiu necesare certificării CertiBPM

Module de studiu și codul elementelor	Module de studiu și titlul elementelor	Criteriu de performanță
BPM.U0	Pachet informativ	5
BPM.U0.E0	Introducere	5
BPM.U0.E1	ECQA și informații pentru certificare (demonstrație pe www.ecqa.org și pe Portalul de Învățare)	2
BPM.U0.E2	Concluzii și referințe	1
BPM.U1	Managementul orientat pe proces	24
BPM.U1.E1	Sistemul de management (ISO 9001:2008)	5
BPM.U1.E2	Managementul proiectelor BPM	5
BPM.U1.E3	BPM și Modelare	5
BPM.U1.E4	Documentarea proceselor de afaceri	4
BPM.U1.E5	Simulare și analiză fundamentală de proces, mijloace și tehnici	5
BPM.U2	BPM și Tehnologia Informației	16
BPM.U2.E1	Alegerea mijloacelor/platformei BPM	5
BPM.U2.E2	Mijloace/platforma BPM	7
BPM.U2.E3	BPM și arhitectura întreprinderii	2
BPM.U2.E4	Sisteme BPM și integrare IT	2
BPM.U3	BPM – aspecte umane. Cadre de lucru și standarde	12
BPM.U3.E1	Factori umani în BPM	4
BPM.U3.E2	Motivarea oamenilor pentru schimbarea procesului	3
BPM.U3.E3	Modele BPM, cadre de lucru și standarde	5
BPM.U4	Specializări BPM	-
BPM.U4.E1	Manager BP pentru procese IT	-
BPM.U4.E2	Manager BP pentru procese de vânzări și marketing	-

Modulele și elementele lor au fost atent descrise pentru a fi consistente în relație cu criteriul de performanță impus de ECQA. ECQA este organismul care eliberează certificatul în urma absolvirii și de aceea criteriile lor sunt foarte importante, mai ales că certificatul are acoperire europeană.

Au fost susținute training-uri CertiBPM în România în regiunea Vest și în București. În total au fost 120 de participanți de la 48 de companii. În Tabelul 1.3 se poate vedea structura celor care au participat la training [62].

Tabelul 1.3. Distribuția participanților la training-ul CertiBPM pe tip de companie

Tipul companiei din eșantion	Număr companii / %	Număr cursanți / %
Automotive	8 / 16,66%	31 / 25,83%
Logistică (transport și distribuție)	7 / 14,58%	12 / 10%
Dezvoltare software	5 / 10,42%	18 / 15%
Telecomunicații	5 / 10,42%	18 / 15%
Producție	8 / 16,66%	15 / 12,5%
Universități (masteranzi și doctoranzi)	3 / 6,25%	10 / 8,33%
Alte industrii (servicii financiare, bănci, turism, administrație publică)	12 / 25%	16 / 13,33%
<i>Total</i>	<i>48 / 100%</i>	<i>120 / 100%</i>

După cum se observă din această statistică, există o orientare puternică către procese, mai ales a celor care provin din medii ingineresti: automotivă și TIC (*Tehnologia informației și a comunicațiilor*), medii dependente de mijloace software. La fel cum s-a arătat în secțiunea 1.1, chiar și în piața locală, cei care caută specializări BPM sunt specialiști care lucrează cu sisteme informatice.

1.5. Domeniul Process Mining în contextul BPM

În cadrul proiectului *CertiBPM* am avut ocazia să întâlnesc metodologia BPM. *CertiBPM* este o certificare utilă pentru a obține cunoștințe de gestionare a unui proiect BPM, cu accent preponderent pe partea de îmbunătățirea de proces prin modelare. Modelarea procesului are două etape. Începe cu identificarea procesului existent (*procesul AS-IS*), și, pornind de la acesta, cu ajutorul limbajului BPMN și al mijloacelor software de modelare, se creează modelul dorit (*procesul TO-BE*). Pentru a identifica modelul de proces existent este nevoie de o procedură stufoasă cu multe interviuri și work-shop-uri cu oamenii implicați. S-a fixat ca obiectiv al cercetării îmbunătățirea de proces prin modelare.

Tehnologia informației are un impact major asupra dezvoltării BPM. Am arătat aceasta atât prin căutarea de pe Google, detaliată în 1.1, cât și din faptul că pe piața internă specialiștii din domeniile tehnice sunt cei mai interesați de metodologia BPM. Deci, există atât o cerere, cât și o ofertă pentru nevoia îmbunătățirii proceselor prin prisma noilor tehnologii software. Această nevoie reprezintă punctul de plecare în cercetarea acestei lucrări.

Tehnologia informației în momentul de față se confruntă cu un fenomen de creștere exponențială a datelor, numit *Big Data* [39], și apare nevoia procesării acestui mare volum de date. În cazul companiilor mari, manipularea informației a dobândit un rol esențial în structurile de management strategice, trecând astfel dincolo de utilizarea imediată a datelor și informațiilor de la nivelul operativ. Informațiile interpretate au devenit astfel suportul esențial pentru planificarea de succes a activităților (mai ales în cazul adaptării organizației la noile situații impuse de mediul extern organizațional) [158].

Setul de tehnici și de mijloace pentru transformarea de date în cunoștințe și informații utile în analiza afacerii sunt grupate sub termenul umbrelă *Business Intelligence (BI)* [29]. O componentă importantă a Business Intelligence este *Data Mining (DM)*. Domeniul *Data Mining* se ocupă cu extragerea tiparelor descoperite din datele brute. Aceste tipare identificate reprezintă cunoștințe folositoare în conducerea afacerii. Apare întrebarea dacă este posibilă îmbunătățirea proceselor în contextul *Data Mining*. Răspunsul la această întrebare l-am identificat în domeniul *Process Mining - PM* (Fig. 1.8).

Van der Aalst [8] definește *Process Mining* ca o disciplină de cercetare în curs de dezvoltare care se află între inteligența artificială și *Data Mining* pe de o parte, și modelarea și analiza proceselor pe de altă parte. *Process Mining* oferă o punte de legătură între *Data Mining* (incluzând inteligența artificială, învățare automată, și descoperirea de cunoștințe) și managementul proceselor orientat către modele (modelarea proceselor, managementul proceselor de afaceri, simulare și verificare).

Domeniul *Process Mining* există în domeniul academic de aproximativ 15 ani. Conștientizarea domeniului de *Process Mining* de către industrie, a început doar în ultimii 2 ani.

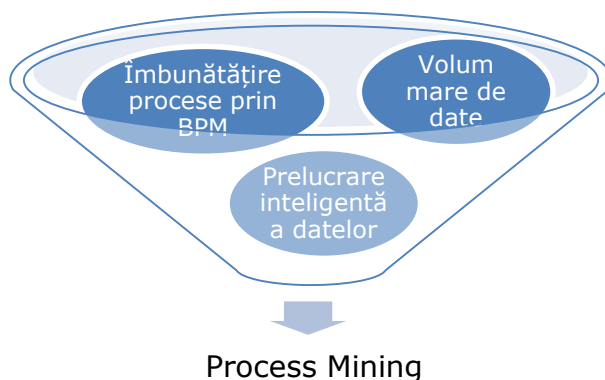


Fig. 1.8. Contextul alegerii Process Mining ca soluție pentru îmbunătățirea proceselor de afaceri

Tehnicile pentru managementul proceselor orientat către modele se concentrează pe procesele derulate integral (end-to-end), dar care nu folosesc date de evenimente, iar modelele sunt făcute de mână și nu sunt corelate cu comportamentul actual observat [22]. Tehnicile de *Process Mining* aduc valoare managementului proceselor orientat către modele, diminuând aceste dezavantaje (eliminând subiectivismul, omisiunile sau alte tipuri de erori umane în cazul modelării manuale). Acesta constă într-un set de tehnici și metode care oferă informații despre procesul analizat.

Cea mai influentă conferință a producătorilor de software BPM și anume BPM|Next desemnează în 2013 drept câștigătoare compania Fluxicon⁴ cu software-ul Disco, un reprezentant software de seamă pentru categoria *Process Mining*. Motivarea câștigătorilor este simplă: „Cele mai multe organizații au procese complexe, care sunt invizibile, astfel greu de gestionat sau de îmbunătățit. Fiecare dintre părțile interesate vede doar o parte a procesului. Descoperirea manuală a procesului prin intermediul workshop-urilor, interviurilor, și revizuirea documentației existente este costisitoare și consumatoare de timp, și rareori reflectă complexitatea reală a procesului. Proces Mining închide acest decalaj făcând procesul real vizibil. Software-ul nostru pentru *Process Mining*, Disco, exploatează la maximum datele IT existente pentru a genera o imagine completă, exactă a procesului, cu perspectivă de acționare imediată. Disco analizează în mod automat fluxurile reale de proces, evidențiază blocajele, arată toate variantele, și permite *reluarea* animată a fluxului de proces, toate făcute interactiv, pornind de la întrebări pertinente asupra procesului”.

În Fig. 1.9 sunt reprezentate categoriile de mijloace software specifice BPM, culese de pe site-ul BPTrends⁵, ce aparține lui Paul Harmon, unul dintre cei mai importanți vizionari BPM. Ceea ce este interesant este că raportul „*The State of Business Process Management*” nu include categoria *Mijloace Process Mining* în versiunea 2012 [83] și martie 2014 [84]. Categoria *Mijloace Process Mining* este recunoscută oficial în mai 2014 de către Paul Harmon [81].

⁴ <http://www.bpmnext.com/bpmnext-2013-presentations/process-mining-discovering-process-maps-from-data/>

⁵ <http://www.bptrends.com>

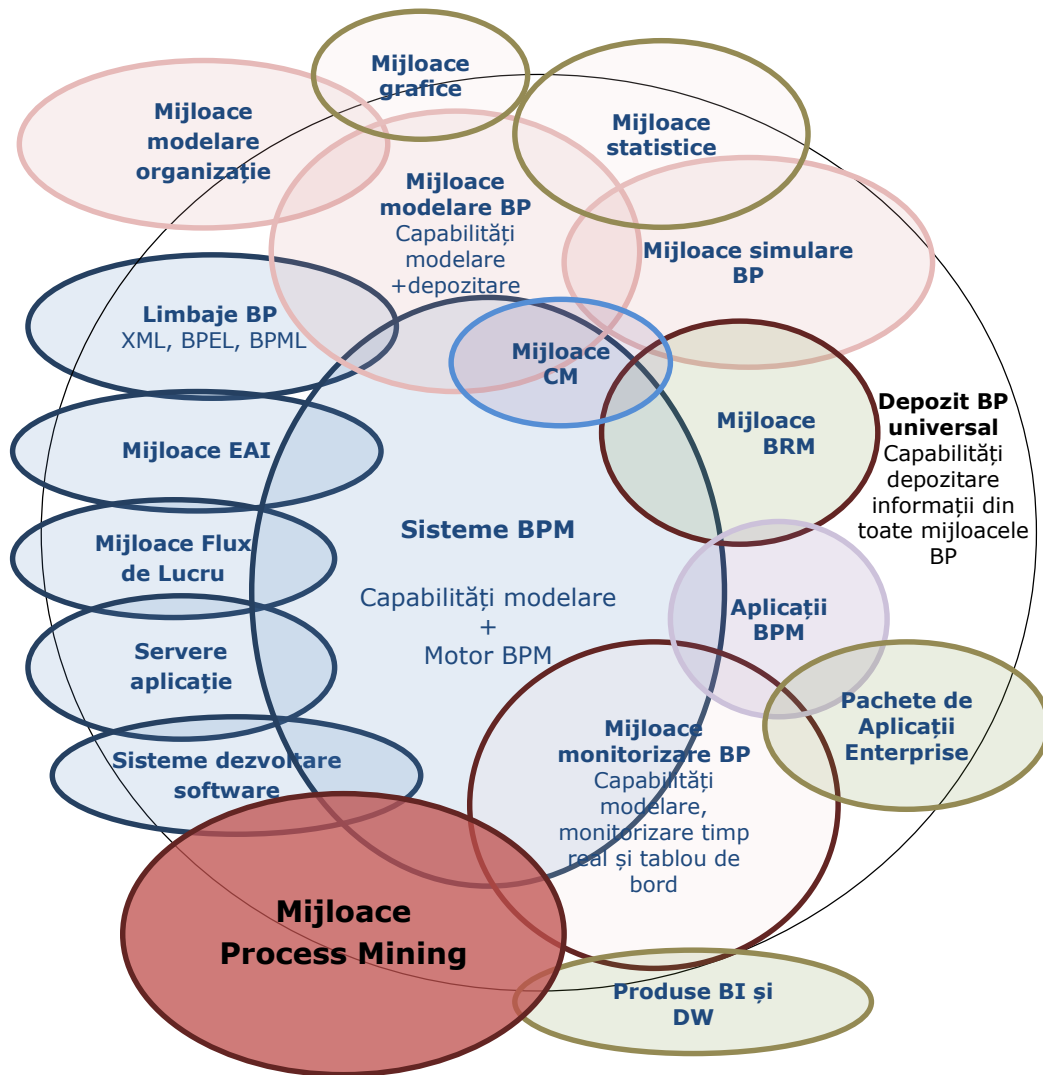


Fig. 1.9. Piața de software BPM conform BPTrends [81]

1.6. Concluzii

În urma cercetărilor teoretice asupra referențialului bibliografic au fost enunțate următoarele concluzii:

- Termenul BPM pune în evidență următoarele:
 - BPM este un termen umbrelă pentru mai multe abordări și metodologii orientate asupra automatizării, controlului sau operaționalizării proceselor de afaceri.

- „BPM reprezintă disciplina sau știința care combină cunoștințe din domeniul tehnologiei informației, precum și din științele managementului, pe care le aplică ulterior în cadrul proceselor operaționale ale afacerii” [8].
- BPM este o continuare a tradiției manageriale, a controlului calității și a tehnologiei informației.
- Modelarea proceselor ocupă un loc central în cazul BPM.
- BPM facilitează introducerea unui sistem de management al calității (SMC) oferind metode de documentare a proceselor.
- BPM conferă o bună organizare unei companii prin identificarea și clasificarea proceselor.
- BPM este strâns legat de tehnologia informației, în special de mijloacele software.
- Datorită abundenței mari de date (*Big Data*), disciplina BPM trebuie regândită. Mijloacele software BPM trebuie regândite prin prisma prelucrării inteligente a datelor. În ajutorul BPM încep să apară tehnici specifice *Business Intelligence: Data Mining și Process Mining*. Acestea sunt tehnici care sunt capabile să gestioneze volume mari de date prin extragerea informațiilor relevante folosite pentru conducerea afacerilor.
- BPMS sunt sisteme software integrate care acoperă multe dintre nevoile unei companii din punct de vedere al BPM. BPMS conțin funcții specifice ciclului de viață BPM.

România are un deficit de cunoștințe în domeniul BPM, existând doar câteva cursuri BPM. Proiectul *CertiBPM* a apărut pentru a răspunde nevoilor de cursuri BPM de pe piața locală, iar în urma cursurilor susținute s-au făcut statistici referitoare la tipul de participanți, identificându-se că specialiștii din companiile ingineresti (automotive, telecomunicații și dezvoltare software) sunt cei mai interesați de disciplina BPM.

Metodologia proiectului *CertiBPM* nu are elemente de actualitate referitoare la informațiile pentru afaceri (*Business Intelligence*), în special *Process Mining*, astfel încât acestea reprezintă punctul de plecare al cercetării realizate în cadrul programului doctoral.

2. CUNOȘTINȚE PRELIMINARE PENTRU REALIZAREA DEMERSULUI DE PROCESS MINING

Prezentul capitol își propune panoramarea cunoștințelor relevante pentru realizarea demersului de *Process Mining (PM)* prin inventarierea, analiza și sinteza referențialului bibliografic. De interes pentru cercetarea prezentă sunt: analiza și prezentarea unor reprezentări (modalități de prezentare) ale modelelor de proces și demersul de *Data Mining (DM)*.

Obiectivul operațional aferent cercetărilor din acest capitol a fost: **OP2** Cercetări asupra referențialului bibliografic pentru caracterizarea demersului de *Process Mining* și a unor metode și mijloace aferente acestuia (modele ale proceselor de afaceri și *Data Mining*)

2.1. Modelarea proceselor de afaceri

Modelarea reprezintă o metodă utilizată în știință și tehnică constând în reproducerea schematică a unui obiect sau sistem sub forma unui sistem similar sau analog în scopul studierii proprietăților și transformărilor sistemului original [114]. Modelul reprezintă un sistem teoretic sau material cu ajutorul căruia pot fi studiate indirect proprietățile și transformările altui sistem, mai complex, cu care primul sistem prezintă o analogie [24].

Un model de proces reprezintă o hartă a acestuia. Similar domeniului cartografiei, unde există diferite tipuri de hărți (geografice, topografice, geofizice, geologice, climatice etc.), un model de proces poate fi reprezentat în diferite moduri, în funcție de notația aleasă. În literatura de specialitate există o mare varietate de notații utilizate pentru a descrie un flux de control [6]: rețele Petri, diagrame BPMN, diagrame de activități UML, Event-driven Process Chain (Lanțul Procesului condus de Evenimente, EPC), rețele cauzale tip C-nets, sisteme de tranziții, lanțuri Markov, modele fuzzy, modele YAWL, modele declarative, algebra de proces, diagrame de flux, modele BPEL ș.a. Apariția unui număr relativ mare de notații a fost determinată de dorința utilizatorilor de a contracara dezavantajele unei soluții în favoarea alteia, dar pentru anumite cazuri concrete, ceea ce a condus la generarea de alte dezavantaje.

Notațiile folosite influențează spațiul căutărilor (realizarea unui model care să surprindă cât mai bine realitatea) și vizualizarea (ceea ce utilizatorii finali vor să fie ușor identificabil și vizibil). Totodată, descoperirea unui proces este dificilă deoarece nu există exemple negative (un jurnal de evenimente, sau inventarul cronologic al unor evenimente, arată doar ce s-a întâmplat și nu arată ce nu se poate întâmpla); jurnalul de evenimente conține doar o parte din comportamentele posibile, datorită concurenței, structurilor repetitive și a deciziilor ce influențează spațiul căutărilor, ceea ce contribuie la dificultatea demersului de descoperire a unui proces, susținut fiind de faptul că nu există o relație clară între mărimea unui model și comportamentul său.

Considerăm oportună prezentare diferitelor modalități de realizare a modelelor de proces, folosind sisteme de notații diferite.

2.1.1. Rețele Petri

Rețelele Petri reprezintă o metodă formală (matematică) folosită pentru modelarea și verificarea sistemelor (concurrente/distribuite). Bazele rețelelor Petri au fost puse de Carl Adam Petri, în anul 1962. O rețea Petri este un graf bipartit ce oferă o reprezentare explicită a stărilor și evenimentelor dintr-un sistem. Rețelele Petri au o reprezentare grafică intuitivă și o semantică formală, ele având avantajul că sunt executabile, fapt datorită căruia sunt folosite în multe aplicații industriale de automatizare [5, 135]. Literatura de specialitate abundă în prezentarea de tehnici de analiză a rețelelor Petri [9, 52, 92].

Definiție (Rețea Petri): O rețea Petri este un triplet $N = (P, T, F)$ unde:

P reprezintă o mulțime finită de *locații* (place),

T reprezintă o mulțime finită de *tranziții* (transition) cu proprietatea

$$P \cap T = \emptyset, \quad (2.1)$$

F - numită *relația de flux* - cu proprietatea:

$$F \subseteq (P \times T) \cup (T \times P), \quad (2.2)$$

și care reprezintă multi-mulțimea arcelor orientate ce leagă locațiile de tranziții și tranzițiile de locații [57].

Definiție (Marcaj): Fie $N = (P, T, F)$ o rețea Petri. Un marcaj al lui N este o funcție:

$$M : P \rightarrow \mathbb{N} \quad (2.3)$$

Definiție (Rețea Petri marcată): Fie $N = (P, T, F)$ o rețea Petri și un marcaj $M : P \rightarrow \mathbb{N}$, unde $M \in \mathcal{B}(P)$, o multi-mulțime peste P . Atunci (N, M) se numește rețea Petri marcată.

Distribuția jetoanelor (punctelor) în locațiile unei rețele reprezintă marcajul rețelei (starea sistemului modelat). În Fig. 2.1, se observă că locația [Start] conține un jeton (punct).

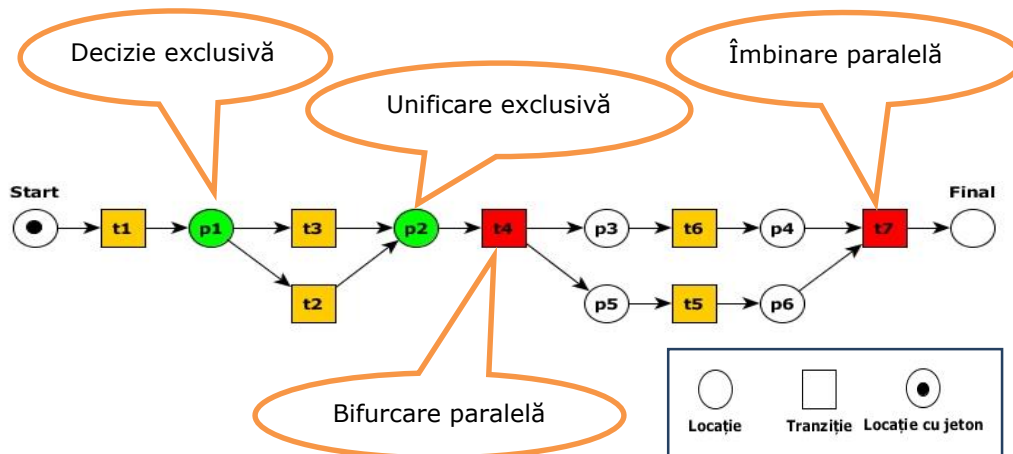


Fig. 2.1. Rețea Petri și elemente de topologie specifice unui proces

O **multi-mulțime** este un concept matematic care reprezintă o extindere a conceptului de mulțime în care fiecare element al mulțimii poate apărea mai mult de o singură dată [156]. Multi-mulțimea mai poartă denumirea de sac (în engleză „bag”), de aceea s-a notat cu \mathcal{B} . Așadar relația:

$$M \in \mathcal{B}(P) \quad (2.4)$$

arată că o rețea Petri marcată poate avea mai multe marcaje.

Elementele mulțimii $P \cup T$ se numesc noduri. Pentru un nod $x \in P \cup T$, premulțimea lui x reprezintă:

$$\bullet x = \{y \mid (y, x) \in F\} \quad (2.5)$$

iar, post-mulțimea lui x reprezintă:

$$x \bullet = \{y \mid (x, y) \in F\}, \quad (2.6)$$

unde

$$F \subseteq (P \times T) \cup (T \times P) \quad (2.7)$$

este relația de flux. De exemplu pentru Fig. 2.1: $\bullet p2 = \{t2, t3\}$, iar $p2 \bullet = \{t4\}$.

Definiție (tranziție validă): Fie (N, M) o rețea Petri marcată. Tranziția $t \in T$ este validă dacă și numai dacă:

$$\bullet t \leq M, \quad (2.8)$$

și se notează cu $(N, M)[t]$, iar cu $\lfloor _ \rfloor$ se notează *regula de declanșare* (din engleză: *firing rule*).

Pentru a observa trecerea dintr-o locație în alta, se folosește conceptul de jeton. Într-o locație pot exista mai multe jetoane. Marcajul rețelei definește starea rețelei la un moment dat și reprezintă un vector ce indică numărul de jetoane aflate în anumite locații. Evoluția sistemului corespunde evoluției marcajelor și are loc prin execuția tranzițiilor, deci o rețea Petri are un comportament dinamic. Acest comportament este dat de regula de declanșare. O tranziție este validă (declanșabilă), dacă toate locațiile sale de intrare conțin câte un jeton. De exemplu, în Fig. 2.1, tranziția $t1$ este declanșabilă deoarece locația de intrare conține un jeton, deci poate fi declanșată. Tranziția $t2$ nu poate fi executată deoarece locația $p1$, care reprezintă locația de intrare nu conține jeton, deci tranziția $t2$ nu este validă.

Execuția unei tranziții presupune următoarele: se consumă (retrage) câte un jeton din fiecare locație de intrare și se produce (depune) câte un jeton în fiecare locație de ieșire.

În contextul *Process Mining*, activitățile corespunzătoare tranzițiilor trebuie etichetate, de aceea apare necesitatea unei rețele Petri etichetate.

Definiție (Rețea Petri etichetată): O rețea Petri etichetată este un tuplu:

$$N = (P, T, F, A, l) \quad (2.9)$$

unde (P, T, F) , fiind definite mai sus, $A \in \mathcal{A}$ reprezintă mulțimea etichetelor de activități, și

$$l \in T \rightarrow A \quad (2.10)$$

este o funcție de etichetare a tranzițiilor. Practic o Rețea Petri etichetată este o rețea Petri clasică căreia îi este atașată o funcție de etichetare pentru tranziții.

Definiție (marcajul accesibil): Fie (N, M_0) o rețea Petri marcată. Se spune despre un marcaj M că este accesibil din marcajul inițial M_0 dacă și numai dacă există o secvență finită de tranziții σ astfel încât:

$$M_0[\sigma \rangle M \quad (2.11)$$

Definiție (graful de accesibilitate): Fie

$$(N, M_0) \text{ cu } N = (P, T, F, A, l) \quad (2.12)$$

cu o rețea Petri etichetată. (N, M_0) definește un sistem de tranziție

$$TS = (S, A', T') \text{ cu } S = [N, M_0], \quad (2.13)$$

$S^{start} = \{M_0\}$, $A' = A$, și

$$T' = \{(M, l(t), M') \in S \times A \times S \mid \exists t \in T (N, M)[t \rangle (N, M')\} \quad (2.14)$$

Deseori se spune că *TS* reprezintă graful de accesibilitate al lui (N, M_0) .

În Fig. 2.2 este reprezentat graful de accesibilitate (fără etichete) rezultat din rețeaua Petri din Fig. 2.1.

Există cazuri în care graful de accesibilitate nu se poate genera, deoarece rețeaua Petri are capacități nelimitate ale locațiilor (de exemplu o tranziție este permanent declanșată). În acest sens, se folosește graful de acoperire care e reprezentat mai jos în Fig. 2.3, graf în care nodurile reprezintă vectori de marcaje posibile. În cazul în care o tranziție este permanent declanșată, locația care are un număr infinit de jetoane este reprezentată prin ∞ în vectorul de marcaj, de exemplu, vectorul de marcaj este $(0,0,0,0,0,0, \infty)$ [124], dacă în nodul final ar fi o buclă către el însuși.

În cadrul *Process Mining*, rețelele Petri reprezintă un model matematic foarte bun ce se poate valida ușor prin jocul cu jetoane. Cu ajutorul rețelelor Petri, se pot valida metodele play-in, play-out și replay [8] ce vor fi descrise în secțiunea 3.1.2.

În continuare se vor prezenta câteva proprietăți ale rețelelor Petri exploatate în *Process Mining*:

- *k*-mărginire: o locație este *k*-mărginită dacă nu există niciun marcaj accesibil cu mai mult de *k* jetoane în *p*. O rețea Petri marcată (N, M_0) este *k*-mărginită, dacă orice locație nu conține mai mult de *k* jetoane pentru orice locație $p \in P$ și orice marcaj:

$$M \in [N, M_0]: M(p) \leq k \quad (2.15)$$

- *Siguranța (1-mărginire)*: o locație sau rețea Petri este sigură dacă și numai dacă este 1-mărginită.

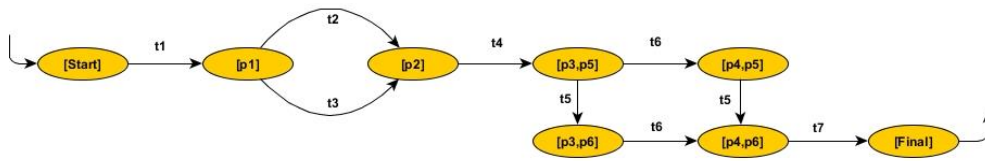


Fig. 2.2. Graful de accesibilitate dezvoltat din rețeaua Petri din Fig. 2.1

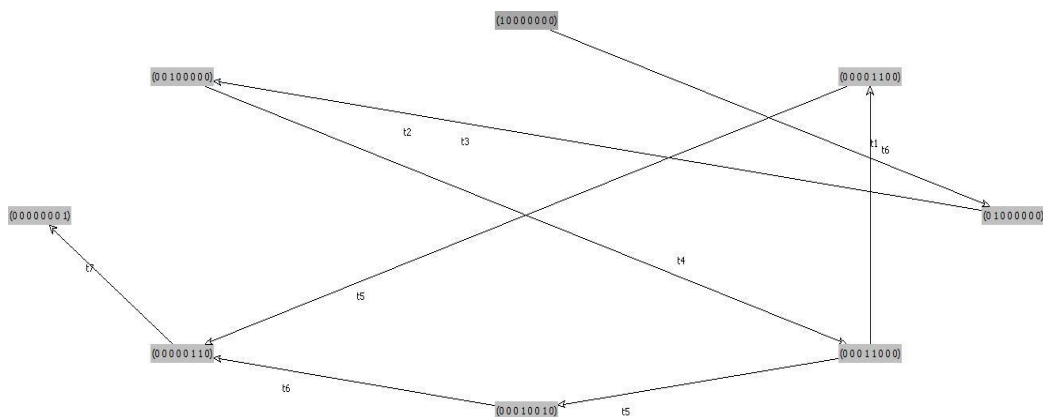


Fig. 2.3. Graful de acoperire dezvoltat din rețeaua Petri din Fig. 2.1

- *Mărginire*: o rețea Petri marcată este mărginită dacă și numai dacă există $k \in \mathbb{N}$ astfel încât rețeaua Petri este k -mărginită.
- *Lipsa blocajelor*: o rețea Petri marcată este fără blocaje, dacă nu există marcaje accesibile moarte. Un marcaj este mort dacă nu există nicio tranziție $t \in T$ astfel încât $M[t]$, adică nicio tranziție nu se mai poate declanșa. O rețea Petri are o locație posibilă moartă dacă are un marcaj accesibil mort. O rețea Petri este fără locații moarte dacă fiecare marcaj accesibil prezintă cel puțin o tranziție.
- *Viabilitatea*: o tranziție $t \in T$ a unei rețele Petri marcate (N, M_0) este viabilă dacă din orice marcaj accesibil este posibil ca tranziția t să aibă loc. O rețea Petri este viabilă, dacă toate tranzițiile ei sunt viabile. Reciproca nu este întotdeauna adevărată.
- *Traseul complet*: reprezintă proprietatea unei rețele Petri de a ajunge într-o locație finală pornind de la o locație inițială.

Rețelele Petri Colorate (RPC – Colored Petri Nets) sunt o extensie a rețelelor Petri clasice [92]. Jetoanele în rețelele Petri Colorate conțin date și au o amprentă de timp, ceea ce înseamnă că jetonul nu se poate consuma decât începând cu data din amprenta de timp. Rețelele Petri se numesc colorate din cauza acestor jetoane ce conțin date, adică au „culoare”; datele se referă la proprietățile acestor jetoane; culorile fac aceste jetoane să aparțină unor clase, ceea ce subliniază unicitatea unui caz, în cadrul unui proces. Rețele Petri sunt un bun suport pentru modelarea proceselor de afaceri [20].

2.1.2. Rețele Flux de Lucru

Pentru modelarea proceselor afacerii se poate folosi o subclasă de rețele Petri denumită rețele Flux de Lucru (*Workflow Nets, WF-net*), adesea utilizate în contextul WfMS și BPMS. Astfel, o rețea Flux de Lucru [5] este o rețea Petri ce are în plus o locație sursă și o locație de final. Locația sursă reprezintă punctul de start al procesului, iar locația finală, reprezintă punctul de final al procesului.

Definiție (Rețea Flux de Lucru): Fie $N = (P, T, F, A, l)$ o rețea Petri etichetată și \bar{t} un identificator ce nu este din $P \cup T$. N este o rețea Flux de Lucru, dacă și numai dacă P conține o locație de start i , astfel încât $\bullet i = \emptyset$, dacă P conține o locație finală o , astfel încât $o \bullet = \emptyset$ și dacă:

$$\bar{N} = (P, T \cup \{\bar{t}\}, F \cup \{(o, \bar{t}), (\bar{t}, i)\}, A \cup \{\tau\}, l \cup \{(\bar{t}, \tau)\}) \quad (2.16)$$

este strâns conectat, adică există o cale directă între orice pereche de noduri ale lui \bar{N} .

Prin rețelele Flux de Lucru, modelele de proces folosite în contextul BPM descriu ciclul de viață al unor cazuri de un anumit tip. Modelul de proces este instanțiat de fiecare dată pentru fiecare caz. Exemple de cazuri ar fi: comenzi online, facturare ș.a. Rețelele Flux de Lucru sunt o reprezentare naturală pentru *Process Mining*. Nu orice rețea Flux de Lucru reprezintă un proces corect. De exemplu, un proces reprezentat de o rețea Flux de Lucru poate prezenta erori, cum ar fi blocaje, activități care nu pot deveni active sau anumite „deșeurii” de jetoane rămase în proces după încetarea acestuia. Prin urmare, în cadrul acestor rețele sunt definite criteriile de corectitudine [15, 126].

Definiție (Corectitudine, din engleză *Soundness*): Fie $N = (P, T, F, A, l)$ o rețea Flux de Lucru cu o locație de start i și o locație finală o . N este corectă (din engleză, *sound*) dacă și numai dacă:

- Este îndeplinită condiția de *siguranță* $(N, [i])$, adică locațiile sale nu pot avea jetoane multiple în același timp;

- Există o *finalizare corespunzătoare*: pentru orice marcaj

$$M \in [N, [i]], o \in M \text{ implică } M = [o] \quad (2.17)$$
- Există *opțiunea de a finaliza* pentru orice marcaj:

$$M \in [N, [i]], [o] \in [N, M] \quad (2.18)$$
- *Lipsa părților moarte*, ce presupune că $(N, [i])$ nu conține tranziții moarte, adică pentru orice $t \in T$, există o secvență de executare ce validează pe t .

O rețea Flux de lucru este corectă dacă și numai dacă rețeaua Petri corespunzătoare „scurt-circuitată” este viabilă și mărginită. Corectitudinea proceselor de afaceri este o problemă intens studiată, anumite abordări fiind prezente în literatură [47, 107, 150].

2.1.3. Rețele cauzale tip C-nets

Rețelele cauzale tip C-nets (Causal Nets – C-Nets), sau rețele-C, au fost introduse pentru prima dată în 2011 [8] și sunt o reprezentare concepută special pentru *Process Mining*, ce are avantajul prezenței elementelor de conectivitate dintre activități. O rețea-C este un graf în care nodurile sunt reprezentate de activități, iar arcele reprezintă dependențele de cauzalitate. O activitate poate să aibă mai multe legături de intrare și mai multe legături de ieșire. Rețelele-C se bazează pe grafurile de dependență, nu sunt executabile semantic, dar schițează posibilitățile de urmat. Se folosesc rețele-C pentru că sunt rezultatul câtorva tehnici de Mining (de exemplu Miningul bazat pe euristici [165]), pentru că se adaptează pe limbaje general acceptate (de exemplu BPMN, EPC, BPEL, YAWL, etc.), pentru că sunt capabile să modeleze porți XOR, AND și OR, neavând nevoie de pași discreți sau de activități duplicate și pentru că evită modele ce nu corespund criteriilor de corectitudine enunțate anterior.

Definiție (Rețea cauzală tip C-nets): O rețea cauzală tip C-nets este un tuplu $C = (A, a_i, a_o, D, I, O)$ unde:

- $A \subseteq \mathcal{A}$ este o mulțime finită de activități
- $a_i \in A$ este activitatea de start
- $a_o \in A$ este activitatea de final
- $D \subseteq A \times A$ este relația de dependență
- Supra-setul lui A :

$$AS = \{X \subseteq \mathcal{P}(A) \mid X = \{\emptyset\} \vee \emptyset \notin X\}, \text{ unde } \mathcal{P}(A) = \{A' \mid A' \subseteq A\} \quad (2.19)$$
- $I \in A \rightarrow AS$ definește o mulțime de posibile legături de intrare per activitate
- $O \in A \rightarrow AS$ definește o mulțime de posibile legături de ieșire per activitate

Astfel încât:

$$D = \{(a_1, a_2) \in A \times A \mid a_1 \in \cup_{as \in I(a_2)} as\} \quad (2.20)$$

$$D = \{(a_1, a_2) \in A \times A \mid a_2 \in \cup_{as \in O(a_1)} as\} \quad (2.21)$$

$$\{a_i\} = \{a \in A \mid I(a) = \emptyset\} \quad (2.22)$$

$$\{a_o\} = \{a \in A \mid O(a) = \emptyset\} \quad (2.23)$$

și toate activitățile în graful (A, D) sunt pe un traseu de la a_i la a_o .

O legătură de activități reprezintă un tuplu (a, as^I, as^O) ce arată apariția unei activități a , cu legătura de intrare as^I și legătura de ieșire as^O . În Fig. 2.4 este reprezentată rețeaua-C transformată din rețeaua Petri din Fig. 2.1.

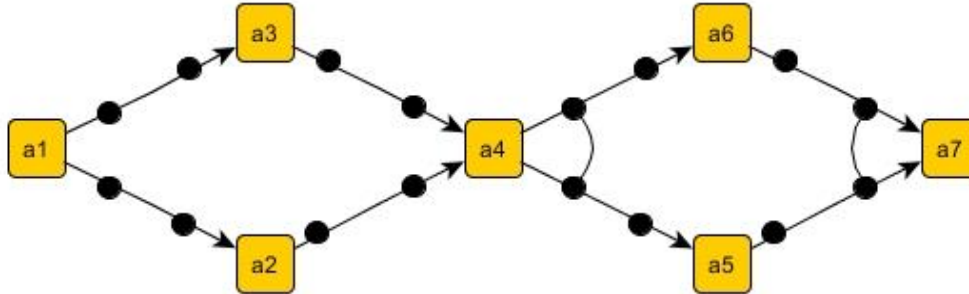


Fig. 2.4. Rețea causală tip C-nets dezvoltată din rețeaua Petri din Fig. 2.1

Definiție (Legătură): Fie $C = (A, a_i, a_o, D, I, O)$ o rețea causală tip C-nets.

$$B = \{(a, as^l, as^o) \in A \times \mathcal{P}(A) \times \mathcal{P}(A) \mid as^l \in I(a) \wedge as^o \in O(a)\} \quad (2.24)$$

este o mulțime de legături de activități. O secvență de legătură σ este o secvență de legături de activități, $\sigma \in B^*$, unde B^* reprezintă mulțimea tuturor secvențelor ce conțin elemente din B și secvența vidă.

Un exemplu de legătură de activități pentru activitatea a_4 ar fi $(a_4, \{a_3\}, \{a_5, a_6\})$. Un exemplu de secvență de legături pentru aceeași rețea ar fi $\langle (a_1, \emptyset, \{a_3\}), (a_3, \{a_1\}, \{a_4\}), (a_4, \{a_3\}, \{a_5, a_6\}), (a_5, \{a_4\}, \{a_7\}), (a_6, \{a_4\}, \{a_7\}), (a_7, \{a_5, a_6\}, \emptyset) \rangle$ ce ar semnifica execuția activităților în ordinea $\langle a_1, a_3, a_4, a_5, a_6, a_7 \rangle$.

O secvență de legătură este validă dacă între legăturile activității precedente și ale activității următoare există întotdeauna un acord. O secvență validă este o secvență de legătură care începe cu activitatea de start a_i , se termină cu activitatea finală a_o , îndepărtează doar „obligațiile” care sunt în așteptare și se încheie fără „obligații” în așteptare. Obligația reprezintă o posibilitate de avansare în rețeaua-C, un corespondent al „jetonului” folosit în rețele Petri. De exemplu, dacă este executată activitatea a_4 , atunci se elimină „obligația” corespunzătoare dinspre activitatea a_3 (dacă a_3 face parte din secvență) și se crează „obligații” de executare pentru activitățile a_5 și a_6 . Rețelele-C asigură semantici de reluare. Pentru a parcurge o rețea-C, se pornește de la activitatea de start și se ajunge în activitatea finală fără obligații în așteptare. Pentru o activitate, legăturile de intrare înlătură obligațiile existente, iar legăturile de ieșire creează alte obligații. Astfel, prin parcurgerea rețelei-C, vom asigura o succesiune de legături valide.

Pentru semantica unei rețele-C luăm în considerare doar secvențele valide, secvențele incorecte nefăcând parte din comportamentul descris de o rețea-C. Semanticile rețelelor-C sunt mult mai declarative pentru că sunt definite cuprinzând secvențe complete, mai degrabă decât o regulă de validare locală. O rețea-C poate fi convertită într-o rețea Flux de Lucru și invers [3]. Legăturile de intrare și de ieșire pot fi modelate prin tranziții discrete, de obicei reprezentate prin dreptunghiuri negre și care nu au etichete. Fiecare execuție corectă a unei rețele Flux de Lucru corespunde unei secvențe valide a unei rețele-C.

Definiție (Corectitudinea unei rețele causale tip C-nets): O rețea causală tip C-nets $C = (A, a_i, a_o, D, I, O)$ este corectă dacă pentru toți $a \in A$ și $as^l \in I(a)$ există un $\sigma \in V(C)$ și $as^o \subseteq A$, astfel încât $(a, as^l, as^o) \in \sigma$ și dacă pentru toți $a \in A$ și $as^o \in O(a)$ există un $\sigma \in V(C)$ și $as^l \subseteq A$, astfel încât $(a, as^l, as^o) \in \sigma$.

Deoarece semantica rețelelor-C conține deja „finalizarea corespunzătoare” și „opțiunea de a finaliza”, va trebui asigurat doar că există secvențe valide și că toate părțile rețelei-C ar putea fi activate de către astfel de secvențe valide.

2.1.4. Lanțul Procesului Conduc de Evenimente (EPCs)

Dacă modelele anterioare sunt îndeosebi modele matematice, următoarele modele ce vor fi prezentate sunt îndeosebi modele de reprezentare ale unui proces de afaceri, care pot fi folosite practic în cadrul unei afaceri. Una dintre primele notații folosite reprezintă notația *Lanțul Procesului Conduc de Evenimente* – *LPE* (Event-Driven Process Chain – EPC), concept dezvoltat la începutul anilor '90 de către profesorul August-Wilhelm Scheer de la Institutul pentru Sisteme Informatice pentru Afaceri, Universitatea Saarland. Pornind de la modelul de referință Y-CIM pentru industriile de producție, a fost dezvoltată metoda *ARIS* (*Architecture of Integrated Information Systems – Arhitectura Sistemelor Informatice Integrate, ASII*), care are la bază metoda EPC pentru reprezentarea proceselor [36]. În prezent, notația EPC este folosită de software-uri comerciale ca ARIS sau SAP R/3.

EPC este un model dinamic care reunește resursele statice de afaceri (sisteme, organizare, date etc.), precum și organizarea lor pentru a oferi o secvență de sarcini sau activități („procesul”), care adaugă valoare afacerii. Există patru tipuri de obiecte folosite în EPC: evenimente, funcții, reguli și resurse (de date, organizaționale, de sistem etc.). Notația standard EPC este reprezentată în Fig. 2.5. Evenimentele reprezintă punctele de declanșare pentru funcții. Evenimentele pot fi de următoarele tipuri: de start, intermediare și de final. Funcțiile reprezintă activități și au doar o singură intrare și o singură ieșire. Este obligatoriu să existe o alternanță de evenimente și activități, astfel încât un eveniment să nu poată urma altui eveniment, iar o funcție să nu poată urma altei funcții. De asemenea, regulile determină conectorii și logica fluxului de proces. Se folosesc următorii conectori: OR-split, OR-join, AND-split, AND-join, XOR-split și XOR-join⁶.

În Fig. 2.6 este redată o diagramă EPC reprezentând conversia rețelei Petri din Fig. 2.1.

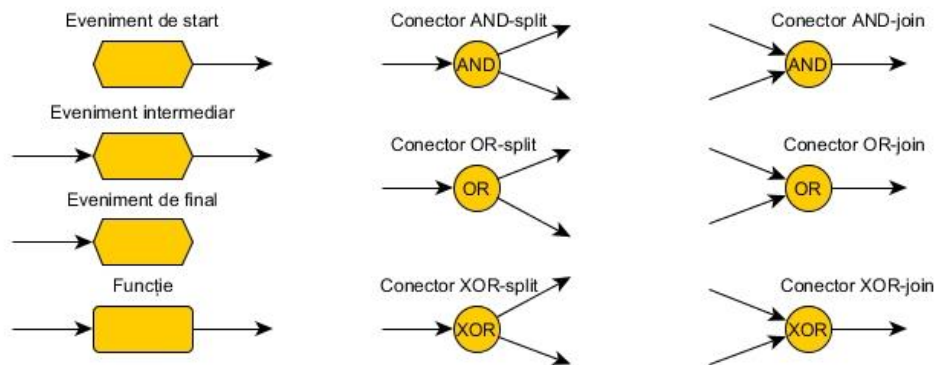


Fig. 2.5. Notația standard EPC

⁶ Split (en) = disociere; Join (en) = unificare; OR = SAU logic; AND = ȘI logic; XOR = SAU Exclusiv logic.

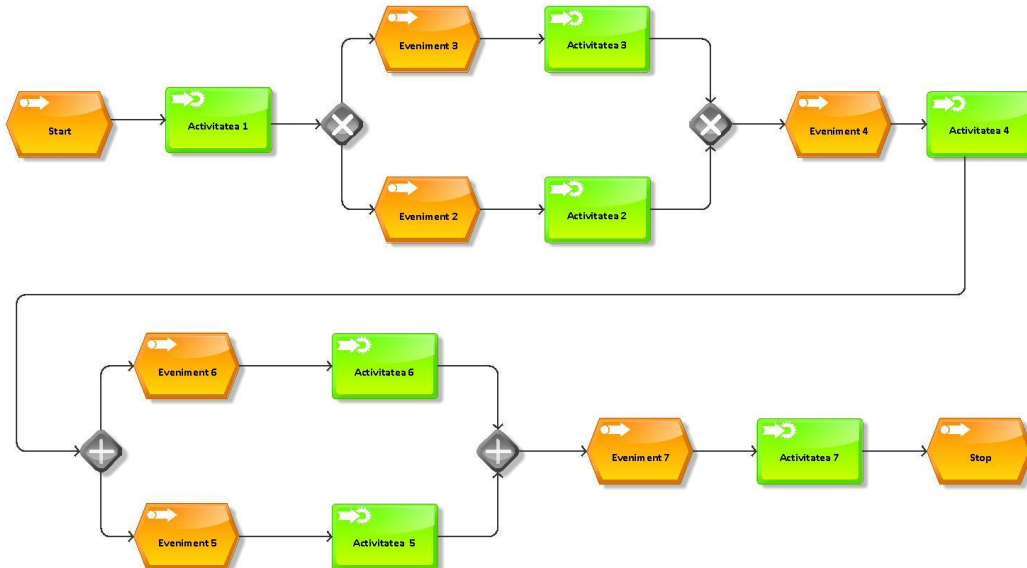


Fig. 2.6. Diagramă EPC dezvoltată din rețeaua Petri din Fig. 2.1

Încă de la apariția metodei EPC, la începutul anilor '90, a existat o dezbateră privind semantica exactă. O caracteristică ce a provocat periodic o dezbateră este non-localizarea semanticii conectorilor OR-join și XOR-join, așa numitul „cerc vicios” [98], un fel de scurt-circuit al unui proces reprezentat prin metoda EPC. Paradoxul „cercului vicios” arată că majoritatea construcțiilor complexe de proces pot introduce multe tipuri de probleme subtile din punct de vedere semantic. În ciuda acestor probleme și a diferitelor notații, conceptele de bază ale diferitelor limbaje pentru reprezentarea proceselor sunt foarte asemănătoare.

Modelele de procese ca EPC sunt utilizate în practică, în timp ce modelele academice sunt rar folosite în practică. Pentru a putea folosi abundența de tehnici de verificare specifice rețelelor Petri pe modele de proces folosite în afaceri, se pune problema transformării modelelor de proces reprezentate prin EPC în rețele Petri. Lohmann și Verbeek [109] au investigat o serie de transformări către rețele Petri, iar pentru fiecare transformare sunt investigate transformarea în sine, construcțiile rețelelor Petri din modelele de afaceri, determinarea problemelor privind transformarea și principalele aplicații ale acestor transformări.

2.1.5. Diagrame YAWL

Limbajul YAWL a fost dezvoltat inițial de către cercetători de la Universitatea Tehnică din Eindhoven și de la Universitatea Tehnică din Queensland [88]. Acronimul YAWL vine de la *Yet Another Workflow Language* (Încă Alt Limbaj pentru Flux de Lucru – IALF).

Ca fundament, limbajul YAWL are un studiu referitor la tipare ale Fluxului de Lucru [4], un studiu care poate fi interpretat ca un răspuns academic privind diferite evaluări ale firmelor prestigioase din domeniul BPM. Este criticat faptul că aceste firme oferă o multitudine de metode pentru a modela un proces/flux de lucru, iar în majoritatea cazurilor aceste metode sunt incompatibile între ele.

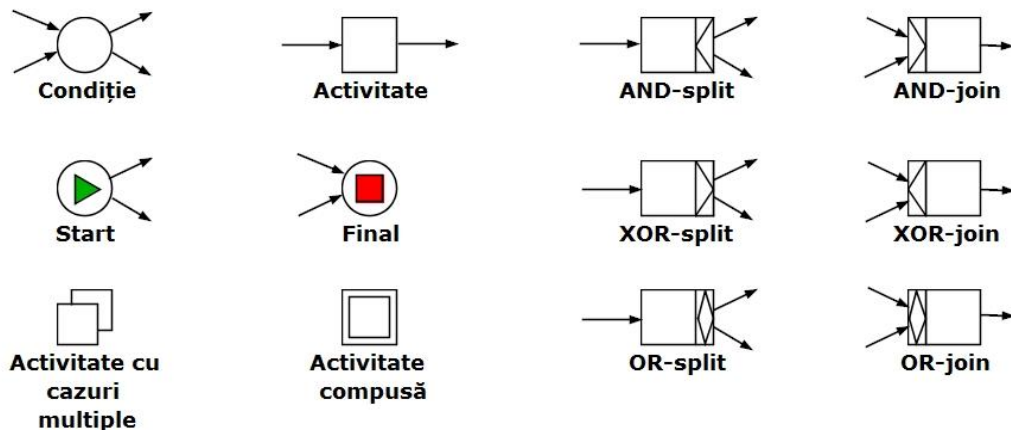


Fig. 2.7. Notăția standard YAWL

Pornind de la tiparele fluxului de lucru identificate și de la problemele folosirii acestor tipare, se propune un nou limbaj, YAWL, care pe lângă metodologie are și suport software open-source⁷. Scopul lui YAWL a fost acela de a oferi suport pentru majoritatea tiparelor, păstrând în același timp limbajul cât mai simplu. În dezvoltarea YAWL s-a pornit inițial de la rețelele Petri, peste care s-au adăugat noi formalisme ce sunt dificil suportate de către rețelele Petri, printre care amintim: instanțe multiple, sincronizări complexe, anulări. YAWL se bazează pe rețele Petri, dar este extins cu caracteristici suplimentare pentru a facilita modelarea fluxurilor de lucru complexe. O explicare detaliată a YAWL este făcută de ter Hofstede et al. [89].

Semantica YAWL este o îmbinare de elemente din rețelele Petri și elemente folosite în BPMN. Se poate observa că elementul Condiție din YAWL este exact o locație dintr-o rețea Petri, iar elementul Activitate are același înțeles ca în cazul rețelelor Petri. Elementele de unificare și disociere (AND, XOR, OR) sunt reprezentări schematice pentru construcții Petri complexe. Ce apare nou sunt activitățile cu cazuri multiple și activitățile compuse, care reprezintă un element de noutate față de rețelele Petri clasice. De asemenea, apare conceptul de regiune de anulare. Prin regiune înțelegem un set de activități care pot fi anulate în totalitate, dacă o anumită condiție nu e îndeplinită sau dacă apare un eveniment excepțional. Datorită suportului regiunilor de anulare, limbajul YAWL are un caracter tranzacțional.

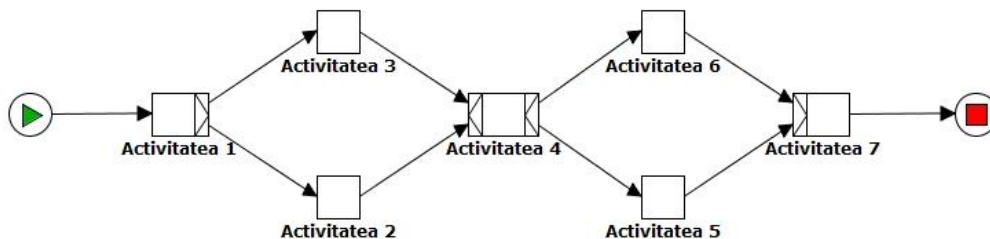


Fig. 2.8. Diagramă YAWL dezvoltată din rețeaua Petri din Fig. 2.1

⁷ <http://www.yawlfoundation.org/>

YAWL permite specificarea altor dimensiuni, pe lângă perspectiva controlului de flux. Astfel, este posibilă definirea acelor actori care pot executa anumite activități bazat pe rolurile avute în organizație. YAWL suportă majoritatea tipurilor de resurse flux de lucru [148]. Un studiu recent arată probleme ale limbajului YAWL [31], fiind criticată aplicabilitatea practică (din cauza fundamentului rețelelor Petri) și slaba popularitate printre practicienii BPM. Într-adevăr, YAWL nu a avut impact mare asupra industriei, fiind folosit doar în câteva companii, printre care: GECKO, Impello, European Defence Agency, MMT Seguros⁸ și în general firme care au avut colaborări cu mediile academice. De asemenea, există anumite mijloace software care folosesc YAWL, printre care amintim: FERP, InsuraPro, Axxerion, FlowConnect și Woped⁸. Un caz aplicativ interesant este crearea unui sistem YAWL folosit în industria filmului, sistem numit YAWL4Film care a fost folosit de compania Porchlight în 2008 [89, 171]. Alt caz interesant este YAWL4Healthcare prezentat în [89].

2.1.6. Diagrame BPMN

Cea mai utilizată metodă de a reprezenta procesele de afaceri o reprezintă diagramele *BPMN*. Inițial, acronimul *BPMN* însemna *Business Process Modelling Notation (Notația Modelării Proceselor Afacerii, NMPA)* pentru versiunile *BPMN 1.x*, dar ulterior acesta a fost redenumit în *Business Process Model and Notation (Notația și Modelul Proceselor de Afaceri, NMPA 2.0)*, odată cu apariția versiunii *BPMN 2.0* [100]. În cadrul abordării BPM, *BPMN (Business Process Model and Notation)* ocupă un loc important ca limbaj al modelării de proces. *BPMN* este un limbaj expresiv, capabil să descrie caracteristicile comportamentale ale unui proces în mod compact într-o diagramă. Versiunea 1.0 a *BPMN* a fost lansată în 2004 de către *Business Process Management Initiative, BPMI (Inițiativa Managementului Procesului de Afaceri, IMPA)* condus de un start-up numit *Intalio*⁹, iar *BPMN 2.0* a fost lansată în 2011 de către *Object Management Group, OMG (Grupul Managementului de Obiect, GMO)* [151]. Ultima versiune pentru *BPMN* este 2.0.2 apărută în Octombrie 2013, iar această versiune a fost publicată de ISO ca ediția standard ISO/IEC 19510/2013¹⁰.

Creatorii de modele de procese sunt interesați în principal de notația grafică. Versiunile *BPMN 1.2* și *BPMN 2.0* au foarte puține elemente grafice diferite. În versiunea *BPMN 2.0* s-a pus accent pe model, mai bine zis pe meta-model, adică pe crearea unui format interschimbabil XML util pentru standardizarea proceselor. *BPMN* a stat în umbră până în 2008, când a fost adoptat de marii jucători din piața BPM: SAP și ORACLE. În paralel, alți mari jucători, IBM și Microsoft, au dezvoltat altă versiune, anume *Business Process Execution Language, BPEL (Limbaj de Execuție a Proceselor de Afaceri, LEPA)*.

În ultimii zece ani, *BPMN* s-a răspândit foarte mult, devenind cea mai utilizată tehnică de modelare a procesului la nivel mondial, susținută fiind de principalele companii furnizoare de soluții de IT¹¹.

⁸ O listă completă se găsește pe site-ul <http://www.yawlfoundation.org/pages/impact/uptake.html>

⁹ <http://www.intalio.com/>

¹⁰ Conform <http://www.omg.org/spec/BPMN/>

¹¹ <http://www.bpmnforum.com/FAQ.htm>

În standardul BPMN 2.0 sunt peste 100 de elemente, dar doar aproximativ 20% dintre acestea sunt folosite frecvent [99]. De precizat că terminologia BPMN a fost adaptată în limba română în cadrul proiectului *CertiBPM* [63] și în cadrul acestui proiect a fost creat un poster cu elementele BPMN 1.x, având toată terminologia tradusă în limba română (vezi ANEXA 1). BPMN este un standard suportat de mulți producători de soluții BPM. El este asemănător diagramelor de activități UML și asemănător EPC. Standardul BPMN conține numeroase elemente, cele mai uzitate fiind redată în Fig. 2.9 și în Fig. 2.10.

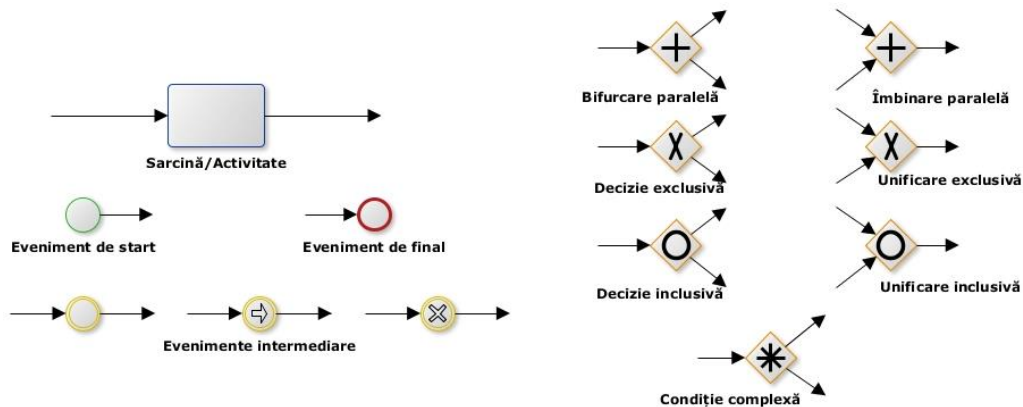


Fig. 2.9. Notăția standard BPMN

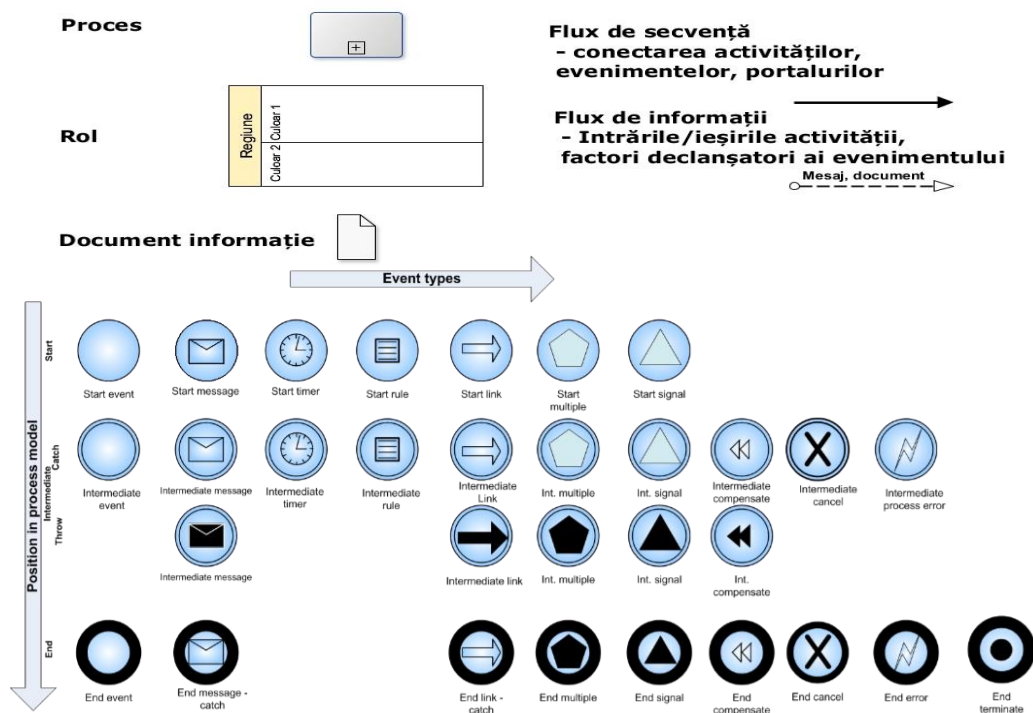


Fig. 2.10. Notății frecvent folosite în reprezentările modelelor de procese, conform BPMN (preluare din materialele de formare *CertiBPM*)

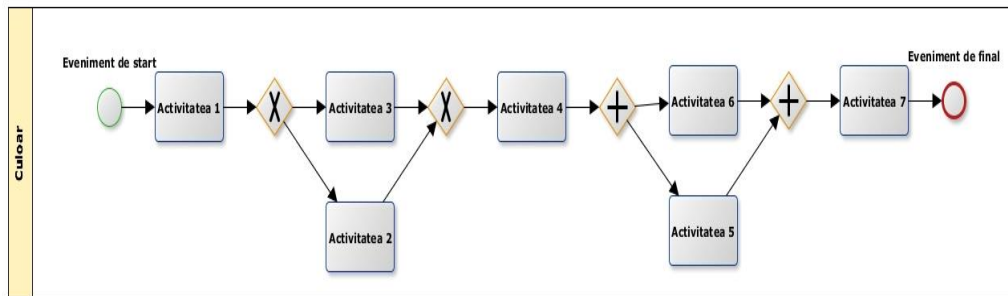


Fig. 2.11. Diagramă BPMN dezvoltată din rețeaua Petri din Fig. 2.1

În termeni BPMN [25], activitățile sunt numite *sarcini* (din engleză, *task*). Asemănător cu YAWL, activitățile pot fi compuse. Logica de rutare nu este comandată din sarcini ca în YAWL, ci din *portaluri* (din engleză, *gateway*) separate. Bifurcările au loc datorită condiționărilor de date. Evenimentele sunt comparabile cu locațiile din rețelele Petri. Evenimentele au un singur arc de intrare (excepție evenimentul de start, care nu are arc de intrare) și un singur arc de ieșire (excepție evenimentul de final, care nu are arc de ieșire). Bifurcările și unificările pot avea loc doar folosind portaluri.

Dintre toate modele de reprezentare a unui proces de afaceri, diagramele BPMN sunt cele mai folosite. În general, comercianții suportă doar un set de bază din totalul de elemente disponibile în BPMN 2.0 și într-adevăr acest set este suficient din punct de vedere practic. Referindu-se la stil, Bruce Silver în [151] compară stilul artistic cu setul de elemente BPMN, afirmând că pentru a scrie o carte e necesar doar un număr limitat de cuvinte, dar dacă cartea se dorește a avea cât mai multe figuri de stil, atunci se va folosi un vocabular mai bogat. Asemenea, un practician BPM trebuie să cunoască cât mai multe din elementele BPMN pentru a descrie cât mai expresiv procesul. În mediul academic, conceptului BPMN îi este recunoscută supremația și există multe mijloace software de conversie a acestui format în alte formate pentru reprezentarea modelelor de proces [95].

Sistemul standardizat de notații BPMN poate fi utilizat:

- În cazul colectării cerințelor aferente noului sistem IT din organizație (prin defalcarea descrierii comportamentului sistemului existent și a interacțiunilor cu utilizatorii). Redarea stării existente inițial definește situația AS-IS (= așa cum este);
- Poate fi utilizat în cazul optimizării unui proces, doar dacă acesta a fost utilizat anterior pentru a descrie starea inițială, AS-IS;
- Dacă procesul este modelat folosind BPMN, modelul acestuia poate fi folosit pentru a simula comportamentul său, utilizând mijloace software dedicate;
- Diagrama de proces realizată folosind sistemul BPMN poate fi transformată automat în diagramă de execuție (codul de program), dar aceasta trebuie definită în detaliu.

Dezvoltarea software tradițională utilizează limbajul de programare pentru a crea o aplicație care poate fi înțeleasă de programatori. Cea mai recentă abordare a dezvoltării de software utilizează diagrame ca punct de plecare sau ca definiție a logicii proceselor de afaceri destinate realizării aplicațiilor software și pot fi direct executate ca program. Faza de transformare către aplicația executabilă este prezentată în Fig. 2.12.

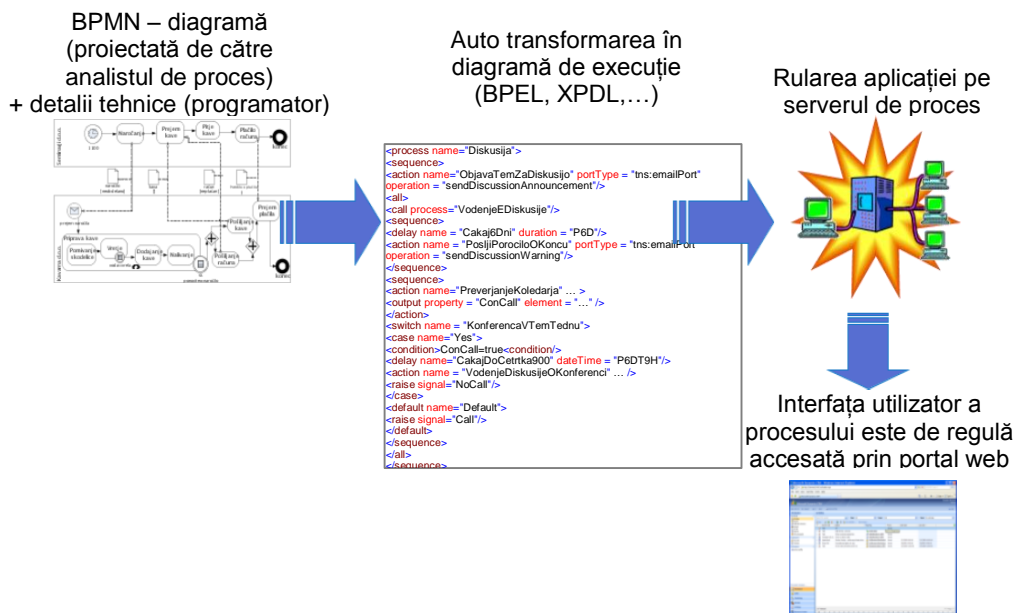


Fig. 2.12. Faza de transformare către aplicația executabilă (preluare din materialele de formare CertiBPM)

După cum s-a menționat, o diagramă de proces poate fi folosită ca un program software: mijloace specifice o pot transforma în cod program; codul este transmis serverului de proces, unde rulează ca un client de server sau o aplicație web; procesul se expune utilizatorului prin interfața sa desktop (Fig. 2.12). Această fază are un specific tehnic pronunțat și necesită muncă susținută a echipei formate din analiști de proces și din programatori de la nivelul organizațional. Datorită acestui mod de operare și de dezvoltare se realizează o mai bună aliniere a cerințelor afacerii și a sistemului IT decât în cazul dezvoltării de aplicații software destinate satisfacerii anumitor nevoi punctuale ale organizației (variante tradițională a dezvoltărilor și implementărilor de software) deoarece diagrama sau modelul de proces poate constitui baza fezabilă a unui program [130].

2.2. Data Mining

Tehnicile de Minerit de Date (*Data Mining*) sunt bazate pe găsirea de tipare relativ simple în seturi de date de mari dimensiuni [80]. Deoarece *Process Mining* este derivat din *Data Mining* și aplicat pentru analiza proceselor, anumite noțiuni de *Data Mining* trebuie însușite pentru a înțelege corect tehnicile *Process Mining*. Tehnicile de *Data Mining* sunt mult mai numeroase, dar în continuare vom prezenta doar pe acelea care aduc valoare în *Process Mining*.

2.2.1. Definiție. Caracteristici. Beneficii

În organizațiile de tip economic, datele și informațiile sunt asociate diferitelor aspecte ale proceselor de afaceri ale acestora, precum: piețe, concurenți, furnizori,

distribuitori, clienți, precum și informațiilor interne legate de procesele de producție, organizarea muncii, depozitare etc. Volumul mare de date dintr-o organizație se referă la activități și procese ale acesteia, deci într-un cadru contextual acestea se transformă în informații și cunoaștere organizațională, sau chiar în înțelepciune.

Pentru multe firme din Occident, procesul de explorare și analiză a datelor nu este unul nou. Tehnicile de *Data Mining* permit realizarea acestor analize mult mai rapid, cu o eficiență și eficacitate sporite. În numeroase cazuri practice ale analizei datelor, tehnicile existente aplicate permit ca datele colectate pentru un anumit scop să poată fi folosite în multe alte scopuri. Astfel, se conturează modul de acțiune al *valorilor intangibile* ale unei organizații [75].

Ca urmare a perfecționării modalităților de colectare și stocare a datelor, numeroase organizații se confruntă cu problema deținerii unor arhive sau spații de stocare de tip baze de date de dimensiuni foarte mari, dar cărora ar trebui să le găsească o utilizare. De exemplu, companiile furnizoare de cărți de credite înregistrează permanent datele referitoare la tranzacțiile comerciale, supermarketurile înregistrau date privind cumpărăturile, utilizarea cupoanelor de reduceri ș.a.m.d. Apariția necesității de a valorifica toate aceste date stocate privind diferite procese de afaceri a condus la apariția și dezvoltarea demersului de *Data Mining* și care este asociat cu automatizarea extragerii și analizei datelor aferente unei organizații [66].

Datele brute sunt rareori folositoare în forma în care ele se găsesc inițial. Valoarea lor reală este dată de posibilitatea de a extrage din ele informații care să sprijine fundamentarea deciziilor sau pentru înțelegerea fenomenelor care guvernează viața organizației. În mod tradițional, analiza datelor era un proces strict manual, ce putea fi realizat de unul sau mai mulți analiști, care deveneau foarte familiari cu datele și furnizau periodic rapoarte, cu ajutorul metodelor statistice. O asemenea abordare a devenit dificil de pus în practică, pe măsură ce volumul de date crește până la dimensiuni uriașe (după cum este descris în prezent prin fenomenul *big data*). Cine ar putea „înțelege” o bază de date cu câteva milioane de obiecte, fiecare având câteva zeci de câmpuri? Pentru a complica situația, volumul de date crește într-un ritm atât de accelerat încât analiza manuală (chiar și atunci când este posibil!) poate foarte ușor să furnizeze erori sau greșeli de interpretare [125].

Prin Data Mining se înțelege procesul de extragere automată a unor informații cu caracter predictiv din marile baze de date. Astfel, pot fi prognozate, previzionate tendințele viitoare și se pot identifica anumite comportamente ale actorilor implicați în derularea unor procese de afaceri, care anterior nu au fost sesizate sau caracterizate (nu au putut fi descrise precis din lipsa datelor), și care le-au scăpat responsabililor din organizație [161].

Tehnicile de Data Mining permit extragerea informațiilor, analiza acestora și realizarea de previziuni, pornind de la date istorice ale organizației. Acest demers poate susține deciziile manageriale (Fig. 2.13).

Comunitatea de cercetători și practicieni interesați de problema automatizării extragerii și analizei datelor (denumită uneori și explorarea de date sau recoltarea informațiilor sau arheologia și prelucrarea datelor) a crescut rapid, cunoscută fiind emulația creată în jurul demersului de descoperire a cunoștințelor din sisteme de baze de date (*Knowledge Discovery in Databases*, KDD). Primul workshop pe tema KDD a avut loc în anul 1989; el s-a transformat ulterior într-o conferință internațională care reunește anual peste 500 de participanți¹².

¹² <http://www.kdnuggets.com/meetings/past-meetings-kdd.html>



Fig. 2.13. Acumularea de cunoaștere ce constituie suportul deciziei de la nivelul managementului unei organizații

Prin definiție, KDD este „un proces non trivial de extragere a informațiilor, date anterior necunoscute și potențial utile”, dar și ca „știința de a extrage informații utile din masive de baze de date”, după Fayyad și alții, 1996 [70]. În acest context, datele sunt o colecție de fapte, iar modelul este considerat un nivel superior de exprimare, care descrie datele sau un subset din acestea.

În analiza datelor, caracteristicile modelelor pe care KDD le identifică trebuie să fie *valide, de noutate, neredundante, folositoare* și în cele din urmă *de înțeles*. În final, este de dorit ca modelele descoperite să fie de înțeles spre a putea fi analizate ulterior, pentru a realiza analize de cauzalitate privind diferitele aspecte ale proceselor de afaceri (analize cauză – efect).

Deoarece extragerea de date (*Data Mining*) este partea centrală a procesului de descoperire de cunoștințe din bazele de date (KDD), termenii *Data Mining* și descoperirea de cunoștințe din baze de date au fost utilizați alternativ de mulți cercetători din domeniu. În ultimul timp însă, se face o distincție clară între cei doi termeni, referitoare la faptul că descoperirea de cunoștințe din bazele de date (KDD) poate fi considerată ca procesul de extragere a informațiilor folositoare și interesante din baza de date. Autorii care fac deosebire între *Data Mining* și KDD consideră KDD ca fiind un proces iterativ și interactiv complex, iar demersul de *Data Mining* ca fiind o etapă în cadrul acestuia.

Procesul de descoperire de cunoaștere (sau cunoștințe, în unele cazuri) din baze de date este prezentat în Fig. 2.14. În concordanță cu cele precizate de Fayyad și colaboratorii (1996), KDD este procesul de folosire, explorare și exploatare a datelor și informațiilor localizate în sisteme de baze de date, urmând etapele: selectare, pre-procesare, transformarea datelor, *Data Mining* și interpretare-evaluare [69]. Astfel, procesul de KDD cuprinde șapte etape, unele dintre acestea vizibile în Fig. 2.14, după cum urmează:

1. Analiza domeniului - este analizată natura datelor din domeniu și se definește obiectivul acțiunii de descoperire. Dacă există cunoștințe anterioare în acest domeniu, acestea sunt evaluate;

2. Selectarea sau segmentarea datelor se realizează în concordanță cu anumite criterii, ceea ce poate însemna eliminarea unor câmpuri sau rânduri din bazele de date (sau ambele);

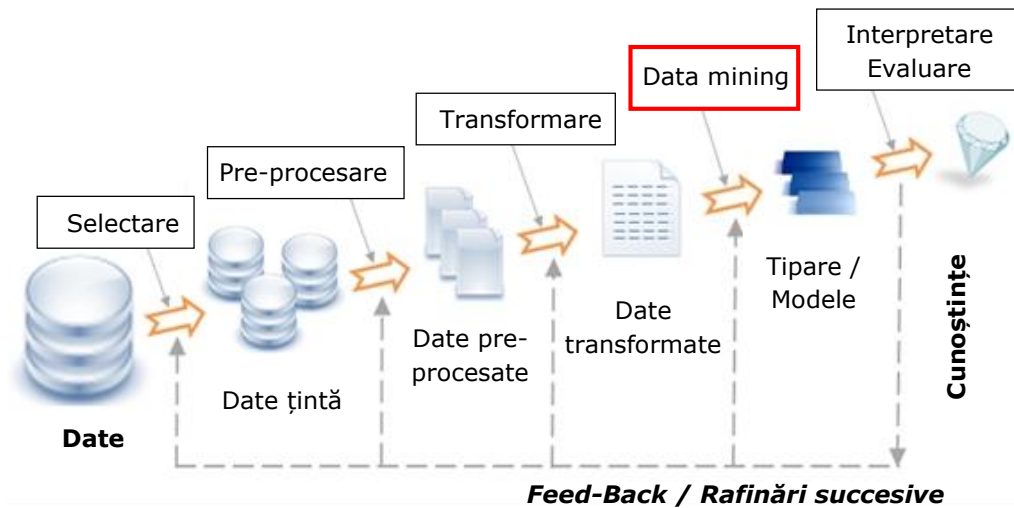


Fig. 2.14. Etapele KDD

3. Pre-procesarea corespunde stadiului de curățare a datelor în care anumite informații sunt îndepărtate, determinându-se de asemenea metode de lucru cu date lipsă;

4. Transformarea datelor presupune realizarea unei reprezentări a datelor care este compatibilă cu algoritmul *Data Mining* ce urmează a fi implementat. Datele sunt analizate cu scopul determinării unor caracteristici pentru reprezentarea lor în concordanță cu obiectivul urmărit;

5. Data Mining este etapa în care se realizează extragerea modelelor pe baza datelor selectate, pre-procesate și transformate anterior. *Data Mining* se concretizează prin aplicarea unor algoritmi specifici pentru extragerea modelelor din date. Calitatea acestei etape depinde foarte mult de calitatea modului de realizare a etapelor precedente. Activitățile și acțiunile suplimentare aferente procesului de KDD, cum sunt prepararea datelor, selectarea datelor, etapa de curățare, integrarea cunoștințelor anterioare necesare, sunt esențiale pentru a asigura că se vor extrage cunoștințe folositoare din date;

6. Interpretarea și evaluarea - Modelele identificate de sistem, în urma algoritmului aplicat, sunt interpretate în cunoștințe care pot fi folosite ca suport decizional (de exemplu, predicțiile și problemele legate de clasificare, sumarizând conținutul bazei de date și explicând fenomenele observate);

7. Consolidarea cunoștințelor descoperite - modelele descoperite sunt puse în folosință. Un mod de utilizare plauzibil îl reprezintă încorporarea cunoștințelor obținute într-un alt sistem pentru acțiuni suplimentare, documentarea modelelor și transmiterea acestora către părți interesate, precum și reaplicarea KDD, folosind drept fundament aceste noi cunoștințe.

În acest context, demersul de *Data Mining* ar putea fi definit ca un **proces de aplicare a unor algoritmi specifici pentru căutarea și extragerea unor șabloane, modele sau regularități ascunse în interiorul unei colecții sau unui sistem de baze de date**. El utilizează metode sofisticate de analiză statistică. Odată găsită, informația ascunsă în spatele datelor brute trebuie prezentată într-o formă accesibilă (cu ajutorul rapoartelor, tabelelor, graficelor etc.). **Statistica este pilonul central al întregii activități de explorare a bazelor de date**. Atât în ceea ce privește validarea ipotezelor, cât și în ceea ce privește analiza exploratorie a datelor

și realizarea predicțiilor, statistica are o importanță fundamentală. Calitatea datelor este critică pentru obținerea unor rezultate coerente din *Data Mining*.

Primul pas în KDD și în *Data Mining* este construirea unui sistem de *selecție, culegere și extragere* a datelor într-o formă cât mai precisă și fidelă. Datele primare de calitate slabă pot compromite întregul proces de analiză. Sistemul cu ajutorul căruia se realizează în fiecare zi activitățile din organizație precum preluarea comenzilor, înregistrările contabile, gestiunea stocurilor etc. poartă numele de **sistem operațional**, iar informațiile pe care acest sistem le creează poartă denumirea de *date operaționale*. Explorarea datelor este folosită pentru a da consistență și substanță acestor date operaționale, care sunt date brute, primare. Este un proces de *descoperire*, în urma căruia ies la iveală informații pe care datele brute nu le pot oferi în forma actuală. Nu sunt utilizate rapoarte cu o structură prestabilită, ci este permis decidentului crearea de interogări ale bazei de date pornind de la informațiile de care are nevoie în momentul respectiv (etape descrise și în Fig. 2.14)

Bazele de date utilizate pentru procesul de analiză și explorare, *Data Mining*, trebuie să fie deosebit de flexibile, deoarece utilizatorul poate face explorarea într-o manieră ad-hoc, fără a urma standarde prestabilite. De asemenea, baza de date trebuie să cuprindă foarte multe *referințe încrucișate*, pentru a putea asocia cu ușurință câmpurile și a extrage rapid rapoarte rezumative. O altă cerință importantă este aceea ca datele să fie cât mai *condensate* cu putință, astfel încât timpul necesar obținerii răspunsului la o interogare să fie cât mai redus posibil (câteva sutimi de secundă). Dacă timpul necesar obținerii unui raport este prea mare (de ordinul secundelor), utilitatea procesului este pusă sub semnul întrebării.

În general, etapele unui proces de *Data Mining* (Fig. 2.15) sunt următoarele [161]:

1. definirea problemei - Pentru o utilizare cât mai eficientă a tehnicilor de explorare trebuie mai întâi specificate foarte clar obiectivele studiului. De exemplu, în cazul unei campanii de direct mailing, obiective precum: creșterea numărului de răspunsuri, creșterea ratei răspunsurilor sau diminuarea costului per răspuns sunt obiective diferite, care vor necesita metode de analiză diferite. Definirea clară a obiectivelor contribuie substanțial la măsurarea eficienței procesului;

2. selectarea tehnicii de analiză ce va fi utilizată, pornind de la problema specifică de rezolvat;

3. selecția și pregătirea datelor. Aceasta este etapa care consumă cel mai mult timp (între 50% și 85% din timpul total alocat proiectului). După cum menționam anterior, calitatea rezultatului final depinde foarte mult de acuratețea datelor de intrare;

4. construirea modelului se realizează ținând seama de tehnicile de analiză alese și de natura datelor disponibile;

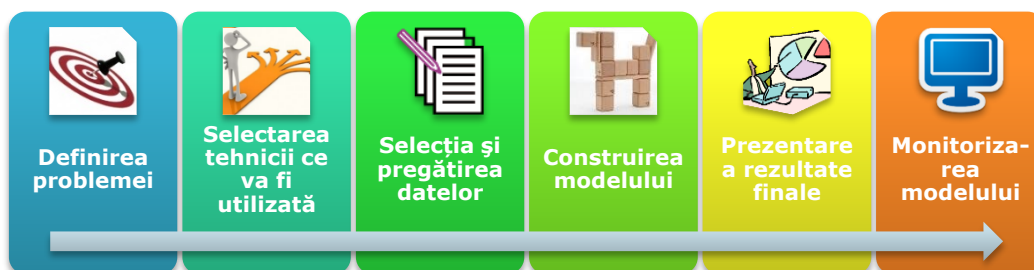


Fig. 2.15. Etapele unui proces de *Data Mining*

5. prezentarea rezultatelor finale se realizează sub forma unor rapoarte conținând text, tabele și grafice pentru a facilita și sprijini deciziile manageriale;

6. monitorizarea modelului are ca scop măsurarea eficienței și utilității sale.

Caracteristicile cheie ale unui proces adecvat de Data Mining [161] sunt:

- ✓ **Utilizare facilă** - Procesul trebuie să fie ușor de realizat și să nu solicite o instruire prea îndelungată a utilizatorilor. Este important ca utilizatorul să se poată concentra asupra propriilor sale necesități și probleme, asupra modului de formulare a interogărilor, și nu asupra modului de utilizare a programului. Cu cât este mai dificilă rularea aplicației software folosită pentru *Data Mining*, cu atât este mai sigur că aceasta nu va fi folosită chiar de către persoanele care ar avea cea mai mare nevoie de ea;
- ✓ **Accesibilitatea** - Valoarea reală a explorării datelor se obține oferind acest instrument celor care au nevoie de el, care ar fi interesați în aflarea de informații și cunoștințe extrem de utile în urma analizei datelor. Managerii de vânzări, de exemplu, îl vor folosi pentru a evalua eficiența agenților de vânzări și a distribuitorilor, pe fiecare tip de client în parte. Managerii de achiziții îl vor folosi pentru a determina evoluția stocurilor pe fiecare tip de material. Pentru contabili, ar putea fi un instrument util în construirea bugetelor și a situațiilor financiare și așa mai departe;
- ✓ **Răspuns rapid** - Este foarte important ca analiza și furnizarea informației să fie realizată în cel mai scurt timp posibil (de ordinul zecimilor sau sutimilor de secundă). Numai în acest fel capacitățile procesului de *Data Mining* vor fi folosite la întregul lor potențial. Decidenții vor fi mai puțin înclinați spre a utiliza programul atunci când știu că trebuie să aștepte secunde bune pentru a obține rezultatul unei interogări, deoarece acest lucru ar însemna pentru ei o pierdere de timp importantă;
- ✓ **De a oferi informație la zi (actualizată)** - Folosirea unor date vechi de săptămâni sau luni reduce considerabil eficiența și utilitatea procesului. Cu cât datele sunt mai actuale, cu atât mai repede poate fi produs un plan de măsuri pentru rezolvarea problemelor sau exploatarea oportunităților identificate prin procesul de *Data Mining*. Cele mai relevante exemple în acest caz sunt cele legate de descoperirea unor pierderi sau comportamente inadecvate (adesea recurgându-se la analiza extremelor), în cazuri precum: combaterea declinului înregistrat la vânzările unor produse, stoparea pierderilor de resurse prin fraude sau defecțiuni aferente unor procese, identificarea și eliminarea unor pierderi financiare prin realizarea unor procese simultan de către operatori diferiți etc.
Beneficiile procesului de Data Mining sunt următoarele:
- ✓ **Costul mai redus al procesării datelor** - Atunci când decidenții obțin datele necesare în urma unor analize făcute de ei înșiși, dispăre necesitatea de a crea aplicații informatice specializate pentru obținerea și raportarea acestor date. Astfel, programatorii își pot concentra atenția asupra altor probleme legate de operarea sistemelor din firmă;
- ✓ **Reducerea cantității de hârtie ocazionate de tipărirea rapoartelor și situațiilor** - Marea majoritate a programelor informatice oferă managerilor sumele totale ce reflectă volumul activității din domeniul lor (de exemplu, vânzările totale, cifra de afaceri, costurile etc.). În cazul procesului de explorare a datelor prin *Data Mining*, managerul poate începe prin a solicita aceste sume totale pe ecranul computerului, apoi poate cere defalcarea lor în funcție de informația (criteriul) pe care dorește să o primească (și pe care el o consideră

relevantă în analiză). Odată ce deține această informație, el va cere eventual tipărirea unui raport, foarte probabil mult mai util decât cel ce prezintă doar totalurile (justificând așadar mai bine costul hârtiei pe care a fost tipărit și nu numai);

- ✓ **Reducerea încărcării sistemului operațional** - Procesul de *Data Mining* utilizează de regulă baze de date care au fost realizate și condensate fie offline, fie în afara orelor de program din firmă. Întrucât procesele de sortare, selectare și raportare sunt mai ușoare pe aceste baze de date, încărcarea sistemului informatic al organizației va fi substanțial redusă.

Ca urmare, managementul organizațiilor trebuie să conștientizeze faptul că procesul de *Data Mining* nu este unul deosebit de complex și de costisitor, iar beneficiile sale sunt extrem de importante pentru crearea avantajului competitiv sustenabil. Astfel, planul de dezvoltare organizațională ar trebui să includă achiziționarea și implementarea sistemelor de acest fel.

2.2.2. Tehnici pentru Data Mining

Există trei clase de activități (Fig. 2.16) relative la demersul de *Data Mining* [161]:

1. Descoperirea, care este procesul de căutare în bazele de date a unor modele ascunse, fără ipoteze și idei preconceptuate asupra naturii acestora. Cu alte cuvinte, inițiativa de a căuta aceste modele îi aparține programului, înainte ca utilizatorul să formuleze interogările;

2. Modelarea predictivă, care presupune utilizarea modelelor descoperite anterior în scopul prezicerii viitorului. Programul permite utilizatorului să introducă înregistrări care au unele câmpuri cu valoare necunoscută; valoarea acestora urmează a fi estimată de program tot cu ajutorul unor șabloane/modele (pattern), identificate anterior;

3. Analiza extremelor, este procesul ce permite detectarea, pe baza modelelor identificate în prima etapă, a situațiilor anormale sau neobișnuite ce caracterizează obiectele din baza de date. Pentru a descoperi cazurile neobișnuite, va trebui să se răspundă mai întâi la întrebările: *Ce înseamnă normalul? Care este regula?* Cazurile care înregistrează o deviere, cu o anumită mărime, de la această regulă, vor fi considerate cazuri neobișnuite. Analiza extremelor merge așadar mai departe decât descoperirea și modelarea predictivă, ajutându-ne nu doar să obținem așa numitele „cunoștințe obișnuite”, ci să identificăm și situațiile anormale.

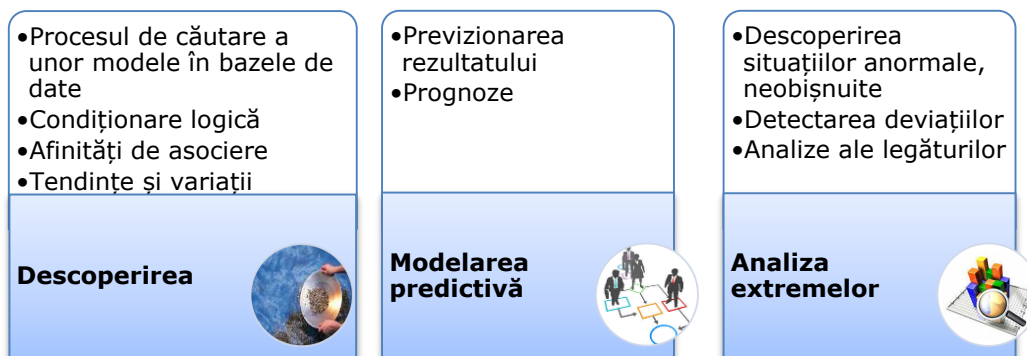


Fig. 2.16. Clase de activități aplicabile în demersul de *Data Mining*

Într-o organizație se pot desfășura trei tipuri de activități de *Data Mining* [161] (Fig. 2.17):

1. Explorarea episodică presupune efectuarea analizei datelor cu ocazia unui anumit eveniment: o campanie de marketing, de exemplu. Se pot analiza datele legate de această campanie pentru a se înțelege o serie de relații de cauzalitate (de exemplu, relația dintre cheltuielile pentru publicitate și volumul vânzărilor) și se poate folosi această analiză pentru a planifica și preziona rezultatele viitoarelor campanii de marketing. De regulă, analiștii sunt cei care realizează activități de *Data Mining* episodice.

2. Explorarea strategică presupune luarea în considerare a unui set de date mai extins, pentru a dobândi o înțelegere de ansamblu asupra unor parametri care caracterizează activitatea firmei (de exemplu, profitabilitatea). În acest caz, explorarea strategică poate răspunde la întrebări de genul: *Care sunt principalele noastre surse de profit? SAU Care este relația dintre portofoliul nostru de produse și segmentele noastre de piață?*

3. În cazul explorării continue, obiectivul este acela de a înțelege ce schimbări au avut loc în organizație și în mediul ei extern pe o anumită perioadă de timp și care sunt cei mai importanți factori care au influențat acele schimbări. Managerul s-ar putea întreba, de exemplu: cum au evoluat vânzările în ultimul semestru, per total și pe tipuri de produse? sau cum s-a modificat nivelul de satisfacție a consumatorilor în ultimele luni și din ce motive?

Pentru realizarea procesului de *Data Mining* pot fi utilizate diferite tehnici de extragere și analiză a datelor [78] (Fig. 2.18):

- **predictive** (bazate pe învățare supervizată), ce utilizează un set de variabile (numite predictor) prin intermediul cărora se realizează predicții relative la valorile continue sau discrete ale altor variabile (numite variabile de decizie). Modelele predictive bazate pe *inteligentă artificială* se construiesc în cadrul unei faze de antrenare, prin care modelul învață să prezică răspunsul potrivit (decizia), când la intrare se prezintă diverse seturi de valori ale predictorilor. După consumarea fazei de antrenare, modelul poate fi folosit în predicție, pentru a rezolva, după caz, fie probleme de *clasificare* (dacă variabila de decizie este nominală sau discretă), fie probleme de *regresie* (dacă variabila de decizie este continuă);



Fig. 2.17. Activități posibile de *Data Mining* în organizație

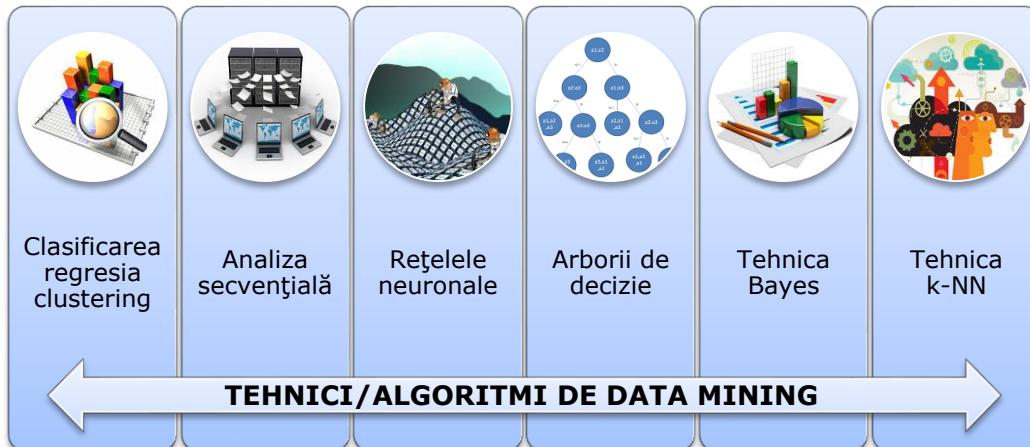


Fig. 2.18. Tehnici de Data Mining

- **descriptive** (bazate pe învățare nesupervizată), destinate extragerii unor șabloane model (pattern, structuri inteligibile) din sistemele de baze de date. Metodele *Data Mining* descriptive formează a doua mare categorie din *Data Mining*. Spre deosebire de modelele predictive, metodele descriptive (precum cele de *clustering*) tratează uniform variabilele, fără să distingă între predictorii și răspuns (decizie), ca atare învățarea nu este supervizată (în sensul învățării din exemple, adică al furnizării de răspunsuri în cadrul fazei de training). Metodele descriptive permit descrierea și explicarea fenomenelor caracteristice sistemului studiat pe baza modelelor descoperite.

Ca urmare, în practica de aplicare a demersului de *Data Mining* se întâlnesc tehnici precum:

1. Clasificarea, regresia sau clustering-ul (asocierea);
2. Analiza secvențială;
3. Rețelele neuronale;
4. Arborii de decizie;
5. Tehnica Bayes;
6. Tehnica k-NN.

2.2.2.1. Clasificarea, regresia și clustering-ul

Clasificarea analizează un set de date și stabilește un ansamblu de reguli pe baza cărora vor fi grupate datele obținute în viitor. Aplicațiile software de *Data Mining* identifică automat clasele (sau „ciorchinii”), studiind șablonul model (pattern) al datelor existente. Odată ce au fost generate clasele se poate stabili, pe baza unor caracteristici precizate, cărei clase îi aparține un anumit obiect din baza de date studiată. De exemplu, o clasă poate fi un segment de piață. Atunci când firma are un client nou ea poate stabili, cunoscând caracteristicile acestuia, în ce segment se încadrează. Astfel, apar *regulile de asociere* care implică existența unor relații între diferite seturi de obiecte din baza de date.

Clasificarea este grupată adesea și cu **regresia**, acestea reprezentând cea mai largă categorie de aplicații, constând în construirea de modele pentru previzionarea apartenenței la un set de clase (clasificare) sau a unor valori (așa cum

se întâmplă în cazul regresiei). Există câteva tehnici dedicate rezolvării problemelor de clasificare și regresie, dintre care arborii decizionali, tehnica Bayes, rețelele neuronale și k-NN se bucură de o largă recunoaștere [122].

Clasificarea sau **analiza de tip cluster** este o tehnică descriptivă utilizată pentru gruparea entităților similare dintr-un set de date sau în egală măsură pentru evidențierea entităților care prezintă diferențieri substanțiale față de un grup. Tehnicile de grupare în clustere se bazează pe algoritmi din sfera rețelelor neuronale, algoritmi demografici, k-NN etc.

Pentru rezolvarea problemelor de clasificare și regresie există o serie de tehnici, iar pentru fiecare dintre acestea sunt disponibili mai mulți algoritmi. Diferența dintre clasificare și regresie este aceea că în primul caz, outputul previzionat aparține unei anumite clase, în timp ce în al doilea caz, outputul estimează valoarea unui atribut. Regresia este utilizată în cazurile în care outputul este definit pe un domeniu foarte larg (chiar infinit) de valori și nu trebuie confundată cu noțiunea de regresie liniară din matematică. De remarcat este faptul că o problemă de regresie poate fi ușor transformată într-una de clasificare și invers. În multe cazuri, mijloacele software pot fi utilizate pentru rezolvarea ambelor tipuri de probleme.

Procesul de asociere sau clustering duce la descoperirea unor asemenea reguli, la diferite niveluri de abstractizare, între obiectele unei baze de date. De exemplu, se poate afla că scăderea vânzărilor este totdeauna însoțită de o simptomatologie definită de evenimente din interiorul sau din afara organizației, și pe această bază se pot studia mai departe motivele care stau în spatele acestei asocieri (analiză de tip cauză-efect).

2.2.2.2. Analiza secvențială sau a succesiunilor

Analiza secvențială sau a succesiunilor duce la descoperirea acelor evenimente care se petrec întotdeauna într-o anumită secvență. Ea studiază datele care apar în tranzacții diferite (spre deosebire de asociație, care studiază relațiile dintre datele ce apar în cadrul aceleiași tranzacții). De exemplu, în urma analizei secvențiale s-ar putea descoperi că majoritatea clienților, care în decursul unei săptămâni cumpără produsele A și B, săptămâna viitoare cumpără produsele C și D.

Analiza asocierilor și succesiunilor, denumită uneori *analiza coșului de cumpărături (basket data)*, generează modele descriptive care evidențiază reguli de corelație între atributele unui set de date.

2.2.2.3. Tehnica rețelelor neuronale

Tehnica *rețelelor neuronale* are la bază două concepte aparținând domeniului inteligenței artificiale. *Neuronul artificial reprezintă unitatea de bază pentru prelucrarea informației în cadrul calculului neuronal*. Prin analogie cu neuronul biologic, el a fost definit ca o unitate ce procesează inputuri informaționale și generează outputuri. Rețeaua neuronală artificială reprezintă un ansamblu de neuroni artificiali, legați prin conexiuni. *Rețelele neuronale sunt sisteme dinamice, al căror comportament poate fi caracterizat prin urmărirea stărilor la momente diferite de timp*. Starea unei rețele la un moment dat este definită de ansamblul nivelurilor de activare a neuronilor și de intensitățile conexiunilor dintre neuroni. În plus față de acești parametri ajustabili, o rețea este definită și de următorii parametri ficși: configurația conexiunilor și tipul funcțiilor de activare.

Rețelele neuronale utilizează un număr mare de parametri pentru a construi un model, care preia și combină un set de inputuri în scopul prognozării comportamentului unei variabile cantitative sau categoriale. În mod generic, valoarea fiecărui nod al rețelei se calculează ca medie ponderată a valorilor nodurilor anterioare. Procesul de construire a modelului implică aflarea acelor ponderi care produc cele mai precise prognoze, utilizând date reale pentru a *antrena* rețeaua. Cea mai comună metodă de antrenare este compararea datelor calculate în rețea cu valorile corecte cunoscute. După fiecare comparare, ponderile sunt ajustate și valorile sunt calculate din nou. După o perioadă îndelungată de astfel de comparații și ajustări, o rețea neuronală devine de regulă un foarte bun predictor.

O rețea neuronală (NN) este un graf orientat $F = \langle V, A \rangle$, care are mulțimea nodurilor $V = \{1, 2, \dots, n\}$ și mulțimea arcelor $A = \{\langle i, j \rangle \mid 1 \leq i, j \leq n\}$, cu următoarele restricții [66]:

1. Mulțimea V este împărțită în submulțimile V_I (noduri de intrare), V_H (noduri ascunse) și V_O (noduri de ieșire)
2. mulțimea arcelor $\{1, 2, \dots, k\}$ este împărțită în niveluri:
 - a. nodurile de intrare pe nivelul 1,
 - b. nodurile ascunse pe nivelurile 2, ..., $k-1$
 - c. nodurile de ieșire pe nivelul k ;
3. orice arc $\langle i, j \rangle$ trebuie să aibă nodul i pe nivelul $h-1$ și nodul j pe nivelul h
4. arcul $\langle i, j \rangle$ primește o etichetă - valoarea numerică w_i
5. nodul i primește o etichetă - funcția f .

În Fig. 2.19 se prezintă un exemplu de rețea neuronală, unde se observă că nivelul nodurilor de intrare ($I1, I2, I3$) este conectat la nivelul nodurilor ascunse ($H1, H2$), care este conectat la nivelul nodurilor de ieșire ($O1$). Ponderea arcelor ($w1, \dots, w8$) reflectă importanța nodurilor legate la intrări [28].

Analiza *limitelor tehnicii* a identificat faptul că rețelele neuronale nu operează decât direct asupra variabilelor numerice. Drept urmare, orice variabilă non-numerică din setul de date care se dorește a fi analizat, va trebui convertită în variabilă numerică înainte de utilizarea sa în instruirea rețelei. În cazul problemelor complexe, utilizatorul este pus în situația de a rezolva un compromis, între a crește numărul de neuroni ascunși, ceea ce poate conduce la o instruire foarte lentă și a accepta o topologie mai simplă, asociată unei soluții mai puțin precise.

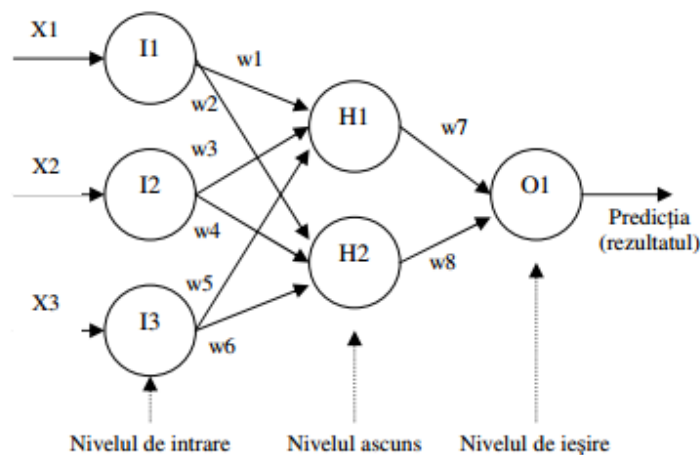


Fig. 2.19. Exemplul unei rețele neuronale cu 6 noduri [132]

Pentru seturi de date cu număr mare de atribute, folosirea rețelelor neuronale devine ne-fezabilă. Determinarea numărului de neuroni ascunși, pentru probleme complexe de clasificare, nu se poate face decât experimental, ceea ce pe de-o parte crește substanțial timpul alocat căutării modelului optim de clasificare, iar pe de altă parte lasă calitatea rezultatelor analizei să depindă de nivelul de experiență al utilizatorului.

Absența componentei descriptive într-un model generat de o rețea neuronală face ca evoluția modelului în etapa de instruire să fie lipsită de transparență pentru utilizator. Datorita acestei caracteristici, tehnica este deseori comparată cu o „cutie neagră”. Totuși, din punct de vedere practic, cel mai mare dezavantaj al rețelelor neuronale este timpul îndelungat necesar pentru o bună instruire, fapt corelat cu necesitatea existenței unui număr relativ mare de instanțe în setul de instruire.

Analiza *avantajelor tehnicii* au identificat următoarele aspecte:

- Rețeaua odată instruită poate realiza predicții rapide pentru instanțe noi. Aceasta caracteristică face ca rețelele neuronale să fie utilizate cu succes în probleme care necesită răspuns în timp real.
- Până în prezent, rețelele neuronale reprezintă metoda cea mai eficientă de modelare a unor relații neliniare.
- Mai mult, aplicațiile de până acum au demonstrat aplicabilitatea acestei tehnici în domenii dificil de modelat, precum vederea electronică sau recunoaștere vocală.

Spre deosebire de celelalte tehnici de *Data Mining*, rețelele neuronale nu restricționează output-ul la un singur atribut. Folosind o arhitectură de rețea potrivită se pot obține predicții simultane pentru mai multe variabile, ceea ce poate însemna o eficientizare semnificativă a proceselor de explorare a datelor.

2.2.2.4. Tehnica arborilor de decizie

Arborii de decizie reprezintă o serie de reguli a căror aplicare conduce către o anumită clasă sau valoare din baza de date. Această tehnică este folosită deseori, deoarece are un grad de precizie rezonabil și, spre deosebire de rețelele neuronale, este mai ușor de înțeles și necesită un timp de construire mai redus. Ar mai trebui precizat că atât arborii de decizie, cât și rețelele neuronale, pot fi utilizate pentru construirea unor modele de regresie. Arborii de decizie pot fi construiți fie de sus în jos, fie de la stânga la dreapta. Nodul de sus respectiv din marginea stângă este nodul rădăcină. Nodurile descendente sunt noduri fii, iar nodurile de jos respectiv din marginea dreaptă sunt noduri terminale. Calea unică de la nodul rădăcină la un nod terminal formează o regulă (reprezentare cu linie punctată în Fig. 2.20). Un arbore de decizie este o structură în care [37]:

- Nodurile interne indică un test pe un atribut;
- Ramurile reprezintă rezultatul unui test;
- Nodurile frunză reprezintă etichete ale claselor.

Arborele de decizie este o tehnică de explorare a datelor cu potențial atât predictiv, cât și descriptiv. Denumirea sa provine de la aceea că rezultatul se prezintă utilizatorului sub forma unui *graf de tip arbore*. Output-ul major, al unui model bazat pe arbori decizionali, este arborele în sine. Procesul de instruire care creează arborele este numit inducție. Aceasta presupune, ca și în cazul rețelelor neuronale, parcurgerea de câteva ori a setului de date de instruire, cu deosebirea că în cazul arborilor, timpul de instruire și implicit numărul de baleieri ale setului de date este mult mai mic decât la rețelele neuronale. Mai precis, numărul de parcurgeri ale setului de instruire este egal cu numărul de niveluri în arbore.

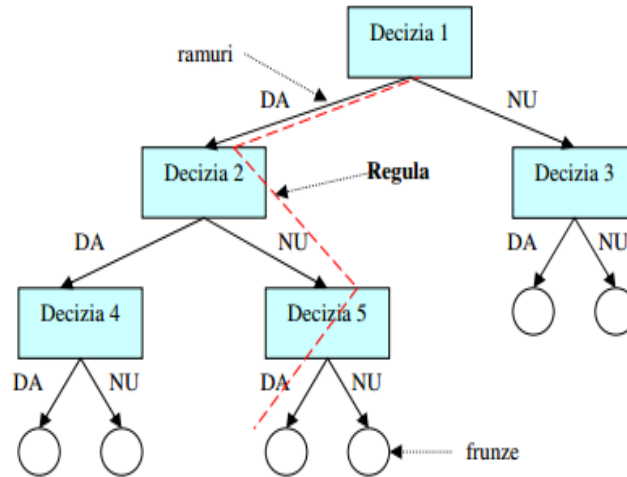


Fig. 2.20. Exemplul unui arbore de decizie [132]

Între limitele acestei tehnici se remarcă faptul că majoritatea algoritmilor nu folosesc întregul set de date indicat de utilizator pentru inducție. Pentru acești algoritmi, construirea arborelui presupune transferul instanțelor din setul de date de instruire în memoria RAM. Dimensiunea limitată a memoriei face ca programul să transfere în RAM numai un subset de date, selectat aleator. În consecință, gradul de reprezentativitate al modelului construit este determinat de capacitatea aplicației de a selecta un subset reprezentativ pentru întreg setul de inducție.

O critică adusă frecvent arborilor decizionali este aceea că algoritmi de inducție nu iau în considerare, la momentul separării, divizării sau splitării (din engleză, split = a diviza), efectul pe care respectiva separare o are asupra viitoarelor splitări. În plus, toate separările se fac secvențial, ceea ce determină dependența fiecărei splitări de cele precedente. Între *avantajele tehnicii* se poate menționa că majoritatea algoritmilor care construiesc arbori decizionali pot fi aplicați fără restricții legate de tipul datelor. Deși variabila dependentă trebuie să fie de natură numerică (în cazul problemelor de regresie) sau categorică (în cazul problemelor de clasificare), pentru majoritatea algoritmilor, variabilele independente pot lua valori în orice domeniu. Tehnica se caracterizează prin capacitate de prelucrare a unor seturi de date cu număr mare de atribute.

Există situații în care o instanță poate fi descrisă printr-un număr relativ mare de atribute, de ordinul sutelor sau chiar miilor. În astfel de situații, explorarea prin tehnica arborilor decizionali reprezintă singura alternativă, cei mai mulți algoritmi fiind capabili să trateze seturi de date cu peste 1000 de coloane.

Algoritmii de construire a arborilor decizionali necesită un număr redus de parcurgeri a setului de date utilizat în inducție. Consecința directă a acestei caracteristici funcționale este rapiditatea procesului de inducție și aplicarea eficientă asupra seturilor mari de date. Forma outputului permite nu numai realizarea de previziuni și clasificări, ci și descrierea relațiilor existente între variabilele independente și variabila dependentă. În plus, forma grafică a outputului facilitează analiza relațiilor. În practica organizațională, există aplicații care permit reprezentarea arborelui sub forma unui set de reguli care, pentru arbori de dimensiuni mari, este mai ușor de înțeles.

2.2.2.5. Tehnica Naive Bayes

Tehnica Naive Bayes este mai puțin implementată în aplicațiile de Data Mining; este o metodă de clasificare care își datorează numele pastorului presbiterian britanic Thomas Bayes (1702- 1761). Teoria probabilităților a lui Bayes, pe care se bazează și tehnica ce-i poartă numele, a fost publicată postum, abia în 1764.

Astfel, Naive Bayes este o tehnică de clasificare cu potențial atât predictiv, cât și descriptiv. Ea permite analiza relației dintre fiecare variabilă independentă și variabila dependentă, prin calcularea unei probabilități condiționate pentru fiecare din aceste relații. Când o nouă instanță se dorește a fi clasificată, predicția se realizează prin combinarea efectelor variabilelor independente asupra variabilei dependente.

Limitele tehnicii rezidă în faptul că pentru instanțele care aparțin setului de date utilizat la calculul probabilităților a priori și al celor condiționale, „predicția” atributului - obiectiv este 100% corectă, însă pentru instanțe din afara setului de date de instruire, eficiența algoritmului este puternic afectată de prezența unor probabilități condiționate, egale sau foarte aproape de zero.

O altă limită a algoritmului provine din prezumția că între atributele independente din setul de date există (teoretic) o independență statistică. Această prezumție stă și la originea adjectivului „naiv” din denumirea algoritmului, având în vedere că independența statistică, de regulă, nu se verifică și în practică.

Algoritmul este limitat din punct de vedere al inputului la date booleene sau categorice. Dincolo de efortul de preprocesare necesar pentru transformarea datelor cu caracter continuu în intervale valorice, așa cum s-a mai menționat, operațiunea este de multe ori dependentă de experiența și chiar intuiția analistului, factori subiectivi care marchează rezultatele explorării.

Ținând seama de faptul că pentru calculul probabilităților nu este nevoie decât de o singură parcurgere a setului de date, algoritmul prezintă avantajul important al unei viteze mari de construire a modelului de clasificare. Un alt avantaj semnificativ este acela că algoritmul prezintă capacitatea de a realiza predicții din informații parțiale. În ciuda sensibilității la caracteristici slab reprezentate în setul de date, pentru realizarea unei predicții, algoritmul nu are obligatoriu nevoie de toate atributele independente, astfel încât cele identificate de analist a fi irelevante pot fi ușor eliminate din algoritm. În fapt, chiar dacă nu s-ar cunoaște nimic despre atributele independente, analistul tot ar putea face o predicție (fără pretenția de a fi foarte precisă) numai pe baza probabilităților a priori.

Modelul obținut prin aplicarea algoritmului are și un conținut descriptiv, care poate fi util analistului. Probabilitățile condiționate aferente fiecărui atribut independent pot fi utilizate în a descrie legătura dintre acestea și atributul - obiectiv.

2.2.2.6. Tehnica k-NN

Tehnica k-NN (prescurtare a expresiei engleze k-Nearest Neighbor, tehnica bazată pe cei mai „aproși vecini”) este una predictivă, de explorare a datelor și care este utilizată cu precădere în probleme de clasificare. Principiul care stă la baza tehnicii este relativ simplu: o instanță nouă este clasificată prin analiza „proximității” sale (sau gradului de similitudine) cu alte instanțe dintr-un set de date cunoscut.

k-NN este o tehnică folosită în special pentru clasificarea datelor în categorii multiple, însă poate fi aplicată inclusiv pentru previzionarea unui atribut - obiectiv de natură numerică (continuă sau discretă), ca rezultat al unor dependențe neliniare. Fie un set de date compus din instanțe care au următoarea structură:

- n atribute numerice independente $\{X_i, i = 1, n\}$;

- m atribute booleene sau categorice independente $\{A_j, j = 1, m\}$;
- un atribut - obiectiv Y , reprezentând variabila dependentă a cărei valoare va trebui estimată pentru noile instanțe.

Pentru a previziona valoarea atributului - obiectiv al unei instanțe noi, algoritmul caută în setul de date k înregistrări „apropiate” de acea instanță, pentru care se cunosc valorile lui Y . Predicția este dată de media valorilor lui Y aferente „vecinilor” identificați, în setul de date. Aplicarea conceptului în practică ridică următoarele probleme:

- Prin ce metodă se stabilește relația de vecinătate dintre două instanțe? Care este metrica utilizată la calculul distanței dintre două instanțe?
- Care este valoarea optimă pentru k ? De câte instanțe similare este nevoie pentru ca media atributelor - obiectiv să se constituie într-o predicție cu grad acceptabil de reprezentativitate?
- Ce metodă de calcul a mediei va fi utilizată pentru predicția lui Y ? Întrebarea este importantă în special pentru cazurile în care atributul - obiectiv nu este de natură numerică, ci logică sau categorială.
- Care din atributele $\{X_i\}$ și $\{A_j\}$ sunt cu adevărat reprezentative pentru predicție? Care din aceste atribute identifică cel mai bine „vecinii” instanței de analizat?

Una dintre *limitele tehnicii* se datorează timpului de calcul, care este direct proporțional cu numărul de instanțe din setul de date. Din acest motiv, pentru seturi mari de date se impune ca în etapa de preprocesare, din setul inițial de date să se selecteze un subset de instanțe cu dimensiuni rezonabile.

Algoritmul lucrează eficient în probleme de clasificare atunci când toate clasele aferente atributului - obiectiv au o reprezentare egală ca pondere în setul de date, fapt care face necesară „îmbogățirea” setului de date original. Algoritmul pe care se bazează tehnica k -NN permite doar realizarea unei estimări a valorii atributului - obiectiv, fără a produce informații suplimentare despre instanța supusă analizei, despre structura setului de date ori despre categoriile de clasificare a atributului - obiectiv.

De cele mai multe ori este dificil de stabilit ce tip de funcție estimează cel mai bine distanța dintre două instanțe. Deși din punct de vedere matematic tehnica permite calculul distanțelor și pentru atribute categoriale și logice, în astfel de cazuri metrica devine puternic influențată de transformările aplicate de analist setului de date în preprocesare. De aceea, k -NN este de preferat a fi utilizată mai mult în situațiile în care pentru toate atributele instanțelor se poate aplica aceeași funcție de distanță.

Între *avantajele tehnicii* se remarcă faptul că aceasta permite clasificarea în multiple clase și modelarea relațiilor neliniare dintre date (în probleme de predicție). Pentru tehnicile care necesită o etapă de învățare a cărei output îl constituie un model predictiv, există riscul ca acest model să devină desuet în timp, iar predicțiile realizate în baza lui să piardă din reprezentativitate.

În cazul k -NN, modelul îl constituie chiar setul de date, care se presupune că odată supus analizei, este deja în forma sa cea mai recentă. Chiar dacă de multe ori pot apărea dificultăți în stabilirea unei metrici eficiente, algoritmul este unul dintre puținele care acceptă ca input, date de natură diferită (continuă, categorială, logică etc.).

2.2.2.7. Comparație între tehnicile de Data Mining

În funcție de cazul concret, anumite tehnici de *Data Mining* sunt mai eficiente decât altele, existând chiar situații în care pentru rezolvarea problemei nu există decât

o unică opțiune (de exemplu, arborii decizionali sunt singura alternativă viabilă pentru analiza seturilor de date cu număr mare de variabile, rețelele neuronale reprezintă unica soluție pentru probleme în care outputul are o formă vectorială etc.). Se impune, deci, o analiză comparativă a acestor tehnici sau algoritmi de *Data Mining*.

Tabelul 2.1 oferă rezultatele unei analize și sinteze a tehnicilor de *Data Mining*, luându-se în considerare 12 caracteristici, care au fost utilizate și în prezentarea anterioară, pentru fiecare din cele patru tehnici de *Data Mining* tratate. S-a notat cu semnul „+” situația în care tehnica analizată de *Data Mining* satisface criteriul curent, iar cu „-” situația contrarie. De departe, arborii decizionali prezintă cele mai multe avantaje, în timp ce rețelele neuronale prezintă gradul cel mai mic de flexibilitate.

În practica organizațiilor, pe măsură ce procesul de *Data Mining* devine o practică obișnuită, rețelele neuronale și arborii de decizie se bucură de tot mai multă apreciere. Deși rețelele neuronale sunt mai complexe în felul lor, utilizarea lor nu cere cunoștințe numeroase și avansate.

Tabelul 2.1. Analiza comparativă a celor patru tehnici de Data Mining

Criteriu de comparație	Tehnica			
	Rețele neuronale	Arbori de decizie	Bayes	k-NN
1. Rapiditate, operativitate în etapa de instruire	-	+	+	-
2. Rapiditate, operativitate în aplicarea modelului	+	+	+	+
3. Instruire eficientă pe seturi largi de date	-	+	+	-
4. Operare eficientă pe seturi de date cu număr mare de atribute	-	+	-	-
5. Capacitate de generare a unor outputuri complexe (mai multe atribute simultan)	+	-	-	-
6. Capacitate de generare a unor outputuri de natură vizuală	-	+	-	-
7. Outputuri cu potențial descriptiv	-	+	+	-
8. Utilizare în probleme de predicție	+	+	-	+
9. Utilizare în probleme de clasificare	+	+	+	+
10. Nu comportă restricții legate de tipul datelor de input	-	+	-	+
11. Soluția (modelul) nu depinde de experiența utilizatorului	-	+	+	+
12. Transparența modelului față de utilizator	-	+	+	-
Scor total	+4 / -8	+11 / -1	+7 / -5	+5 / -7

2.3. Concluzii

Process Mining și *Data Mining* sunt complementare, astfel încât folosirea acestora împreună duce la rezultate benefice în conducerea afacerilor. Avantajul tehnicilor *Process Mining* constă în faptul că ele combină punctele forte ale *Data Mining* și ale modelării de proces specifice BPM (Fig. 2.21). Modelele de proces se creează automat prin intermediul datelor obținute în timp real din sistemele informatice, obținându-se astfel instant imaginea procesului, ceea ce ajută managementul în luarea deciziilor. De asemenea, tehnicile *Data Mining* găsesc diferite tipare în datele analizate (reguli, arbori de decizie, clusterare ș.a.), date utile managementului.

În acest capitol au fost prezentate două domenii ale cunoașterii, care contribuie substanțial la realizarea demersului de *Process Mining*. Astfel, pe baza cercetărilor asupra referențialului bibliografic s-au realizat sinteze și analize asupra:

1. Modelării proceselor de afaceri;
2. Principalelor tehnici utilizate în *Data Mining* (ca parte integrantă a KDD) și rolul acestora în descoperirea de cunoștințe.

Așa cum s-a arătat, implementarea BPM în organizații presupune modelarea grafică a proceselor, ce poate fi realizată prin diferite moduri (capitolul 2.1), după o documentare minuțioasă a acestora (activitate realizată în grupuri eterogene de specialiști din domeniul managementului organizației și IT). Cea mai folosită metodă de reprezentare grafică a proceselor este BPMN 2.0 (secțiunea 2.1.6), metodologia *CertiBPM* făcând referire doar la aceasta. Însă pentru a aplica tehnici specifice *Process Mining* asupra proceselor modelate în BPMN, este nevoie de modele matematice abstracte, iar diagramele BPMN nu sunt potrivite în acest sens. Rețelele Petri și rețelele cauzale tip C-nets sunt cele mai indicate pentru tehnicile de *Process Mining*, cu mențiunea că rețelele Flux de Lucru sunt rețele Petri special concepute pentru procese de afaceri. *Descoperirea* de procese și *verificarea conformității* lor operează cu modele de proces reprezentate prin rețele Petri ce se pot converti în format BPMN, și invers. În funcție de contextul analizat, un analist de proces trebuie să aleagă reprezentarea de model cea mai potrivită.

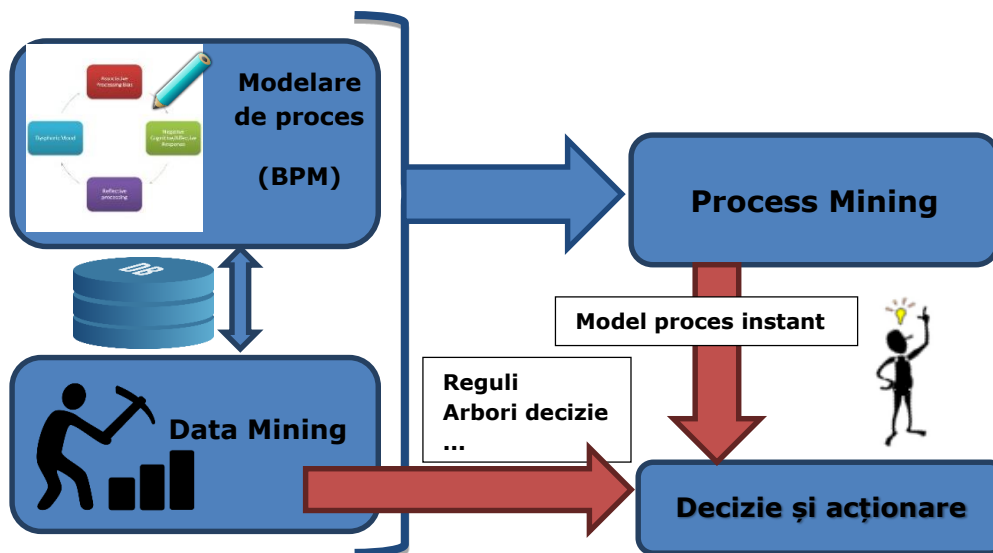


Fig. 2.21. Luarea deciziilor cu ajutorul *Process Mining* și *Data Mining*

Diagramele YAWL și EPC sunt și ele folosite în modelare, dar au o răspândire mai redusă decât diagramele BPMN. Pentru un analist de proces este foarte important să cunoască diferențele dintre aceste modele, precum și avantajele și dezavantajele fiecărei categorii în parte.

Procesele de afaceri și activitățile aferente acestora generează și sunt confruntate cu un volum impresionant de date (virtual, fiecare companie își înregistrează toate tranzacțiile operate), informații, dar care au determinat existența unor depozite, baze de date sau arhive de dimensiuni mari (care ocupă spații, reale sau virtuale, extinse pentru realizare) și care aparent nu au utilitate decât pentru raportări, evaluări, diagnoze, bilanțuri de activități. Pe de altă parte, managerii au conștientizat că o mare parte a cunoașterii acumulate de organizații de-a lungul ciclului lor de viață și oglindită în înțelepciunea lor dobândește atributul de **active intangibile**, sau mai precis **cunoaștere tacită**, având un potențial important de utilizare în susținerea deciziilor strategice. Ca urmare, în procesul de analiză, informația ce poate fi identificată, explorată și extrasă (ceea ce definește demersul de *Data Mining*) din depozite, sisteme de baze de date (date, informații, cunoaștere din trecut) își poate găsi utilitatea în prezent:

- pentru construirea unor modele de previziune și/sau prognoză a proceselor de afaceri;
- la identificarea unor legături sau relații între informațiile și cunoștințele (înregistrări) stocate în depozite, baze de date;
- în clasificarea sau identificarea unor ierarhii cu privire la înregistrările existente;
- pentru realizarea unei descrieri privind informațiile și cunoștințele (înregistrări) stocate în depozite, baze de date.

Toate aceste acțiuni fac apanajul tehnicilor de *Data Mining*, care permit extragerea datelor, informațiilor și realizarea de previziuni pornind de la date istorice ale organizațiilor, generând informații și cunoștințe noi, capabile să susțină mai bine decizia managerială, în general, și pe cea strategică, în special.

Analiza anumitor tehnici de *Data Mining* (capitolul 2.2.2), fără a cuprinde însă toate aspectele și tehnicile prezente în literatura de specialitate, ci doar a celor relevante pentru cercetarea de față, a scos la iveală limite ale fiecăreia, dar și avantaje în aplicare. Sinteza acestor puncte tari și slabe, realizată în cazul a patru tehnici de *Data Mining*, a demonstrat superioritatea arborilor de decizie, care se bucură de o mai mare răspândire și implementare în diferite organizații, pentru aplicații în domenii cum sunt: marketing, vânzări, servicii financiare, comportamentul consumatorilor, analiza riscurilor de investiții, prognoze etc.

Tehnicile de *Data Mining* au cel mai mare grad de utilitate în Miningul punctelor de decizie (secțiunea 3.7.1) și în filtrări de jurnale de evenimente (de exemplu, ca punct de plecare pentru clusterizarea din Metodologia PDM-Healthcare – secțiunea 5.2.2).

3. CERCETĂRI TEORETICE PRIVIND PROBLEMATICA REALIZĂRII PROIECTELOR DE PROCESS MINING

În cadrul prezentului capitol sunt prezentate analiza și sinteza cunoștințelor ce susțin implementarea demersului de *Process Mining*, prin intermediul unui proiect specific. Astfel, s-au creat premisele de cunoaștere pentru fundamentarea metodologiei propuse și adoptate în vederea derulării proiectelor de *Process Mining* în organizații.

Obiectivul operațional aferent cercetărilor din acest capitol a fost: **OP3** Cercetări teoretice pentru identificarea elementelor cheie care fac parte dintr-un proiect de *Process Mining*.

3.1. Descrierea generală a proiectelor de Process Mining

3.1.1. Process Mining

Tehnicile *Process Mining* folosesc date din evenimente pentru a *descoperi* modele de proces, pentru a *verifica conformitatea* modelelor predefinite și pentru a *extinde* aceste modele cu informații despre gâtuiuri (*bottlenecks*), decizii și folosirea resurselor. Aceste tehnici operează pe evenimente observate, iar nu pe modele manuale. Modelele sunt învățate și îmbogățite cu informații din jurnale de evenimente.

Tehnicile de *Process Mining* extrag cunoștințe și modele din jurnale de evenimente. Pe lângă acestea, folosind tehnici specifice din sfera *Process Mining*, procesul de afaceri poate fi monitorizat sau îmbunătățit. Fiecare eveniment se referă la o activitate a procesului. Acesta stochează informații despre numele activității, operatorul care a finalizat activitatea, momentul în care a avut loc sau orice alte informații referitoare la activitatea respectivă.

Totalitatea evenimentelor convergente pe execuția unui proces formează un caz (o instanță de proces). Un caz se referă la o singură execuție a procesului.

Domeniul de *Process Mining* oferă o serie de tehnici care analizează modelele de procese ascunse în „amprentele” lăsate de către utilizatori în cadrul sistemelor informatice. Aceste „urme” sunt numite jurnale de evenimente, iar sistemele informatice capabile să genereze jurnale prin integrarea aplicațiilor, resurselor și proceselor sunt numite Sisteme Informatice pentru Conștientizarea Proceselor (*Process-Aware Information System, PAIS*).

Practica prezentului a evidențiat trei tipuri de *Process Mining*: *descoperire*, *conformitate* și *îmbunătățire*. Tehnicile de descoperire se referă la extragerea modelelor de procese din jurnalele de evenimente. Pe lângă descoperire, *Process Mining* analizează abaterile între „comportamentul” dorit al unui proces și cel real al acestuia, astfel realizându-se analize de conformitate. Ultima categorie de *Process Mining* se referă la procesul de îmbunătățire pe baza informațiilor existente în jurnalele de evenimente.

3.1.2. Relația dintre modelul de proces și jurnalul de evenimente

Între jurnalul de evenimente, ce reprezintă „realitatea” și modelul de proces corespunzător trebuie stabilită o relație. Relația a fost definită prin termenii *Interpretare-internă (Play-in)*, *Interpretare-externă(Play-out)* și *Reluare (Replay)* [8] și este reprezentată schematic în Fig. 3.1. Semnificația termenilor este următoarea:

- *Interpretarea-internă* preia ca date de intrare un jurnal de evenimente și generează din acesta un model de proces. Această activitate este specifică *descoperirii*.
- *Interpretarea-externă* preia ca date de intrare un model de proces și generează comportament. Această activitate este specifică analizei și simulării.
- *Reluarea* preia ca date de intrare atât un model de proces, cât și un jurnal de evenimente. Scopul acestei activități este: *verificarea conformității, îmbunătățirea* modelului cu adăugarea de diferite perspective (pentru miningul punctelor de decizie, pentru miningul organizațional ș.a., *suport operațional, construirea de modele predictive*.

3.2. Imaginea de ansamblu a unui proiect de Process Mining

Un proiect de *Process Mining* se constituie ca o provocare la nivelul managementului organizației, dar și în ansamblul ei. Etapele sale sunt:

1. **Obținerea unui jurnal de evenimente** care este acceptat în format XES. Pentru realizarea sa, anumite surse de date trebuie integrate, ceea ce va conduce la definirea jurnalului. Datele pot fi din aceeași bază de date sau din baze de date diferite, pe servere de mail, forumuri, wiki-uri ș.a.

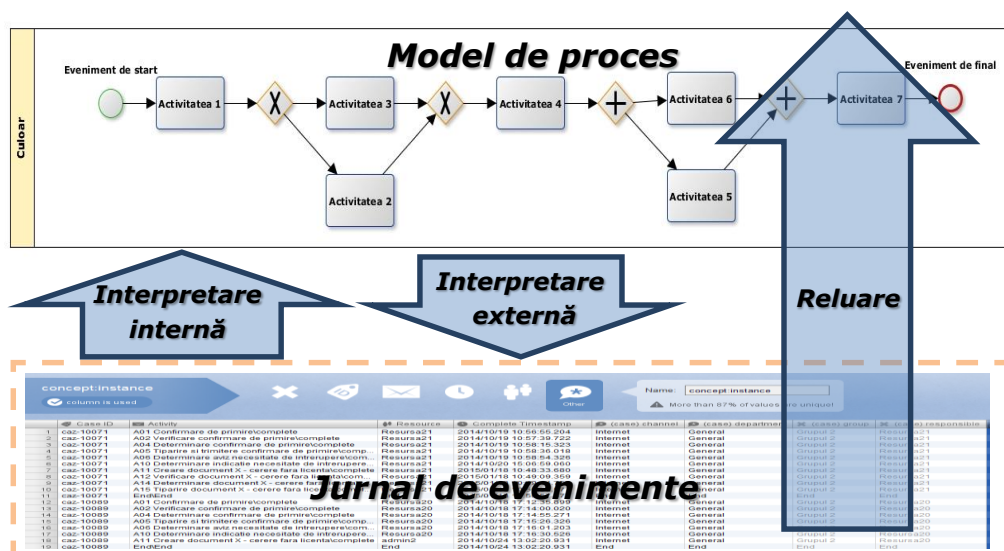


Fig. 3.1. Relația dintre modelul de proces și jurnalul de evenimente

2. **Crearea sau descoperirea modelului de proces:** jurnalul de evenimente este supus câtorva algoritmi de descoperire și, astfel, se obține o rețea Petri. Printre cei mai utilizați algoritmi în această etapă sunt: algoritmul α , algoritmul ILP, algoritmul HeuristicsMiner (rețeaua Petri se generează prin conversia rețelei euristice obținute), algoritmul Inductive Miner ș.a. Totodată, sunt acceptate și modele realizate manual.
3. **Conectarea evenimentelor din jurnal cu activități din model** este etapa în care are loc verificarea conformității modelului descoperit/creat cu informațiile din jurnalul de evenimente (validarea modelului). În această etapă se observă diferențele dintre comportarea modelului și jurnal și are loc o aliniere a celor două.
4. **Extinderea modelului prin adăugarea de perspective** organizaționale, temporale, cazuistice ș.a. În această etapă are loc îmbunătățirea modelului cu adăugare de date despre diferite perspective. Dacă modelele de proces ar fi o hartă cu drumuri, acest pas s-ar putea corela cu adăugarea punctelor de interes. Din punct de vedere organizațional se pot crea rețele sociale pentru a se observa anumite detalii: împărțirea activităților între participanți, descoperirea rolurilor participanților în funcție de tipare, descoperirea grupurilor din care participanții fac parte. Din punct de vedere temporal se pot observa timpii de așteptare, timpii de activitate, iar din analiza acestora pot rezulta anumite gâtuiuri (*bottleneck*) ale procesului. Din punct de vedere cazuistic se poate determina calea ce va fi urmată de derularea unui caz, în funcție de anumite decizii din proces. Se mai pot adăuga și alte perspective, ca de exemplu costuri asociate activităților.
5. **Obținerea modelului final**, care oferă o perspectivă holistică asupra întregului proces. Aceasta poate genera noi idei de îmbunătățire. Modelul astfel obținut poate fi folosit ca sursă pentru diagnosticare, re-inginerie, suport operațional. Trebuie conștientizate limitările (sub- sau supra-potrivirea), ele rezultând din ponderarea celor patru criterii de calitate: generalizarea, simplitatea, precizia și potrivirea [41]. Altă observație importantă, de care trebuie să se țină seama în definirea modelului final, este aceea că modelul obținut tinde să fie prea descriptiv datorită discrepanțelor dintre acesta și jurnal. În acest caz, modelul trebuie să fie îmbunătățit pentru a surprinde realitatea cât mai bine (indicator de calitate al modelului).

Schematic, pașii proiectului de *Process Mining* sunt descriși în Fig. 3.2.

3.3. Pregătirea datelor

Pregătirea datelor este un proces complex care are rolul de a obține jurnalul de evenimente, aceasta fiind prima etapă din proiectul de *Process Mining*. Pregătirea datelor este o activitate de complexitate ridicată, fiind prima dintre provocările demersului de *Process Mining* [13].

3.3.1. Surse de date

Jurnalul de evenimente reprezintă sursa de intrare pentru *Process Mining*, motiv pentru care acest jurnal trebuie să fie de calitate ridicată.

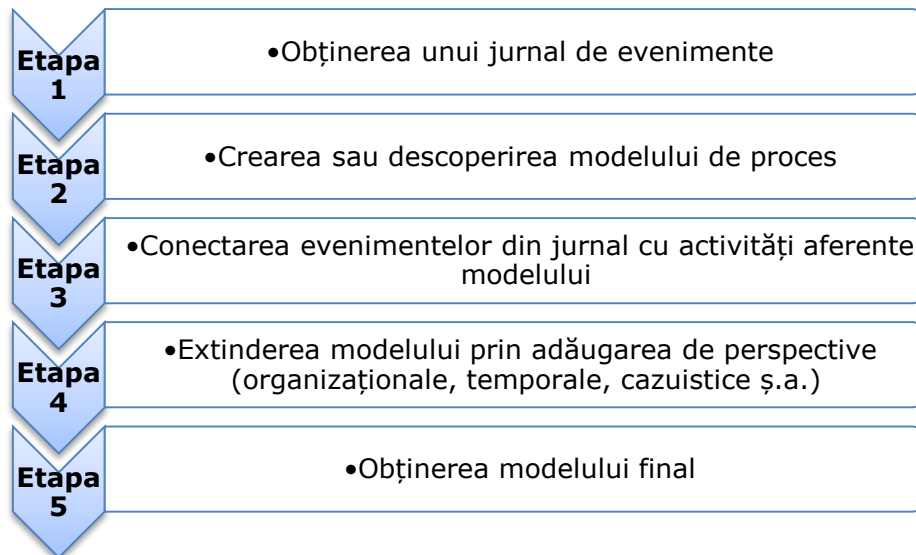


Fig. 3.2. Etapele unui proiect de Process Mining

Provocarea constă în extragerea datelor dintr-o varietate de surse: baze de date, mesaje, emailuri, tranzacții etc. Ca urmare, datele provenite din surse primare sunt brute, „necizelate” și ascunse în diferite surse. Datele pot fi împrăștiate din motive tehnice (de exemplu, datorită existenței unor sisteme de date vechi, incompatibile cu sistemele informatice actuale) sau organizatorice (de exemplu, date sau informații disponibile și utilizate doar la nivel departamental). Totodată, datorită faptului că un proces de afaceri se poate extinde în afara organizației, datele pot fi partajate de mai multe organizații [51]. Adesea serviciile web sunt utilizate ca surse potențiale de date [123].

Sursele de date pot fi structurate și bine descrise de metadate (adică date despre date). Din păcate, în multe situații, datele sunt nestructurate sau metadatele lipsesc.

În Fig. 3.3 se prezintă modul de pregătire a datelor în vederea aplicării demersului de *Process Mining*.

În contextul BI și al demersului de *Data Mining*, expresia Extragere, Transformare, Încărcare (*Extract, Transform, Load, ETL*) [78] este folosită pentru a descrie procesul ce implică:

- (a) extragerea datelor din surse externe,
- (b) transformarea datelor pentru a se potrivi nevoilor operaționale (se ocupă cu probleme sintactice și semantice, asigurând în același timp un nivel predefinit de calitate) și
- (c) de încărcare în sistemul destinație care poate fi un depozit de date sau o bază de date relațională.

Un depozit de date (*Data Warehouse, DW*) este un depozit logic unic, în care sunt stocate datele tranzacționale și operaționale ale unei organizații. Dacă organizația dispune de un depozit de date, atunci probabil că informațiile pentru crearea unui jurnal de evenimente sunt deja existente. Depozitele de date utilizate pentru Procesarea Analitică Online (*Online Analytical Processing, OLAP*) nu conțin foarte multe date legate de proces, pentru că nu le sunt specifice stocarea evenimentelor de afaceri și ordonarea lor.

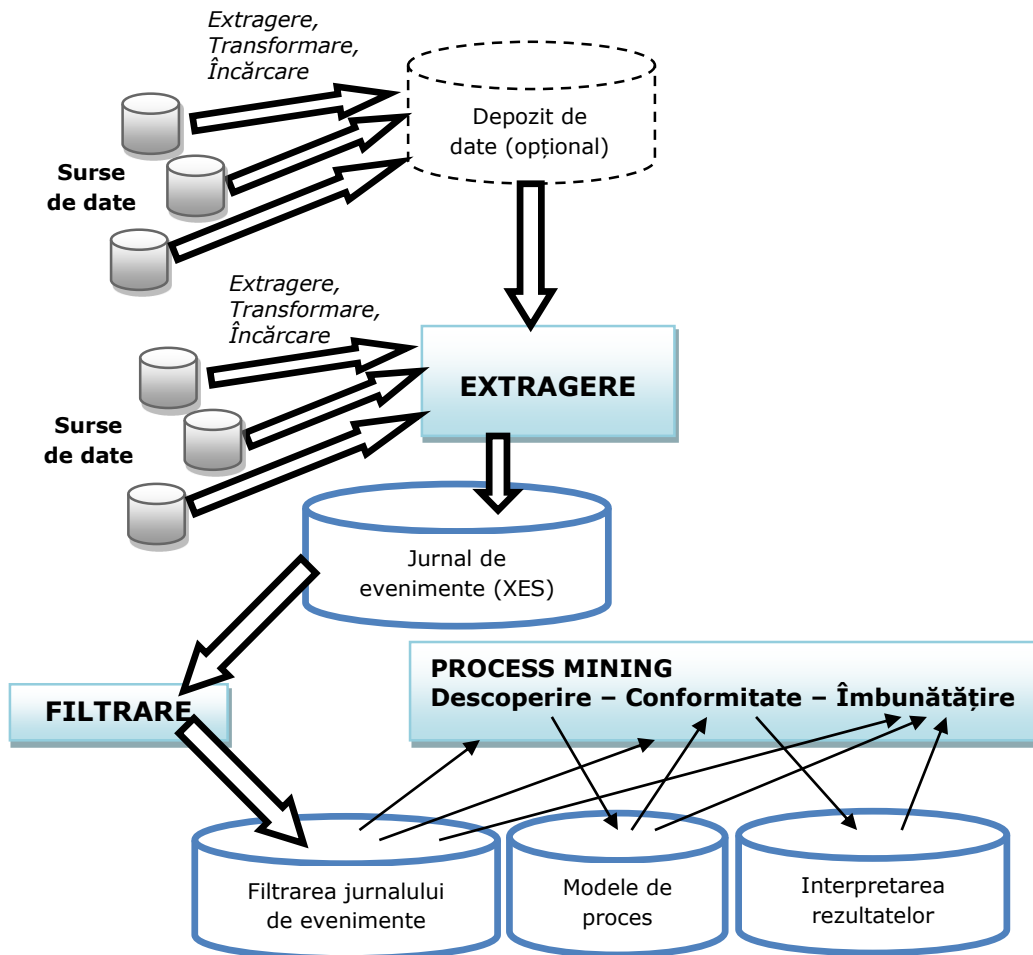


Fig. 3.3. Pregătirea datelor folosite pentru Process Mining

Jurnalele de evenimente trebuie create în funcție de contextul analizat [14]. În analiza inițială trebuie puse întrebări referitoare la context, pentru determinarea elementelor importante ale analizei. De exemplu, dacă se dorește o analiză a procesului de vânzări pe o anumită perioadă/regiune, din toate datele disponibile se vor alege doar acelea din care reiese contextul analizat.

3.3.2. Definirea termenilor folosiți

Pentru a construi un jurnal de evenimente trebuie cunoscute elementele constitutive ale acestuia. Aceste elemente vor fi descrise și definite în continuare.

- **Evenimentul** reprezintă o întâmplare (înregistrată în jurnal) [13], un fenomen instantaneu, având asociată o colecție de atribute și referințe. Atributele au un nume și o valoare. Referințele au nume și un identificator ce se referă la un anumit obiect (persoană, caz, mașină etc.).
- **Atributele evenimentului** pot fi: amprenta de timp, activitatea asociată, resursa asociată (cine execută evenimentul), tipul tranzacției asociate (start, finalizare, suspendare, reluare, întrerupere etc.), costul asociat și altele, în funcție de contextul analizat.
- **Clasificatorul** este o funcție care mapează (cartografiază) atributele unui eveniment pe o etichetă utilizată în modelul de proces rezultat. Acest concept a apărut din faptul că un eveniment este deseori asociat cu activitatea lui, ceea ce nu e corect, deoarece un eveniment are mai multe atribute, așa cum am arătat mai sus. De aceea, clasificatorul reprezintă modul în care este identificat evenimentul. De exemplu, un eveniment poate fi identificat după cuplul activitate și resursă.
- **Secvența** este modul cel mai natural de a prezenta un traseu într-un jurnal de evenimente. La descrierea semantică operațională a rețelelor Petri și a sistemelor de tranziție, comportament modelat este în termeni de secvențe.
- **Traseul** este o secvență finită de evenimente, cu proprietatea că fiecare eveniment din traseu este conținut doar o dată.
- **Cazul** reprezintă un tip de dată ce are o colecție de atribute, dar care are un atribut special obligatoriu și anume traseul. Cazul se mai numește și instanța procesului.
- **Jurnalul de evenimente** reprezintă un set de cazuri, cu proprietatea că fiecare eveniment apare cel mult o dată. Setul reprezintă doar elementele unice dintr-o secvență. Multi-setul reprezintă elementele unice dintr-o secvență, fiecărui element fiindu-i asociată frecvența. Este important ca jurnalul de evenimente să fie ordonat. Dacă jurnalul conține amprente de timp, atunci evenimentele dintr-un traseu trebuie să fie ordonate.
- **Un traseu simplu** este o secvență de activități. Un **jurnal de evenimente simplu** este un multi-set de trasee din mulțimea totală de activități. Aceste notații simplificate au apărut tocmai din faptul că activitatea în cele mai multe cazuri se confundă cu evenimentul, adică activitatea este clasificatorul. Pentru a transforma un jurnal de evenimente într-un jurnal de evenimente simplu, se identifică clasificatorul, iar acesta este aplicat pe secvențe.
- **Activitatea** reprezintă un pas bine definit dintr-un proces [13], mai exact acțiunile ce se întâmplă în acest pas.

3.3.3. Formate suportate

Formatul actual suportat pentru jurnalele de evenimente este XES (eXtensible Event Stream – FEE: Flux Extensibil de Evenimente) adoptat de Grupul Operativ IEEE pentru *Process Mining* [74]. Predecesori formatului XES au fost formatele MXML (Mining eXtensible Markup Language – LMEM: Limbaj de Marcare Extensibil pentru Mining) și SA-MXML (Semantically Annotated Mining eXtensible Markup Language – LMEM-NS: Limbaj de Marcare Extensibil pentru Mining Notat Semantic).

Formatul XES conține elemente de jurnal, traseu și eveniment, exact cum au fost ele definite în secțiunea anterioară (3.3.2). Un document XES este un fișier XML ce conține un jurnal ce poate avea orice număr de trasee. Fiecare traseu descrie o listă secvențială a evenimentelor care corespund unui caz particular. Jurnalul, traseele

sale și evenimentele pot avea orice număr de atribute, iar atributele pot fi imbricate. Există cinci tipuri de bază: String, Data, Int, float, și boolean. Acestea tipuri corespund tipurilor XML standard: *xs:string*, *xs:dateTime*, *xs:long*, *xs:double* și *xs:boolean*. XES nu prevede un set fix de atribute obligatorii pentru fiecare element (jurnal, trasee și evenimente). Un eveniment poate avea oricare număr de atribute. Cu toate acestea, pentru a oferi semantica unor astfel de atribute, jurnalul are referințe către așa-numitele extensii. Utilizatorii pot defini ei înșiși extensii, pentru a surprinde cât mai bine contextul analizat. Există cinci extensii standard:

1. *Concept extension*: definește atributul *name* pentru trasee și evenimente. Pentru trasee, atributul reprezintă identificatorul de caz, iar pentru evenimente, atributul reprezintă numele activității.
2. *Life-cycle extension*: definește atributul *transition* pentru evenimente.
3. *Organizational extension*: definește trei atribute standard pentru evenimente: *resource* (resursa ce a executat evenimentul), *role* (rolul resursei) și *group* (grupul/departamentul din care face parte resursa).
4. *Time extension*: definește atributul *timestamp* pentru evenimente, adică timpul la care s-a desfășurat evenimentul.
5. *Semantic extension*: definește atributul *modelReference* pentru toate elementele din jurnal. Se folosește pentru clasificarea facilă a informațiilor.

Meta-modelul XES¹³ este explicat schematic (diagrame UML) în Fig. 3.4.

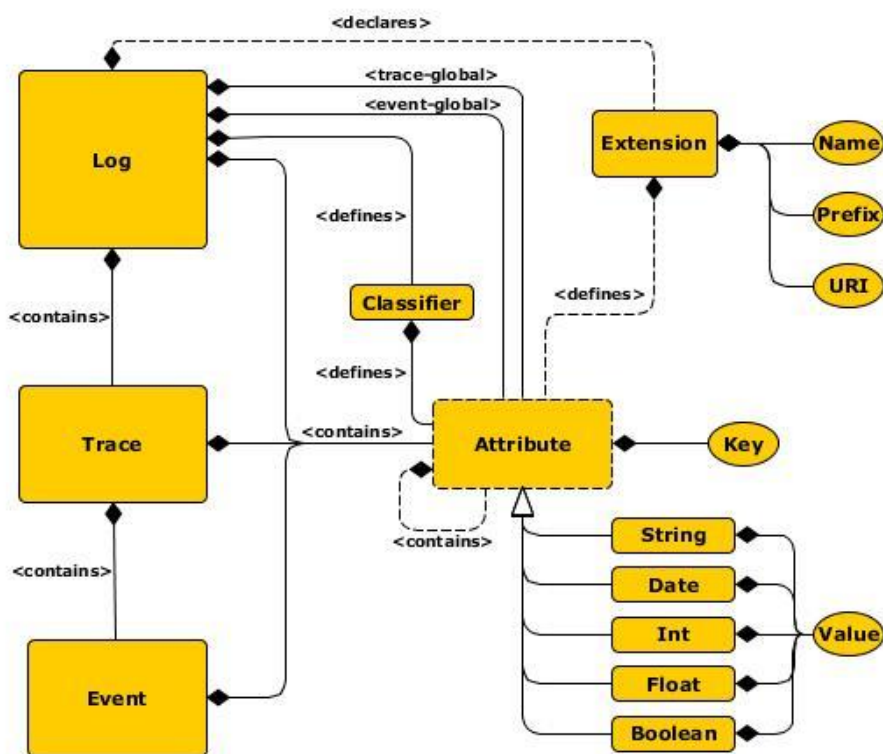


Fig. 3.4. Meta-modelul XES explicat schematic prin diagrame UML [74]

¹³ Preluat din <http://www.xes-standard.org/media/xes/xesstandarddefinition-2.0.pdf>

Calitatea ridicată a jurnalelor de evenimente este o cerință fundamentală pentru *Process Mining*. De aceea extragerea jurnalelor de evenimente este foarte dificilă și interesantă. Provocările întâlnite sunt următoarele:

1. *Corelarea evenimentelor* care se regăsesc în mai multe sisteme informatice. Evenimentele trebuie grupate după caz.
2. *Ordonarea evenimentelor* pe caz după amprenta de timp. De multe ori, datele ce se găsesc în sisteme informatice diferite folosesc „ceasuri interne”, iar o sincronizare a timpului este obligatorie. La unele evenimente se înregistrează doar data, iar timpul nu, de aceea ordonarea este dificilă.
3. *Cazurile nu se încadrează în intervalul de timp definit pentru jurnal*. Cazurile incomplete care încep înainte și se sfârșesc după intervalul de timp trebuie eliminate.
4. *Definirea dimensiunii surselor de date*. Unele baze de date au foarte multe tabele și apare întrebarea care sunt acele tabele folosite din care se poate extrage jurnalul de evenimente?
5. *Existența diferitelor niveluri de detalii din sursele de date*. Unele surse de date au foarte multe informații care nu sunt utile analizei.

O altă problemă o reprezintă *aplatizarea datelor din jurnalele de evenimente*. Datele sursă pentru jurnalele de evenimente se găsesc în numeroase tabele. Între aceste tabele există diferite relații, de tipul chei primare și străine. Aceste tabele sunt reunite, atunci când ele se doresc a fi extrase în jurnale. Această operație creează o aplatizare a datelor, adică din câteva tabele se obține un singur tabel. Aplatizarea datelor într-un jurnal de evenimente poate fi comparată cu agregarea multidimensională în *Online Analytical Processing (OLAP)*, fiind de asemenea o activitate complexă.

Pentru a pregăti datele pentru *Process Mining* se recomandă:

1. *Crearea unui depozit de date* ce conține informații orientate pe proces, despre evenimentele relevante. Depozitul de date ar trebui să evite stocarea datelor agregate, ele trebuind să fie adunate la un loc ca evenimente brute de afaceri. În depozitele tradiționale de date, evenimente sunt agregate în date cantitative, împiedicând astfel analiza de proces.
2. În funcție de contextul analizat, *datele se vor aplatiza* pentru a produce un jurnal de evenimente (de exemplu, în format XES) din perspectiva dorită. De obicei procesul este analizat din mai multe perspective, iar pentru fiecare perspectivă se pot executa diferite tehnici de *Process Mining*.

Tendențele actuale în domeniu sunt de a aranja datele în noi formate. Inspirat din cuburile multi-dimensionale OLAP, a apărut ideea de *cuburi de proces* [10, 110]. Cuburile de proces se pot folosi la *Process Mining* comparativ. Un cub de proces conține toate datele dintr-un eveniment pentru o situație particulară, dar datele sunt organizate în mai multe dimensiuni de tipul: timp, departament, locație, cost, gen, nivel, prioritate etc. Acest cub ne ajută să comparăm datele de eveniment și procesele de-a lungul diferitelor dimensiuni. În cazul *Process Mining*, anumite date trebuie configurate: fiecare rând (intrarea) reprezintă un eveniment; se va stabili care coloană reprezintă cazul, care coloană reprezintă activitatea, care coloană reprezintă timpul, iar opțional se va stabili care coloană reprezintă resursa. Din analiza cuburilor, putem compara diferite procese la diferite intervale de timp, la diferite locații, pentru diferiți clienți. Cuburile de proces, asemenea cuburilor OLAP, pot suporta operații de: filiere, grupare pe subcuburi și detalieri. Datele din cuburi OLAP sunt înrudite, dar câteodată datele sunt agregate, fiind nefolositoare pentru *Process Mining* (evenimentele sunt pierdute).

Formatul XES în contextul „Big data”, adică a procesării unui volum foarte mare de date, are deficiențe, și de aceea s-a propus un nou format *Relational XES*, compatibil XES [56].

3.3.4. Calitatea datelor extrase

În prezent există o nevoie mare de a asigura calitatea datelor extrase [33]. Disciplina *Process Mining* a evoluat, iar acum se pune problema calității datelor.

În manifestul pentru *Process Mining*, primul principiu este: „*GP1: Datele de evenimente ar trebui tratate ca cetățeni de clasă înaltă*” [13]. În cadrul aceluiași principiu este definit și un calificativ pentru maturitatea jurnalelor de evenimente, spectrul variind de la ***** (5 stele) pentru jurnale de evenimente excelente calitativ la * (o stea) pentru jurnale de evenimente de slabă calitate. La cel mai scăzut nivel de maturitate, jurnalele de evenimente sunt de proastă calitate, de exemplu, evenimentele înregistrate nu pot corespunde realității sau evenimentele pot lipsi. Un exemplu tipic este un jurnal de evenimente în care evenimentele sunt înregistrate manual.

La cel mai înalt nivel de maturitate, jurnalele de evenimente sunt de o calitate excelentă (de exemplu, de încredere și complete) și evenimentele sunt bine definite. În acest caz, evenimentele (și toate atributele lor) sunt înregistrate automat și au semantică clară (de exemplu, evenimentele se pot referi la o ontologie agreată). Tehnicile de *Process Mining* sunt direct influențate de calitatea jurnalului de evenimente, iar anumite provocări au apărut recent relativ la caracteristicile procesului: date voluminoase (număr ridicat de cazuri), eterogenitate cazurilor (număr mare de trasee distincte), granularitatea evenimentelor (număr ridicat de activități), flexibilitatea procesului și devierea lui (schimbări ale mediului extern influențează schimbări evolutive sau momentane).

Analizând calitatea datelor din jurnale de evenimente s-au identificat mai multe categorii de probleme și anume [33]:

- **Date lipsă** - anumite tipuri de informații obligatorii lipsesc din jurnal. Datele care lipsesc reflectă mai ales o problemă a activității de înregistrare;
- **Date greșite** - deși datele sunt furnizate în jurnal, se poate dovedi că, pe baza informațiilor de context, acestea au fost înregistrate incorect;
- **Date imprecise** - în acest caz apare o problemă de precizie în înregistrarea datelor. De exemplu, dacă amprentele de timp nu sunt înregistrate exact, pot apărea probleme cu ordonarea evenimentelor;
- **Date irelevante** - înregistrările analizate conțin date irelevante, iar jurnalele de evenimente trebuie filtrate.

Aceste patru categorii de probleme au fost puse în relație cu elementele ce constituie un jurnal de activități și astfel au fost identificate 27 de clase de probleme de calitate, sintetizate de către Bose et al. [33], fiind reprezentate schematic în Fig. 3.5.

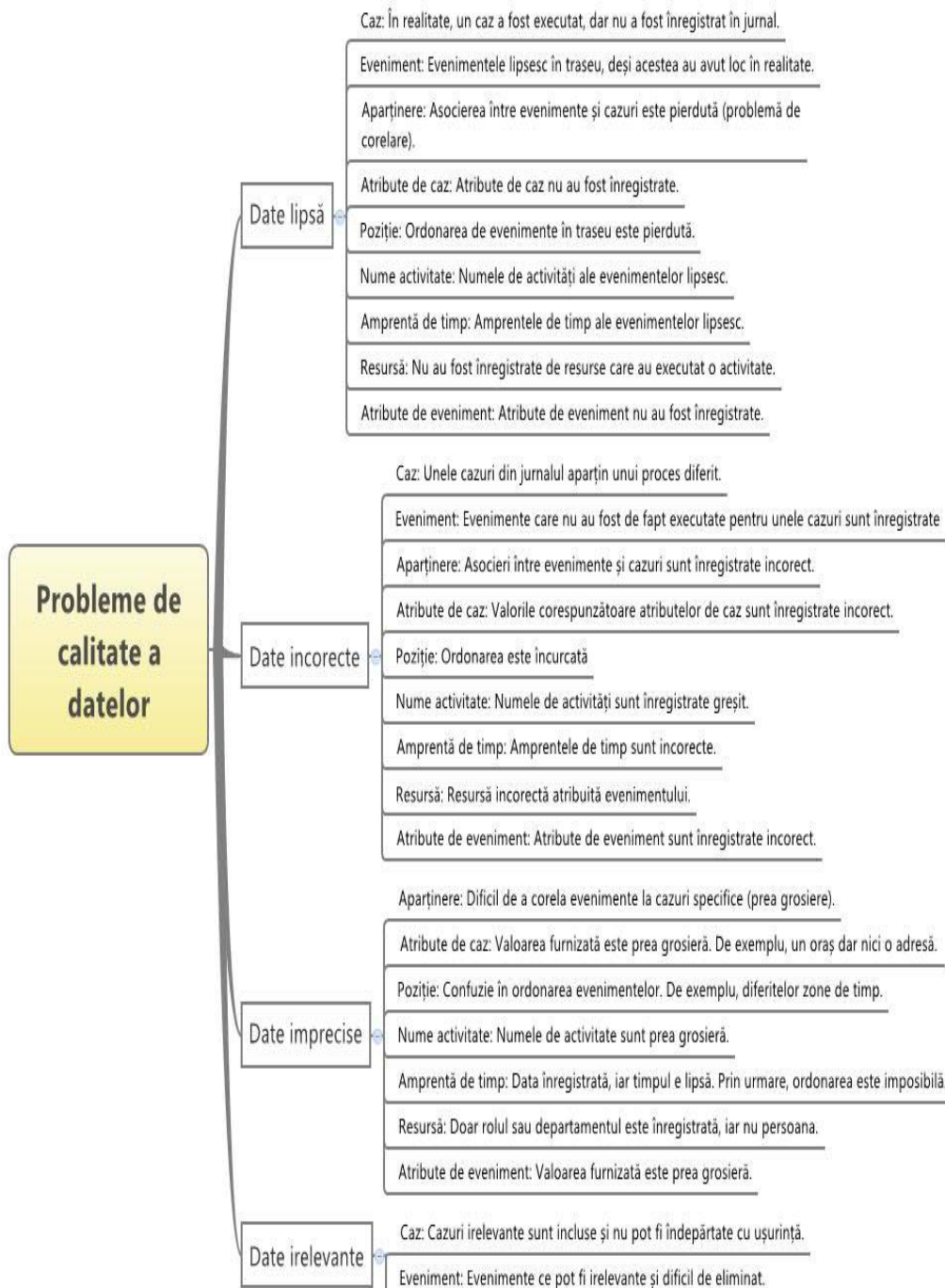


Fig. 3.5. Probleme de calitate ale datelor din jurnalele de evenimente

3.3.5. Directive pentru înregistrarea evenimentelor

Aceste directive se referă la bune practici ce pot îmbunătăți înregistrarea evenimentelor ce urmează a fi extrase într-un jurnal [11]. Există numeroase probleme legate de calitatea datelor, iar cauza principală a imperfecțiunilor sau defectelor o reprezintă proasta înregistrarea a acestora. După cum a fost definit anterior (în capitolul 3.3.2), evenimentele sunt lucruri ce se întâmplă și sunt reprezentate de referințe și atribute. Referințele au nume și un identificator ce se referă la un anumit obiect, iar atributele au un nume și o valoare. Directivele și descrierea acestora este prezentată în Tabelul 3.1.

În acest context, cercetătorii de date trebuie să nu pornească în analiza datelor cu date de tip „gunoi”, pentru că rezultatele de ieșire vor reprezenta tot „gunoi”, ceea ce în limbajul comun specialiștilor de IT este desemnat prin denumirea de *GIGO, Garbage In, Garbage Out* [111].

Tabelul 3.1. Directive pentru înregistrarea evenimentelor [11]

Cod	Descriere
GL1	Numele referințelor și ale atributelor trebui să aibă semantică clară, de exemplu, ar trebui să aibă același înțeles pentru toate persoanele implicate în crearea și analizarea datelor despre evenimente. Diferiți participanți interesați ar trebui să interpreteze datele despre evenimente în același mod.
GL2	Ar trebui să existe o colecție structurată și gestionată pentru numele referințelor și ale atributelor. În mod ideal, numele sunt grupate ierarhic (cum ar fi o taxonomie sau ontologie). Un nou nume de referință sau de atribut poate fi adăugat numai după ce există consens privind valoarea și semnificația sa.
GL3	Referințele trebuie să fie stabile (de exemplu, identificatorii nu trebuie reutilizați sau să depindă de context). De exemplu, referințele nu ar trebui să fie dependente de timp, de regiune, sau de limbă. Unele sisteme creează diferite jurnale în funcție de setările de limbă. Acest lucru complică inutil analiza.
GL4	Valorile atributelor trebuie să fie cât de precise posibil. Dacă valoarea nu are precizia dorită, acest lucru trebuie indicat în mod explicit. De exemplu, dacă pentru unele evenimente numai data este cunoscută, dar nu timpul exact, atunci acest lucru trebuie precizat în mod explicit, printr-un calificativ.
GL5	Incertitudinea cu privire la producerea evenimentului, la referințele lor sau la atributele lor ar trebui capturate prin calificative corespunzătoare. De exemplu, din cauza unor erori de comunicare, unele valori pot fi mai puțin fiabile decât de obicei. Incertitudinea este diferită de imprecizie.
GL6	Evenimentele ar trebui cel puțin să fie parțial ordonate. Ordinea evenimentelor poate fi stocată în mod explicit (de exemplu, folosind o listă) sau implicit, prin intermediul unui atribut care denotă timpul evenimentului. În cazul în care înregistrarea amprentelor de timp nu este de încredere sau imprecisă, pot exista totuși modalități de a ordona evenimente bazate pe cauzalități observate (de exemplu, utilizarea altor date).
GL7	Dacă este posibilă stocarea informației de tranzacție, cu atât mai bine (start, completare, anulare, programare, suspendare, reluare, atribuire, retragere ș.a.). Dacă evenimentele de start și de completare sunt prezente, atunci se va putea calcula durata activităților

Cod	Descriere
GL8	Ar trebui să se efectueze regulat verificări automate de consistență și de corectitudine pentru a asigura acuratețea sintactică a jurnalului de evenimente. Ar trebui verificat dacă referințele sau atributele lipsesc, ori dacă numele de referințe/atribute sunt cele convenite. Asigurarea calității evenimentelor este un proces continuu (pentru a evita degradarea calității jurnalului în timp).
GL9	Trebuie asigurată comparabilitatea jurnalelor de evenimente în timp, precum și a diferitelor grupuri de cazuri sau variante de proces. Procesul de înregistrare în sine nu ar trebui să se schimbe în timp (fără a fi raportate schimbările). Pentru Mining comparativ de proces, este vital ca aceleași principii de exploatare să fie utilizate.
GL10	Este interzis a agrega evenimente în jurnalul de evenimente folosit ca intrare pentru procesul analizat. Agregarea trebuie făcută în timpul analizei și nu înainte (din moment ce nu poate fi refăcută). Datele de evenimente ar trebui să fie cât mai „neprelucrate” cu putință
GL11	Este interzis a scoate evenimente pentru a asigura proveniență. Reproducibilitatea este esențială pentru procesul de Mining. De exemplu, un student nu trebuie șters din baza de date după ce a renunțat la un curs, deoarece acest lucru poate duce la inducerea în eroare a rezultatelor analizei. De asemenea: concerte nu sunt șterse, ci sunt anulate; angajații nu sunt șterși, ci sunt concediați etc.
GL12	Trebuie să se asigure intimitatea fără a pierde corelații semnificative. Date sensibile sau private ar trebui să fie eliminate cât mai devreme posibil (de exemplu, înainte de analiză). Cu toate acestea, dacă este posibil, ar trebui să se evite scoaterea corelațiilor. Hashing-ul poate fi un instrument puternic în compromisul dintre intimitate și analiză.

3.4. Descoperirea procesului

Următoarea etapă a proiectului de *Process Mining* reprezintă descoperirea (*discovery*) procesului. Ca date de intrare se folosește jurnalul de evenimente, iar ca date de ieșire, va rezulta un model de proces în formatul unei rețele Petri (se folosesc rețele Petri pentru că acestea modelează foarte bine concurența). Trebuie precizat că descoperirea procesului în contextul *Process Mining* se referă la modelarea procesului prin tehnici specifice de *Process Mining*.

3.4.1. Algoritmul α

Algoritmul α este un algoritm de bază pentru descoperirea unui proces și face parte din tehnica *interpretării interne (play-in)*, prezentată în secțiunea 3.1.2. Algoritmul a apărut în 2004 [21], dar bazele lui au fost puse de o cercetătoare româncă, Laura Mărușter, în teza ei de doctorat [115] prin care a aplicat metode de învățare automată pentru a înțelege procesele de afaceri. Obiectivul reprezintă descoperirea procesului din înregistrările de evenimente. Acest proces va fi reprezentat sub formă de rețea Petri. O înregistrare de evenimente reprezintă un set de trasee. Un traseu este o secvență de nume de activități.

Algoritmul α a fost unul dintre primii algoritmi ce puteau trata concurența. Algoritmul α este un algoritm simplu, clasic, el fiind fundamentul pentru alți algoritmi avansați.

Ca intrare pentru algoritmul α avem un jurnal de evenimente simplu L peste \mathcal{A} , unde $L \in \mathbb{B}(\mathcal{A}^*)$. Ieșirea algoritmului α reprezintă o rețea Petri marcată, $\alpha(L) = (N, M)$. Algoritmul α caută în jurnalul de evenimente șabloane particulare. Există patru tipare ce pot fi descoperite din jurnal, iar aceste tipare sunt rezultate din relațiile de ordonare a activităților.

Definiție (Relații de ordonare conform jurnalului): Fie L un jurnal de evenimente peste \mathcal{A} , unde $L \in \mathbb{B}(\mathcal{A}^*)$. Fie $a, b \in \mathcal{A}$, două activități oarecare. Atunci următoarele relații sunt definite:

- *Relația de succesiune:*
 $a >_L b$ (a este urmat/succedat de b) (3.1)
 dacă și numai dacă există un traseu $\sigma = \langle t_1, t_2, \dots, t_n \rangle$ și $i \in \{1, \dots, n-1\}$
 astfel încât $\sigma \in L$ și $t_i = a$ și $t_{i+1} = b$;

- *Relația de cauzalitate:*
 $a \rightarrow_L b$ (a cauzează/este strict urmat de b) (3.2)
 dacă și numai dacă $a >_L b$ și $b \not>_L a$;

- *Relația de exclusivism:*
 $a \#_L b$ (a este exclusiv/indiferent față de b) (3.3)
 dacă și numai dacă $a \not>_L b$ și $b \not>_L a$

- *Relația de paralelism:*
 $a \parallel_L b$ (a este paralel cu b) (3.4)
 dacă și numai dacă $a >_L b$ și $b >_L a$.

Să considerăm următorul jurnal de evenimente $L = [\langle a, b, d, e, f, g \rangle^{10}, \langle a, c, d, f, e, g \rangle^5, \langle a, c, d, e, f, g \rangle^2]$. Pentru acest jurnal de evenimente, relațiile de ordonare ar fi următoarele:

$$>_L = \{(a, b), (a, c), (b, d), (c, d), (d, f), (d, e), (e, f), (e, g), (f, e), (f, g)\}$$

$$\rightarrow_L = \{(a, b), (a, c), (b, d), (c, d), (d, e), (d, f), (e, g), (f, g)\}$$

$$\#_L = \{(a, a), (a, d), (a, e), (a, f), (a, g), (b, b), (b, c), (b, e), (b, f), (b, g), (c, b), (c, c), (c, e), (c, f), (c, g), (d, a), (d, d), (d, g), (e, a), (e, b), (e, c), (e, e), (f, a), (f, b), (f, c), (f, f), (g, a), (g, b), (g, c), (g, d), (g, g)\}$$

$$\parallel_L = \{(e, f), (f, e)\}$$

Pornind de la aceste relații de ordonare, se poate construi matricea amprentei (*footprint*) jurnalului de evenimente L :

Tabelul 3.2. Matricea amprentei jurnalului de evenimente L

	a	b	c	d	e	f	g
a	#	→	→	#	#	#	#
b	←	#	#	→	#	#	#
c	←	#	#	→	#	#	#
d	#	←	←	#	→	→	#
e	#	#	#	←	#	∥	→
f	#	#	#	←	∥	#	→
g	#	#	#	#	←	←	#

Relațiile de ordonare pot fi folosite la descoperirea de tipare și la reconstituirea procesului original. Tiparele fac legătura între relațiile de ordonare și elemente ale rețelelor Petri. Tiparele sunt:

- Tipar de secvență: $a \rightarrow b$ (3.5)

- Tipar de decizie exclusivă (*XOR-split*): $a \rightarrow b, b \rightarrow c$ și $b \# c$ (3.6)

- Tipar de unificare exclusivă (*XOR-join*): $a \rightarrow c, b \rightarrow c$ și $a \# b$ (3.7)

- Tipar de bifurcare paralelă (*AND-split*): $a \rightarrow b, a \rightarrow c$ și $a \parallel b$ (3.8)

- Tipar de îmbinare paralelă (*AND-join*): $a \rightarrow c, b \rightarrow c$ și $a \parallel b$ (3.9)

Definiție (Algoritm α): Fie L un jurnal de evenimente peste $T \subseteq \mathcal{A}$. $\alpha(L)$ este definit astfel:

$$(1) \quad T_L = \{t \in T \mid \exists \sigma \in L \ t \in \sigma\} \quad (3.10)$$

$$(2) \quad T_I = \{t \in T \mid \exists \sigma \in L \ t = \text{prim}(\sigma)\} \quad (3.11)$$

$$(3) \quad T_O = \{t \in T \mid \exists \sigma \in L \ t = \text{ultim}(\sigma)\} \quad (3.12)$$

$$(4) \quad X_L = \{(A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall a \in A \forall b \in B \ a \rightarrow_L b \wedge \forall a_1, a_2 \in A \ a_1 \#_L a_2 \wedge \forall b_1, b_2 \in B \ b_1 \#_L b_2\} \quad (3.13)$$

$$(5) \quad Y_L = \{(A, B) \in X_L \mid \forall (A', B') \in X_L \ A \subseteq A' \wedge B \subseteq B' \implies (A, B) = (A', B')\} \quad (3.14)$$

$$(6) \quad P_L = \{p_{(A,B)} \mid (A, B) \in Y_L\} \cup \{i_L, o_L\} \quad (3.15)$$

$$(7) \quad F_L = \{(a, p_{(A,B)}) \mid (A, B) \in Y_L \wedge a \in A\} \cup \{(p_{(A,B)}, b) \mid (A, B) \in Y_L \wedge b \in B\} \cup \{(i_L, t) \mid t \in T_I\} \cup \{(t, o_L) \mid t \in T_O\} \quad (3.16)$$

$$(8) \quad \alpha(L) = (P_L, T_L, F_L) \quad (3.17)$$

Astfel, T_L reprezintă setul de activități din jurnalul de evenimente L și este notat cu T deoarece se referă la tranzițiile modelului descoperit. T_I sunt setul de activități care apar primele într-un traseu. T_O sunt setul de activități care apar ultimele într-un traseu. $p_{(A,B)}$ reprezintă locația ce trebuie construită, unde A reprezintă setul de tranziții de intrare $\bullet p_{(A,B)} = A$, respectiv B , setul de tranziții de ieșire $p_{(A,B)} \bullet = B$. Toate elementele din setul A trebuie să fie urmate de toate elementele din setul B . Elementele din setul A nu trebuie să se urmeze unele pe altele și nici elementele din setul B nu trebuie să se urmeze unele pe altele. După ce se obțin toate perechile (A, B) ce formează mulțimea X_L , se va construi mulțimea Y_L prin care se selectează doar perechile maximale. Oricărui element (A, B) din mulțimea Y_L îi corespunde o locație $p_{(A,B)}$. P_L reprezintă mulțimea tuturor locațiilor rețelei care au fost descoperite, la care se mai adaugă locația de start i_L și locația finală o_L . Mulțimea F_L conține toate arcele de legătură dintre locații și tranziții, respectiv tranziții și locații. Rezultatul este o rețea Petri $\alpha(L) = (P_L, T_L, F_L)$, așa cum a fost definită în secțiunea 2.1.1.

După aplicarea algoritmului α asupra logului L obținem:

$$T_L = \{a, b, c, d, e, f, g\}$$

$$T_I = \{a\}$$

$$T_O = \{g\}$$

$$X_L = \{(\{a\}, \{b\}), (\{a\}, \{c\}), (\{a\}, \{b, c\}), (\{b\}, \{d\}), (\{c\}, \{d\}), (\{b, c\}, \{d\}), (\{d\}, \{e\}), (\{d\}, \{f\}), (\{e\}, \{g\}), (\{f\}, \{g\})\}$$

$$Y_L = \{(\{a\}, \{b, c\}), (\{b, c\}, \{d\}), (\{d\}, \{e\}), (\{d\}, \{f\}), (\{e\}, \{g\}), (\{f\}, \{g\})\}$$

$$P_L = \{p_{(\{a\}, \{b, c\})}, p_{(\{b, c\}, \{d\})}, p_{(\{d\}, \{e\})}, p_{(\{d\}, \{f\})}, p_{(\{e\}, \{g\})}, p_{(\{f\}, \{g\})}, i_L, o_L\}$$

$$F_L = \{(a, p_{(\{a\}, \{b, c\})}), (p_{(\{a\}, \{b, c\})}, b), (p_{(\{a\}, \{b, c\})}, c), (b, p_{(\{b, c\}, \{d\})}), (c, p_{(\{b, c\}, \{d\})}), (p_{(\{b, c\}, \{d\})}, d), (d, p_{(\{d\}, \{e\})}), (d, p_{(\{d\}, \{f\})}), (p_{(\{d\}, \{e\})}, e), (p_{(\{d\}, \{f\})}, f), (e, p_{(\{e\}, \{g\})}), (f, p_{(\{f\}, \{g\})}), (p_{(\{e\}, \{g\})}, g), (p_{(\{f\}, \{g\})}, g), (i_L, a), (g, o_L)\}$$

$$\alpha(L) = (P_L, T_L, F_L)$$

Rețeaua Petri descoperită din jurnalul de evenimente $L = [\langle a, b, d, e, f, g \rangle^{10}, \langle a, c, d, f, e, g \rangle^5, \langle a, c, d, e, f, g \rangle^2]$ este reprezentată grafic în Fig. 3.6. Rețelele Petri descoperite cu ajutorul algoritmului α sunt rețele Flux de Lucru.

Modelul de proces ce a fost descoperit trebuie evaluat pentru a se observa conformitatea lui relativ la comportamentul înregistrat. Calitatea unui algoritm pentru descoperirea unui proces este măsurată de cuantificarea măsurii în care modelul rezultat poate reproduce comportamentul din jurnal, adică potrivirea reluării (din englezul *replay fitness*). Cele mai importante criterii de calitate pentru construirea de modele sunt [8, 41, 43]:

- *Potrivirea* – se referă la faptul că modelul creat ar trebui să permită comportamentul observat din jurnalul de evenimente. Un model are potrivire perfectă dacă toate traseele din jurnal se pot executa integral;
- *Simplitatea* – arată că modelul creat trebuie să fie cât mai simplu posibil. Având la bază principiul *Briciului lui Occam* [53], acesta enunță faptul că modelul ce explică comportamentul din jurnal într-un mod cât mai simplu, este considerat cel mai bun model;
- *Precizia* – se referă la faptul că modelul creat nu ar trebui să permită un comportament complet diferit față de ceea ce a fost observat în jurnalul de evenimente. Un model este precis dacă nu permite „prea mult” comportament;
- *Generalizarea* – se referă la faptul că modelul creat trebuie să generalizeze comportamentul observat în jurnalul de evenimente, astfel încât să mai permită și alt comportament.

Între cele patru criterii trebuie găsit un echilibru, după cum se poate observa în Fig. 3.7. Cele patru criterii sunt antinomice două câte două: pe de-o parte potrivirea cu simplitatea, pe de altă parte precizia cu generalizarea. Vom arăta ulterior că dintr-un singur jurnal de evenimente se pot descoperi mai multe modele. Printre acestea unele sunt foarte exacte, dar nu sunt simple, altele sunt simple, dar nu exacte. Unele modele sunt foarte generale și ele permit mai multe comportamente decât cele descrise în jurnalul de evenimente, iar altele sunt precise și nu mai permit alte comportamente.

Când se modelează un proces, se pornește de la un jurnal de evenimente care, de obicei, este parțial. Din motive de performanță (un proces poate să fi fost executat ani de zile) se alege doar o perioadă mai restrânsă pentru analiză. Din această perioadă de analiză pot rezulta modele precise, dar care nu sunt generale, pentru că nu tot intervalul a fost luat în considerare. Alegerea punctului de echilibru dintre cele patru criterii face ca descoperirea unui proces să fie o activitate foarte provocatoare.

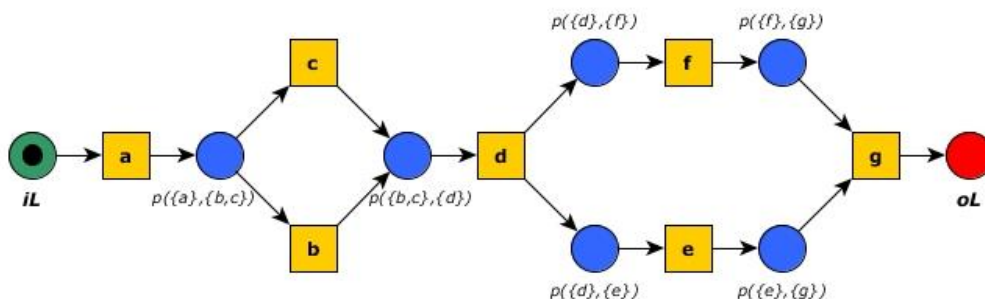


Fig. 3.6. Rețeaua Petri descoperită cu algoritmul α din jurnalul de evenimente L

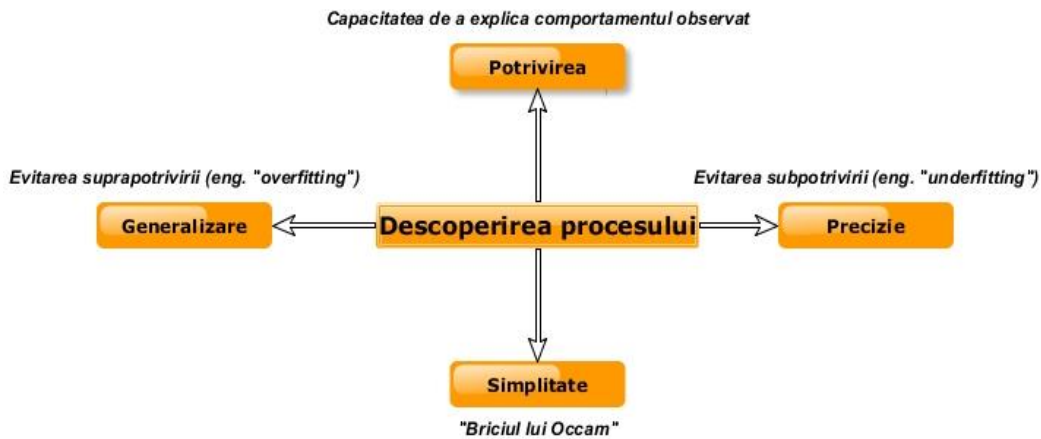


Fig. 3.7. Cele patru dimensiuni calitative care caracterizează un model de proces descoperit

În [117] sunt prezentate limitările algoritmului α :

- Algoritmul are deficiențe dacă jurnalul de evenimente nu este complet din punctul de vedere al relațiilor de ordonare;
- Algoritmul are probleme atunci când sunt descoperite cicluri scurte (de una sau două activități). Tot în [117] și în [118] sunt prezentate și extensii ale algoritmului ce pot rezolva această problemă, algoritmul ce poartă numele α^+ ;
- Algoritmul nu poate descoperi dependențe non-locale, rezultate din diferite construcții de decizii. Această problemă a fost rezolvată în [167], iar algoritmul se numește α^{++} ;

Algoritmul nu tratează problema frecvenței traseelor, de aceea algoritmul este sensibil la „zgomote” (perturbații) și la jurnalele care nu sunt complete.

3.4.2. Algoritmi avansați

Sunt multe alternative de algoritmi pentru *Process Mining*. Algoritmii sunt aleși în funcție de problema studiată, ținându-se cont de criteriile de calitate discutate mai sus: potrivire, precizie, generalizare și simplitate. Dar sunt și alți factori care influențează alegerea algoritmului: viteza, utilizarea memoriei, reprezentarea aleasă, flexibilitatea, abilitatea de a gestiona zgomotele, asumarea faptului că jurnalul este complet.

Pe lângă algoritmul α [21] și extensiile lui α^+ [118] și α^{++} [167] vom prezenta în continuare alți algoritmi folosiți în descoperirea de proces, algoritmi ce au apărut în ultimii 10 ani.

3.4.2.1. Algoritmul HeuristicsMiner

Algoritmul *HeuristicsMiner* este prezentat în [164], iar acesta extinde algoritmul α prin faptul că ține cont de frecvența relațiilor de ordonare pentru a calcula tabele de dependență/frecvență din jurnalul de evenimente și folosește metode euristice pentru a converti această informație într-o rețea euristică. Algoritmul conține trei etape:

1. Mining pentru graful de dependență – implică construcția matricii de dependență în funcție de frecvența relațiilor de ordonare cu formula de mai jos, iar apoi construcția grafului de dependență:

$$|a \Rightarrow_L b| = \begin{cases} \frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1}, \text{dacă } a \neq b \\ \frac{|a >_L a|}{|a >_L a| + 1}, \text{dacă } a = b \end{cases} \quad (3.18)$$

unde

$$|a >_L b| = \sum_{\sigma \in L} L(\sigma) \times |\{1 \leq i < |\sigma| \mid \sigma(i) = a \wedge \sigma(i+1) = b\}| \quad (3.19)$$

$|a \Rightarrow_L b|$ poartă numele de valoarea de dependență, iar $|a >_L b|$ se numește valoarea succesiunii directe și reprezintă frecvența cu care activitatea a este urmată de activitatea b .

Pentru jurnalul $L = [\langle a, b, d, e, f, g \rangle^{10}, \langle a, c, d, f, e, g \rangle^5, \langle a, c, d, e, f, g \rangle^2]$ se consideră următoarele matrici cu valorile succesiunii directe și cu valorile de dependență:

Pornind de la matricea amprentei (Tabelul 3.2), se poate construi graful de dependență (Fig. 3.8), care ține cont doar de relațiile de cauzalitate (\rightarrow) din graf. De remarcat că acest graf nu ține cont de relațiile de paralelism (\parallel) și de relațiile de exclusivitate ($\#$). Pe fiecare arc s-a trecut valoarea succesiunii directe (din Tabelul 3.3) și valoarea de dependență (din Tabelul 3.4).

Tabelul 3.3. Matricea valorilor succesiunii directe pentru jurnalul de evenimente L

$ a >_L b $	a	b	c	d	e	f	g
a	0	10	7	0	0	0	0
b	0	0	0	10	0	0	0
c	0	0	0	7	0	0	0
d	0	0	0	0	12	5	0
e	0	0	0	0	0	12	5
f	0	0	0	0	5	0	12
g	0	0	0	0	0	0	0

Tabelul 3.4. Matricea valorilor de dependență pentru jurnalul de evenimente L

$ a \Rightarrow_L b $	a	b	c	d	e	f	g
a	0,00	0,91	0,88	0,00	0,00	0,00	0,00
b	-0,91	0,00	0,00	0,91	0,00	0,00	0,00
c	-0,88	0,00	0,00	0,88	0,00	0,00	0,00
d	0,00	-0,91	-0,88	0,00	0,92	0,83	0,00
e	0,00	0,00	0,00	-0,92	0,00	0,39	0,83
f	0,00	0,00	0,00	-0,83	-0,39	0,00	0,92
g	0,00	0,00	0,00	0,00	-0,83	-0,92	0,00

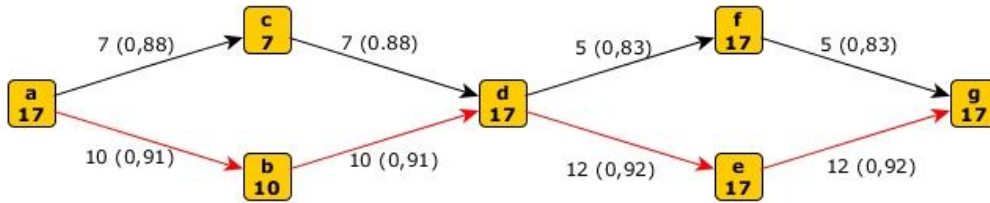


Fig. 3.8. Graful de dependență rezultat din jurnalul de evenimente L

Valorile de dependență se pot compara cu o anumită valoare de prag, iar dacă valoarea de dependență este sub acest prag, atunci arcul se poate ignora. Astfel, dacă în Fig. 3.8 se consideră ca valoare de prag 0,9, atunci graful de dependență ar conține doar arcele marcate cu roșu și anume (a,b) , (b,d) , (d,e) și (e,g) , restul fiind ignorate. Numerele de sub numele unei activități reprezintă frecvența de apariție a activității în jurnalul de evenimente L .

Rețelele cauzale tip C -nets (prezentate în capitolul 2.1.3) reprezintă notația agreată ce va fi folosită în continuare în cadrul algoritmului pentru *Mining heuristic*, și nu rețelele Petri.

2. Învățarea disocierilor și a unificărilor prin două clase de abordări:

- Abordări euristice*, folosind o fereastră de timp înainte și după fiecare activitate. Legăturile se pot determina prin numărarea seturilor de activități de intrare și de ieșire. O fereastră de timp poate fi considerată prin diferite mărimi, acestea reprezentând câte activități din fața/de după activitatea analizată sunt luate în considerare [44].
- Abordări de optimizare bazate pe reluare*. Pornind de la un set de legături de intrare și de ieșire, se pune problema dacă realitatea se poate relua corect. Mulțimea de legături de intrare și de ieșire sunt finite, de aceea se poate stabili o funcție care să găsească cel mai bun set de legături [165].

3. **Vizualizarea în formatul dorit.** De obicei, este nevoie de conversia rețelei cauzale tip C -nets în formate de tipul: BPMN, EPC, rețea Flux de Lucru etc. Mai multe despre conversia rețelelor cauzale tip C -nets se găsesc în [3].

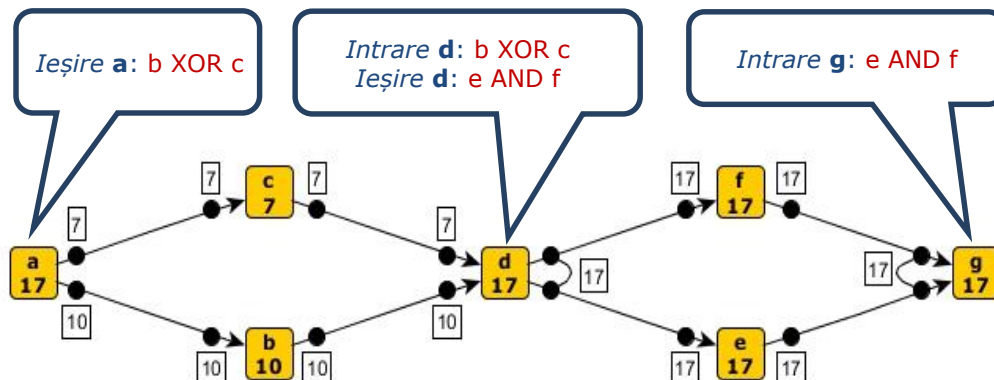


Fig. 3.9. Învățarea disocierilor și a unificărilor într-o rețea cauzală tip C -Nets adnotată cu frecvențe

3.4.2.2. Algoritmul Fuzzy Miner

Algoritmul Fuzzy Miner a fost introdus în 2007 de Günther și Aalst [73] și este apropiat de algoritmul *HeuristicsMiner*. Algoritmii existenți tind să funcționeze pe procese structurate, dar nu reușesc să furnizeze rezultate bune pentru procese nestructurate. Algoritmul aduce o nouă perspectivă asupra modelului de proces prin care sunt evidențiate importanța activităților și relațiile dintre acestea.

Algoritmul propune o abordare din mai multe perspective, având ca fundament metrici de proces deduse din jurnal. Günther definește două clase de metrici: *semnificația* (care arată importanța relativă a comportamentului sau cât de importante sunt anumite activități și arce) și *corelarea* (care arată cât de apropiate sunt două evenimente între ele).

Metricile referitoare la *semnificație* sunt grupate în:

- *Semnificația unară*:
 - *semnificația frecvenței* – cu cât o activitate este mai frecventă, cu atât ea este mai importantă;
 - *semnificația de rutare* – cu cât diferă predecesorii unui nod față de succesorii lui;
- *Semnificația binară*:
 - *semnificația frecvenței* – cu cât două activități succesive sunt mai frecvente, cu atât este mai importantă relația dintre ele;
 - *semnificația distanței* – arată măsura cu care diferă o relație în comparație cu importanța nodurilor sursă și țintă.

Metricile referitoare la *corelare* sunt în binare:

- *corelarea proximității* (determină activitățile ce urmează la scurt timp una de alta),
- *corelarea inițiatorului* (determină persoanele ce execută activități succesive),
- *corelarea punctului de final* (compară nume de activități pentru activități succesive),
- *corelarea tipurilor de date* (compară activități succesive ce partajează cantități mari de date) și
- *corelarea valorilor* (compară activități succesive din punctul de vedere al valorilor altor atribute).

Pornind de la aceste metrici se aplică trei metode de transformare asupra modelului de proces care va fi simplificat:

- (i) rezolvarea conflictelor;
- (ii) filtrarea arcelor (se elimină arcele mai puțin semnificative);
- (iii) agregarea și abstractizarea (se elimină nodurile mai puțin semnificative).

Algoritmul Fuzzy Miner stă la baza programului Disco (4.2.2) creat tot de același autor.

3.4.2.3. Algoritmul Genetic Process Mining

Algoritmul Genetic Process Mining [119] este derivat din aria inteligenței artificiale și folosește o abordare iterativă revoluționară care imită procesul de evoluție naturală. Algoritmul conține patru pași, ca orice algoritm genetic:

- (i) inițializare;
- (ii) selecție;
- (iii) reproducere;
- (iv) terminare.

Algoritmi genetici încep cu o populație inițială de indivizi, în cazul nostru un posibil model de proces. Fiecare individ are asociată o valoare de potrivire care îi arată calitatea. Funcția de potrivire determină cât de bine individul este capabil să reproducă comportamentul din jurnal. Populațiile evoluează prin selectarea celor mai potriviți indivizi și prin generarea de noi indivizi folosind operatori genetici precum încrucișarea (combinarea părților a doi sau mai mulți indivizi) sau mutația (modificare aleatoare a unui individ). Algoritmul genetic este robust din punct de vedere al „zgomotelor” (perturbațiilor) și dă rezultate bune chiar dacă jurnalul este incomplet.

Ca mod de reprezentare, algoritmul folosește matricea cauzală. Fiecare individ are propria sa matrice cauzală. Algoritmul caută acele modele care sunt complete și precise. Un model este complet dacă conține toate traseele din jurnal. Un model este precis dacă el nu mai permite alt comportament decât cel din jurnal.

Pentru jurnale foarte mari, algoritmul nu este deosebit de eficient și va necesita un timp foarte îndelungat pentru a identifica un model de o potrivire acceptabilă. Algoritmul suportă implementări paralele, astfel că jurnalul se poate împărți în mai multe părți, iar fiecare parte se poate distribui pentru calcul paralel, separat [38]. Această abordare îmbunătățește eficiența algoritmului.

Îmbinarea *algoritmului HeuristicsMiner* cu *algoritmul Genetic Process Mining* poate aduce rezultate bune în cadrul *Process Mining*. Astfel, se descoperă modelul cu ajutorul *algoritmului HeuristicsMiner*, iar acest model reprezintă populația inițială pentru *algoritmul Genetic Process Mining*. Acest algoritm îmbunătățit este apărut în 2015 și se numește *ProDiGen* [32].

Algoritmul Genetic Process Mining este un algoritm clasic și el stă la baza altor algoritmi precum *Evolutionary Tree Miner (ETM)* [43, 67] și *Honey Bee Genetic Algorithm (HBGA)* [149].

3.4.2.4. Algoritmi bazați pe teoria regiunilor

Teoria regiunilor stabilește o conexiune între sistemele de tranziții și rețele Petri prin așa-numita sinteză de rețea. Sistemele de tranziții sunt un alt suport pentru reprezentarea unui model de proces, având următoarele avantaje: conțin toate stările posibile și conțin toate tranzițiile posibile. Dezavantajul sistemelor de tranziții este faptul că un model de proces complex reprezentat cu acestea devine foarte voluminos, întrucât el conține toate stările și tranzițiile. Pentru a remedia acest dezavantaj, regiunile bazate pe stări sintetizează un model mai compact din respectivele sisteme de tranziții. Prima aplicare a teoriei regiunilor în *Process Mining* este citată în [54].

Printre subclasele de algoritmi reprezentanți ai algoritmilor bazați pe teoria regiunilor sunt *Regiuni Bazate pe Stări (State-Based Regions, RBS)* și *Regiuni Bazate pe Limbaj (Language-Based Regions, RBL)*. Acești algoritmi descoperă modelul de proces în două faze. Prima fază se reduce la construcția sistemului de tranziție din jurnalul de evenimente și apoi la identificarea regiunilor din sistemul de tranziție. A doua fază o reprezintă conversia sistemului de tranziție cu regiunile descoperite în rețea Petri. Ideea de bază este aceea că o regiune corespunde unei locații de rețea Petri. Activitățile unei regiuni pot fi clasificate ca activități de intrare, de ieșire și de non-trecere față de regiune. O uniune de două regiuni conduce la o nouă regiune. Algoritmul de descoperire a procesului prin *Regiuni Bazate pe Stări* este descris în [18].

Algoritmul de descoperire a procesului prin *Regiuni Bazate pe Limbaj* este similar celui precedent, doar că acesta nu are ca intrare un sistem de tranziții, ci un limbaj [168]. În acest caz, limbajul se referă la un set finit de secvențe. Se definește un limbaj \mathcal{L} peste un alfabet T , iar obiectivul este de a identifica rețeaua Petri $N(L)$, în

care tranzițiile corespund simbolurilor din alfabetul limbajului și în care fiecare cuvânt reprezintă o secvență validă.

Problema se reduce la un sistem de inecuații de tipul:

$$1 \cdot c + A' \cdot x - A \cdot y \geq 0 \quad (3.20)$$

unde A și A' reprezintă două matrici $|\mathcal{L}| \times |\mathcal{T}|$, iar (x, y, c) reprezintă o regiune bazată pe limbaj.

Orice soluție la această inecuație se poate adăuga la rețeaua Petri. Se pune problema alegerii celor mai interesante locații cu ajutorul unei funcții de scop. Aceasta se traduce printr-o problemă de optimizare, care se poate soluționa folosind metoda Programării Liniare în Numere Întregi (*Integer Linear Programming, ILP*) [168].

3.4.2.5. Algoritmul Inductive Miner

Bazele *algoritmului Inductive Miner* au fost puse în lucrarea [102]. S-a pornit de la ideea că până la momentul apariției algoritmului nu exista unul care să returneze un model complet (fără părți moarte sau alte anomalii), care să gestioneze comportamentul ocazional și să aibă un timp de procesare scurt. Algoritmul folosește arbori de proces (*process tree*), care sunt o reprezentare abstractă pentru un bloc structurat complet al unei rețele flux de lucru. Un arbore reprezintă un limbaj, o frunză descrie o activitate, iar un nod descrie relația de combinare dintre copiii lui. În cadrul acestui algoritm, pentru relația de combinare dintre copiii unui nod sunt considerați patru operatori: \times (selecție exclusivă), \rightarrow (compoziția secvențială), \wedge (compoziție paralelă intercalată) și \cup (operator secvență repetitivă).

Algoritmul Inductive Miner funcționează recursiv: (i) selectarea operatorului rădăcină care se potrivește cel mai bine pentru jurnalul L ; (ii) divizarea activităților din jurnalul L în seturi disjuncte; (iii) împărțirea jurnalului L folosind aceste seturi în sub-jurnale. Aceste sub-jurnale sunt supuse procesului de Mining recursiv până când un sub-jurnal conține doar o singură activitate.

În literatura de specialitate se explică modul de aplicare a principiul Pareto asupra jurnalului de evenimente, adică 80% din comportamentul observat poate fi explicat cu un model restrâns definit ca 20% din modelul ce descrie întreg comportamentul. A rezultat *Algoritmul Inductive Miner-infrequent* [103] ce folosește filtrarea acelor arce care sunt rare. Astfel, dintr-un nod sunt alese doar acele arce de ieșire pentru care raportul dintre frecvența lor și frecvența celui mai des întâlnit arc de ieșire este mai mare decât un factor k . Filtrarea are loc de fiecare dată, înainte de pasul de divizare a activităților. În acest context, sunt realizate și alte filtrări cum sunt filtrările: asupra arcelor puțin frecvente din cicluri, asupra activităților izolate, asupra traseelor goale, asupra divizărilor de jurnal.

3.4.2.6. Alți algoritmi

În literatură sunt prezentați foarte mulți algoritmi pentru descoperirea de proces printre care remarcăm (denumirea lor este doar în engleză, sau acronim constituind parte a limbajului de specialitate din domeniu): *AGNEs*, *Episode Miner*, *Declare Miner*, *LTL Miner*, *Lattice Miner*, *Conformal Process Graph*, *Hidden Markov Models*, *Stochastic Task Graph Miner*, *Fusion Miner* ș.a. Domeniul este relativ nou, având în vedere că algoritmul α a apărut în 2003, dar acest domeniu este intens cercetat, dovadă fiind multitudinea de algoritmi noi apăruti în ultimii ani. Un studiu recent prezintă comparativ acești algoritmi din punctul de vedere al calității modelelor obținute [163].

3.5. Verificarea conformității

A treia etapă a proiectului de *Process Mining* îl reprezintă conectarea evenimentelor din jurnal cu activități din model. Este etapa în care se verifică conformitatea modelului creat și se realizează alinieri între jurnal și model. Când se pune problema verificării conformității, aceleași patru dimensiuni (generalizarea, precizia, potrivirea și simplitatea) joacă un rol important în compararea jurnalului cu modelul. Totuși, în cadrul verificării conformității, reluarea în contextul potrivirii are un caracter dominant. Potrivirea (*fitness*) reprezintă măsura în care comportamentul din jurnal este posibil, conform modelului. Dacă se urmărește reluarea unui caz, urmărind tranzițiile print-un model de proces, de multe ori se observă nereguli. Din moment ce procesul s-a executat, se poate „forța” și în cazul reluării, ca procesul să fie finalizat, iar toate neregulile din timpul parcurgerii modelului vor fi notate.

Obiectivul verificării conformității este de a descoperi asemănările și diferențele dintre comportamentul modelat și cel observat, atât global, cât și local. Conformitatea globală observă problemele holistice între model și jurnal. Conformitatea locală identifică în model unele noduri cu probleme, iar în jurnal unele cazuri cu probleme. Non-conformitatea trebuie interpretată pornind de la scopul modelului. Astfel, modelul poate fi *descriptiv*, în cazul în care captează și prezice (prognozează) realitatea, iar în acest caz non-conformitatea se traduce în acțiuni de îmbunătățire a modelului pentru ca acesta să surprindă mai bine realitatea. Totodată, modelul poate fi *normativ*, în cazul în care este folosit pentru a influența sau a controla realitatea, iar în acest caz non-conformitatea se poate traduce în *devieri nedorite* sau *devieri dorite*. Trebuie analizate atent devierile dorite, sau mai bine spus acele devieri care duc la un rezultat bun neașteptat, pentru că ele sunt un punct de pornire spre progres. De reținut este faptul că și conformitatea poate fi privită din două perspective: (1) modelul nu captează comportamentul real, adică modelul este greșit (modelul se dorește a fi *descriptiv*) și (2) realitatea deviază de la modelul dorit, adică jurnalul de evenimente este greșit (modelul se dorește a fi *normativ*).

Verificarea conformității este folosită pentru *aliniera afacerii* și pentru *audit*. Totodată, aceasta este folosită și pentru evaluarea performanței algoritmilor de descoperire a proceselor, pentru conformitatea cu specificațiile (unui software, serviciu etc.) și pentru a repara modelele care nu sunt aliniate cu realitatea.

Alinierea afacerii, îndeosebi cu sistemele informatice, se referă la sincronizarea afacerii cu suportul oferit de sistemele IT [86]. Modul în care se derulează procesele de afaceri în prezent este o consecință a evoluției, sau mai bine spus a revoluțiilor tehnologice în zona informației și a comunicării. Managerii și specialiștii din domeniul IT recunosc în consens că dimensiunea sau componenta tehnologică a unei afaceri este un factor determinant al succesului acesteia. În acest context, se observă că mijloacele software pe care o afacere le are la dispoziție sunt rareori personalizate. Astfel, afacerea este „încorsetată” de aceste mijloace software, existând o nevoie stringentă ca procesele de afaceri să beneficieze rapid de un suport tehnologic adecvat, întrucât climatul în care se desfășoară afacerea este în continuă schimbare. Ca urmare, procesele sunt încetinite pentru că suportul lor IT nu apare la timp sau nu este adecvat. În acest context, *Process Mining*, prin etapa de verificare a conformității, sprijină demersul de cunoaștere exactă a proceselor și a nevoilor actuale ale afacerii și astfel permite o mai bună reacție de răspuns pentru a alinia mai bine afacerea, prin prisma proceselor sale, la sistemele informatice. Verificarea conformității poate fi făcută atât offline, cât și online, adică la rulare, conducând la generarea de alerte instantanee.

Auditul este efectuat pentru a stabili validitatea și fiabilitatea informațiilor despre organizația și procesele asociate acesteia. Scopul este verificarea dacă procesele afacerii sunt executate în anumite limite impuse de manageri, de stat sau de alți stakeholderi (părți interesate). În ultimii ani s-a pus un foarte mare accent pe verificarea conformității, mai ales din partea organismelor guvernamentale, prin impunerea unor reglementări legislative cum sunt: Sarbanes-Oxley (SUA) [34], Basel II/III (UE), J-SOX (Japonia), C-SOX (Canada) ș.a. De asemenea, standardul ISO 9001:2008, revizuit pentru 2015 [90], impune organizațiilor să-și modeleze procesele operaționale. Multe organizații au implementat acest standard doar la nivelul documentației, dar în realitate procesele de producție sunt foarte diferite de procesele modelate. Astfel, verificarea conformității este un bun suport pentru audit.

Tehnicile cele mai cunoscute pentru verificarea conformității sunt: folosirea amprentelor de cauzalitate, folosirea reluării bazate pe jetoane și verificarea conformității bazată pe aliniere. Acestea vor fi prezentate în continuare.

3.5.1. Verificarea conformității folosind amprente de cauzalitate

Conform [8], metoda se referă la matricea amprentei ce derivă din aplicarea algoritmului α (prezentat în capitolul 3.4.1). Dacă amprenta jurnalului de evenimente și modelul coincid, atunci situația este o potrivire perfectă, iar conformitatea bazată pe amprentă este 1. Cuantificarea nivelului de conformitate poate lua valori între 0 și 1 (0 este valoarea ce corespunde cu cea mai proastă potrivire, iar 1 corespunde situației de potrivire perfectă).

În acest caz, verificarea conformității are loc astfel:

- Pentru fiecare model obținut se va crea matricea amprentei, iar aceasta se va compara cu matricea amprentei obținută cu algoritmul α .
- Se măsoară numărul de diferențe și se raportează la numărul total de celule din tabelă. Rezultatul este scăzut din 1 (ce reprezintă potrivirea perfectă). Valoarea obținută astfel este folosită la evaluarea modelului. Formula de calcul a indicelui conformității bazat pe amprente ($C_{amprentă}$) este:

$$C_{amprentă} = 1 - \frac{\text{Număr diferențe}}{\text{Număr total celule matrice}} \quad (3.21)$$

Să notăm rețeaua Petri din Fig. 2.1 cu P_1 și să considerăm o altă posibilă rețea Petri P_2 , descoperită tot din jurnalul de evenimente L și reprezentată în Fig. 3.10. Matricele amprentei pentru rețelele Petri P_1 și P_2 sunt reprezentate în Tabelul 3.5, iar în Tabelul 3.6 avem diferențele dintre cele două matrici ale amprentelor. Se observă că matricea diferenței conține 6 elemente diferite.

Astfel, indicele conformității bazate pe amprente pentru acest caz este:

$$C_{amprentă} = 1 - \frac{6}{49} = 0.88$$

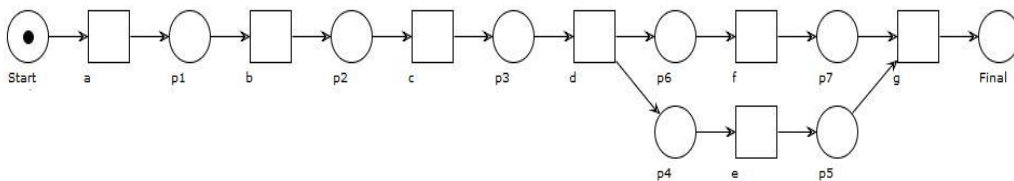


Fig. 3.10. Altă rețea Petri descoperită din jurnalul de evenimente L (P_2)

Tabelul 3.5. Matricea amprentelor rețelelor Petri P_1 și P_2

P_1	a	b	c	d	e	f	g
a	#	→	→	#	#	#	#
b	←	#	#	→	#	#	#
c	←	#	#	→	#	#	#
d	#	←	←	#	→	→	#
e	#	#	#	←	#		→
f	#	#	#	←		#	→
g	#	#	#	#	←	←	#

P_2	a	b	c	d	e	f	g
a	#	→	#	#	#	#	#
b	←	#	→	#	#	#	#
c	#	←	#	→	#	#	#
d	#	#	←	#	→	→	#
e	#	#	#	←	#		→
f	#	#	#	←		#	→
g	#	#	#	#	←	←	#

Tabelul 3.6. Matricea amprentei cu diferențele dintre rețelele Petri P_1 și P_2

P_1 vs. P_2	a	b	c	d	e	f	g
a			→:#				
b			#:→	→:#			
c	←:#	#:←					
d		←:#					
e							
f							
g							

Se poate și diagnostica modelul obținut prin analiza comparării rezultatelor din celule (ceea ce demonstrează cum ar trebui să fie față de cum este în momentul de față). Cu această metodă se pot compara modelele de proces între ele, jurnalele de evenimente între ele, dar și jurnale de evenimente cu modele de procese.

Dezavantajele acestei metode sunt următoarele:

- (i) frecvențele nu sunt folosite;
- (ii) comportamentul este considerat doar prin prisma relației de succesiune directă;
- (iii) surprinde fidel, într-o singură metrică, doar potrivirea, generalizarea și precizia, dar nu și simplitatea.

Similar, acest tip de diagnosticare a conformității se poate extinde și se pot folosi alte matrice, cum sunt matricea valorilor succesiunii directe sau matricea valorilor de dependență (prezentate în capitolul 3.4.2.1) și astfel, se va ține cont și de frecvența activităților.

3.5.2. Verificarea conformității folosind reluarea bazată pe jetoane

Metoda a fost cercetată inițial în lucrarea de dizertație a Annei Rozinat [141], iar apoi aprofundată de către aceeași cercetătoare [142, 143, 146]. Metoda are următoarea logică:

Se încearcă parcurgerea modelului de proces din jurnalul de evenimente folosind jocul cu jetoane, iar dacă parcurgerea nu poate avea loc, atunci se forțează avansarea, plasându-se un *jeton absent*. Jetoanele ce rămân în rețea după consumarea jetonului din locația finală se numesc *jetoane rămase*. Există și o metrică ce cuantifică potrivirea (*fitness*) la nivel de traseu:

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c}\right) + \frac{1}{2} \left(1 - \frac{r}{p}\right) \quad (3.22)$$

$$0 \leq fitness(\sigma, N) \leq 1 \quad (3.23)$$

unde N reprezintă rețeaua Flux de lucru, σ reprezintă traseul din N , p reprezintă numărul *jetoanelor produse* (*produced tokens*), c reprezintă numărul *jetoanelor consumate* (*consumed tokens*), m reprezintă numărul *jetoanelor absente* (*missing tokens*), r reprezintă numărul *jetoanelor rămase* (*remained tokens*). *Jetoanele absente* m sunt acele jetoane consumate în timp, ele fiind adăugate artificial. *Jetoanele rămase* r sunt acele jetoane produse, dar neconsumate, acestea rămânând în rețeaua flux de lucru după consumarea jetonului în locația de final. Potrivirea este cuantificată printr-un indice între 0 și 1, unde valorile apropiate de zero arată o potrivire foarte mică (sau deloc) dintre model și jurnal, iar o valoare egală cu unu arată o potrivire perfectă.

Există doi invarianti (proprietăți ce rămân neschimbate în cazul unei transformări):

- La orice moment în toată rețeaua flux de lucru sau în orice locație din rețea:

$$p + m \geq c \geq m \quad (3.24)$$

- La final în toată rețeaua sau în orice locație din rețea:

$$r = p + m - c \quad (3.25)$$

La inițializare, un jeton este produs pentru locația sursă, deci $p = 1$. La finalizare, un jeton este consumat din locația de final (chiar dacă acesta nu este acolo), deci:

$$c' = c + 1 \quad (3.26)$$

Potrivirea, la nivel de jurnal, are următoarele relații de determinare:

$$fitness(L, N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}}\right) \quad (3.27)$$

$$0 \leq fitness(L, N) \leq 1 \quad (3.28)$$

De remarcat este faptul că formula potrivirii la nivel de jurnal este similară cu cea la nivel de traseu, diferența constând în luarea în calcul a sumei totale a jetoanelor produse, consumate, absente și rămase din reluarea fiecărui traseu descris în jurnal, raportate la o multi-mulțime de trasee din cadrul acestuia.

De exemplu, $\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}$ se referă la numărul total de jetoane absente atunci când este reluat întregul jurnal de evenimente, $L(\sigma)$ reprezentând frecvența traseului σ , iar $m_{N,\sigma}$ reprezentând numărul de jetoane absente pentru o singură instanță a lui σ .

Să considerăm un jurnalul de evenimente $L_1 = [\langle a, b, d, e, f, g \rangle^{10}, \langle a, d, f, e, g \rangle^5, \langle a, b, c, d, e, f, g \rangle^2]$ și modelul de proces P_2 reprezentat în Fig. 3.10. Vom relua traseul $\sigma_1 = \langle a, b, d, e, f, g \rangle$ și vom număra jetoanele produse, consumate, absente și rămase (Fig. 3.11).

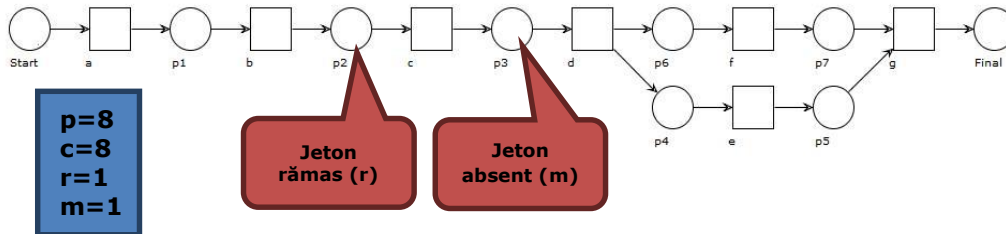


Fig. 3.11. Verificarea conformității la nivel de traseu prin metoda reluării bazate pe jetoane pentru rețeaua Petri P_2

Potrivirea la nivel de traseu pentru acest caz este:

$$fitness(\sigma_1, P_2) = \frac{1}{2} \left(1 - \frac{1}{8}\right) + \frac{1}{2} \left(1 - \frac{1}{8}\right) = 0,875$$

În mod similar, se va calcula potrivirea și pentru celelalte două trasee:

- Pentru traseul $\sigma_2 = \langle a, d, f, e, g \rangle$, avem $p = 7$, $c = 7$, $r = 1$, $m = 1$ și $fitness(\sigma_2, P_2) = 0,857$
- Pentru traseul $\sigma_3 = \langle a, b, c, d, e, f, g \rangle$, avem $p = 9$, $c = 9$, $r = 0$, $m = 0$ și $fitness(\sigma_3, P_2) = 1$, adică potrivire perfectă.

Ținând cont de frecvența activităților și de numărul de jetoane obținut pentru fiecare traseu, se poate calcula potrivirea la nivel de jurnal pentru jurnalul L_1 pentru rețeaua P_2 , astfel:

$$fitness(L_1, P_2) = \frac{1}{2} \left(1 - \frac{1 \times 10 + 1 \times 5 + 0 \times 2}{8 \times 10 + 7 \times 5 + 9 \times 2}\right) + \frac{1}{2} \left(1 - \frac{1 \times 10 + 1 \times 5 + 0 \times 2}{8 \times 10 + 7 \times 5 + 9 \times 2}\right) = 0,887$$

Potrivirea jurnalului L_1 , peste rețeaua P_2 de 0,887 poate fi interpretată în 2 moduri:

- Jurnalul L_1 are o potrivire de 0,887, adică 11,3% din evenimente deviază;
- Modelul de proces are o potrivire de 0,887, adică modelul nu poate explica 11,3% din comportamentul observat.

Limitările acestei metode sunt următoarele:

- abordarea de bază a reluării presupune tranziții vizibile și etichetate unic;
- valorile conformității sunt câteodată prea optimiste din cauza mulțimii mare de jetoane;
- luarea de decizii locale poate duce la rezultate derutante, pentru că tehnica reluării nu oferă o anumită cale prin model. O soluție în acest caz ar fi luarea în considerare, la momentul deciziei, a „celei mai scurte căi”, soluție prezentată în continuare.

3.5.3. Verificarea conformității bazată pe aliniere

Această metodă a fost introdusă și detaliată în lucrarea [12]. Alinierea dintre jurnalul de evenimente și modelul de proces este foarte importantă atunci când trebuie verificată conformitatea sau trebuie analizată performanța. Prin analiza alinierii se pot diagnostica devierile de la comportamentul modelat, iar severitatea fiecărei deviații poate fi cuantificată.

Stabilirea unei alinieri între un model de proces și un jurnal de evenimente trebuie să stabilească, implicit, o relație între deplasările din jurnal și deplasările din model. Se definește \gg ca simbol pentru lipsa de mișcare.

Fie A_L un set de activități din jurnalul L , iar mișcarea acestora este dată de relația:

$$A_L^{\gg} = A_L \cup \{\gg\} \quad (3.29)$$

Similar, A_M un set de activități din modelul M , iar mișcarea acestora este dată de relația:

$$A_M^{\gg} = A_M \cup \{\gg\} \quad (3.30)$$

De menționat că atât evenimentele, cât și activitățile au asociate etichete pentru a putea relaționa activitățile din model cu evenimentele din jurnal.

Un pas din aliniere este reprezentat de perechea $(x, y) \in A_L^{\gg} \times A_M^{\gg}$, astfel încât:

$$\bullet (x, y) \text{ este o deplasare în jurnal dacă } x \in A_L \text{ și } y = \gg \quad (3.31)$$

$$\bullet (x, y) \text{ este o deplasare în model dacă } y \in A_M \text{ și } x = \gg \quad (3.32)$$

$$\bullet (x, y) \text{ este o deplasare în amândouă (jurnal și model) dacă } x \in A_L \text{ și } y \in A_M \quad (3.33)$$

$$\bullet (x, y) \text{ este o deplasare nepermisă dacă } x = \gg \text{ și } y = \gg \quad (3.34)$$

Mulțimea tuturor deplasărilor permise este:

$$A_{LM} = \{(x, y) \in A_L^{\gg} \times A_M^{\gg} \mid x \in A_L \vee y \in A_M\} \quad (3.35)$$

Fie $\sigma_L \in L$ un traseu din jurnalul L , iar $\sigma_M \in \beta(M)$ un traseu (mai exact o secvență de execuție de la început la final) din modelul M . O *alinieri* dintre σ_L și σ_M este o secvență $\gamma \in A_{LM}$, astfel încât proiecția pe primul rând (ignorând \gg) corespunde lui σ_L , respectiv proiecția pe al doilea rând (ignorând din nou \gg) corespunde lui σ_M .

Fie traseul $\sigma = \langle a, b, c, d, f, g \rangle$ din jurnalul de evenimente L și modelul de proces P_1 (din Fig. 2.1). În Fig. 3.12 este reprezentată alinierea dintre traseul σ și o rulare a modelului de proces P_1 .

Evaluarea unei alinieri se face cu ajutorul unei funcții standard de distanță $\delta_S \in A_{LM} \rightarrow \mathbb{N}$, care consideră doar deplasările permise. Pentru orice $x \in A_L^{\gg}$ și $y \in A_M^{\gg}$ avem:

$$\delta_S(x, y) = \begin{cases} 1, & \text{dacă } x \in A_L \text{ și } y = \gg \\ 1, & \text{dacă } y \in A_M \text{ și } x = \gg \\ 0, & \text{dacă } x \in A_L \text{ și } y \in A_M \text{ și } x = y \\ \infty, & \text{dacă } x \in A_L \text{ și } y \in A_M \text{ și } x \neq y \end{cases} \quad (3.36)$$

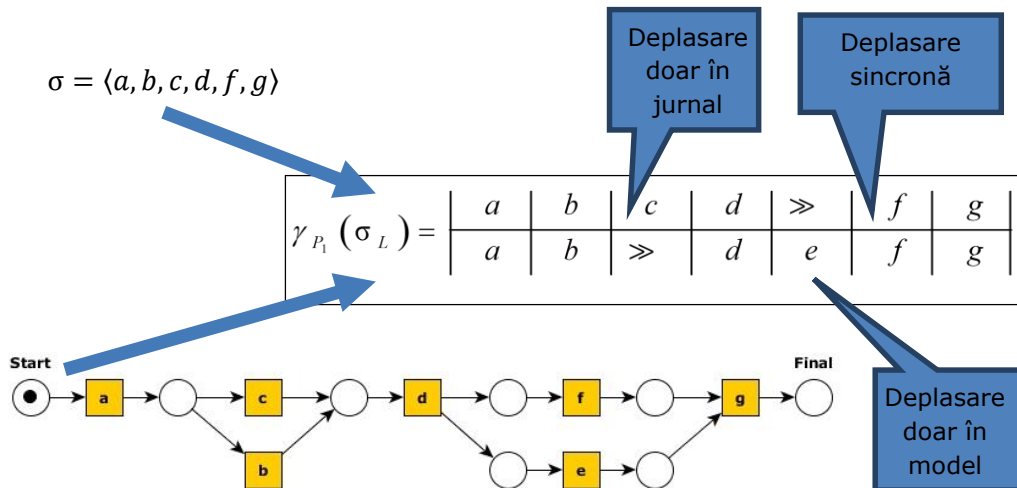


Fig. 3.12. Verificarea conformității prin alinierea unui traseu cu o rulare a modelului de proces

Pornind de la această formulă, pentru alinierea de mai sus dintre traseul σ_L și modelul P_1 , avem costul:

$$\delta_S(\gamma) = \delta_S(a, a) + \delta_S(b, b) + \delta_S(c, \gg) + \delta_S(d, d) + \delta_S(\gg, e) + \delta_S(f, f) + \delta_S(f, f)$$

$$\delta_S(\gamma) = 0 + 0 + 1 + 0 + 1 + 0 + 0 = 2$$

Costul total al alinierii dintre un jurnal de evenimente L și un model M este:

$$fcost(L, M) = \sum_{\sigma_L \in L} \delta(\gamma_M(\sigma_L)) \quad (3.37)$$

Între un anumit traseu și un model pot exista mai multe alinieri. Este important să aflăm alinierea optimă, adică acea aliniere a cărei cost asociat este minim. Orice structură de cost este posibilă. Până acum ne-am referit doar la activități similare, dar costul poate fi asociat și în funcție de resursele ce execută activitatea, de timpul cât durează aceasta sau de alte date asociate evenimentului din jurnal. În acest caz de verificare a conformității nu avem nicio restricție referitoare la notația procesului (de exemplu, sunt posibile tranziții discrete sau ca două tranziții să aibă aceeași etichetă).

Această metodă de verificare a conformității se poate folosi la diagnosticare (după cum a fost arătat), dar cu ajutorul ei se poate calcula și potrivirea. Potrivirea se calculează prin compararea costului unei alinieri optime cu cel mai prost scenariu, adică cu suma dintre numărul de evenimente care produc o deplasare doar în jurnal, notat cu $move_L(L)$, și cel mai scurt drum posibil de la starea inițială la cea finală, notat cu $move_M(M)$, conform relației

$$fitness(L, M) = 1 - \frac{fcost(L, M)}{move_L(L) + |L| \times move_M(M)} \quad (3.38)$$

Pentru traseul $\sigma = \langle a, b, c, d, f, g \rangle$ (considerăm că jurnalul are un singur traseu) și pentru modelul de proces P_1 , potrivirea este:

$$fitness(\sigma, P_1) = 1 - \frac{2}{6 + 6} = 0,833$$

Avantajele alinierii jurnalului cu modelul constau în:

- (i) comportamentul observat este direct corelat cu cel modelat;
- (ii) alinierea este foarte flexibilă, putând fi realizată după orice structură de cost;
- (iii) se pot obține diagnostice detaliate;
- (iv) după aliniere, alte atribute ale calității se pot investiga relativ ușor.

3.6. Suportul operațional detecția, predicția și recomandarea

Există patru întrebări generice referitoare la studiul datelor:

1. Ce s-a întâmplat?
2. De ce s-a întâmplat?
3. Ce se va întâmpla?
4. Care este cea mai bună variantă care se poate întâmpla?

Primele 2 întrebări sunt legate de trecut. Suportul operațional se concentrează pe întrebările generice 3 și 4. Întrebarea 3 este specifică predicției (de exemplu, se pune problema când se va finaliza un caz, sau dacă respectivul caz se va respinge), iar întrebarea 4 este specifică recomandării. Suportul operațional începe cu datele „pre mortem”.

- *Primul pas al suportului operațional este detecția sau semnalarea că un lucru nu merge bine: ceva se întâmplă acum, cazul deviază acum, termenul limită a expirat. Ca urmare, trebuie întreprinse acțiuni, măsuri imediate.*
- *Al doilea pas este predicția: va devia cazul, va fi respins cazul, când va expira termenul?*
- *Al treilea pas este recomandarea: ce activitate ar trebuie executată, cine ar trebui să execute activitatea?*

Un model de proces se poate converti automat la un *sistem de tranziție*. Pentru fiecare activitate avem un eveniment de start și un eveniment de stop. Se vor identifica și stările unei activități. Detecția funcționează astfel: o persoană lucrează la un caz într-un anumit moment. În același timp, se creează un model normativ. Ceea ce s-a executat este trimis spre analiză sistemului de suport operațional. Dacă ceea ce s-a executat și ceea ce este în modelul normativ diferă, atunci este semnalată o încălcare. Utilizatorul sau administratorul este informat că o astfel de abatere a avut loc. Încălcările ce pot fi detectate sunt de tipul: a avut loc execuția unei alte activități SAU o activitate a fost executată prea târziu ș.a.

În cazul predicției se va crea un model predictiv bazat pe analiza jurnalului de evenimente. Acesta se va folosi de către sistemul de suport operațional care, pe baza analizei cazului ce se execută, va furniza date previzionale despre acesta (cum ar fi momentul când se va termina cazul sau ce se va executa). Pentru această analiză se folosește sistemul de tranziție cu notații (pentru fiecare stare se completează timpul când s-a intrat în stare și când s-a ieșit din stare, timpul petrecut în stare, timpul rămas în stare. Pentru a obține o predicție cât mai bună, avem nevoie de a cunoaște foarte bine contextul, adică dacă mai există și alte cazuri similare care sunt în execuție sau așteaptă să fie executate.

În cazul recomandării, se pune întrebarea care este următoarea activitate care va fi executată și cine o va executa. Se creează un model de recomandare care este obținut din analiza jurnalului de evenimente, iar ceea ce s-a executat este comparat cu modelul de recomandare. O recomandare are un obiectiv specific: minimizarea costurilor, minimizarea timpului rămas, folosirea resurselor. Stările viitoare recomandate se pot adăuga la modelul de predicție și se pot crea scenarii care se pot compara. În situații practice, dacă o activitate se poate previziona, atunci ea poate fi și recomandată.

Suportul operațional este cel ce face trecerea din *Process Mining* offline în *Process Mining* online.

3.7. Alte perspective ale Process Mining

Modelele de proces pot fi comparate cu niște hărți [1]. Dacă se vizualizează o hartă, în cadrul acesteia sunt folosite abstractizări (de exemplu, drumurile nesemnificative nu sunt reprezentate), agregări (de exemplu, mici entități sunt grupate împreună sub un nume general – cartiere, orașe), mărimi și culori (evidențierea entităților importante), amplasări (poziționarea elementelor are o justificare clară). Pornind de la aceste elemente specifice hărților, se dorește adăugarea „punctelor de interes” pe modelul de proces.

O hartă este o reprezentare a realității, dar nu se confundă cu realitatea. Modelele de proces nu sunt statice, ele trebuie să prindă „viață”. Pe o hartă se pot poziționa mai multe elemente: blocaje de trafic, proprietăți de vânzare, localizarea mijloacelor de transport în comun etc. Pe un model de proces se pot adăuga: strangulări, costuri, deviații, personal etc.

Extinzând comparația modelelor cu hărțile, un sistem de navigare (GPS) este un instrument util care afișează recomandări, previziuni, detectări, exact ceea ce reprezintă suportul operațional pentru *Process Mining*. Un sistem de navigare se poate adapta la condițiile actuale și nu forțează șoferul să ia o anumită cale. Un sistem informatic pentru conducerea proceselor ar trebui să aibă aceleași caracteristici.

Cele trei tipuri de *Process Mining*, și anume descoperirea, verificarea conformității și îmbunătățirea, au fost aplicate până acum doar din perspectiva fluxului de control, așa cum este sintetizat în Fig. 3.13, dar demersul de *Process Mining* nu trebuie limitat doar la acest tip de aplicații, el putând fi utilizat cu succes și din alte perspective: organizațional, operațional, din perspectiva datelor (puncte de decizie) [17] sau altele. *Din analiza tuturor perspectivelor și prin reunirea tuturor elementelor identificate se poate obține un model integrat de proces, din perspectivă holistică* [8]. Implementarea *Process Mining* din multiple perspective se poate realiza dacă evenimentele din jurnal au atributele specifice atașate (după cum s-a precizat în secțiunea 3.3.3). În Fig. 3.14 s-a reprezentat legătura dintre atributele evenimentului și proces, din punct de vedere al formatului datelor.

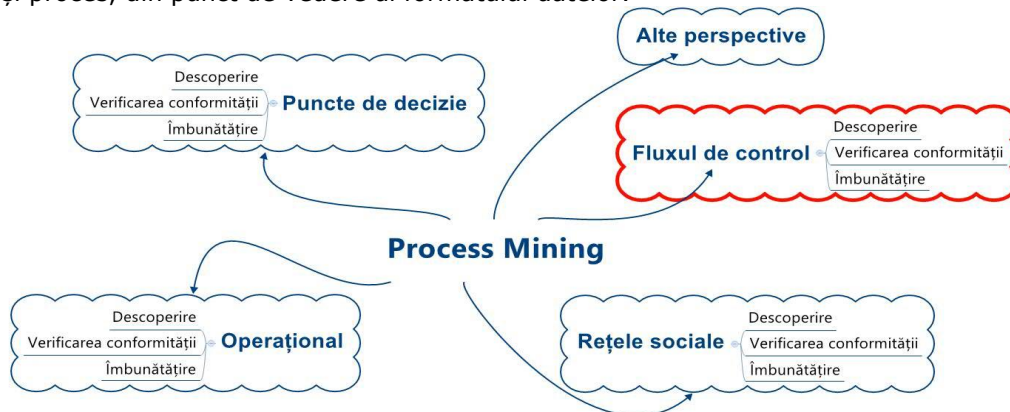


Fig. 3.13. Perspective ale Process Mining

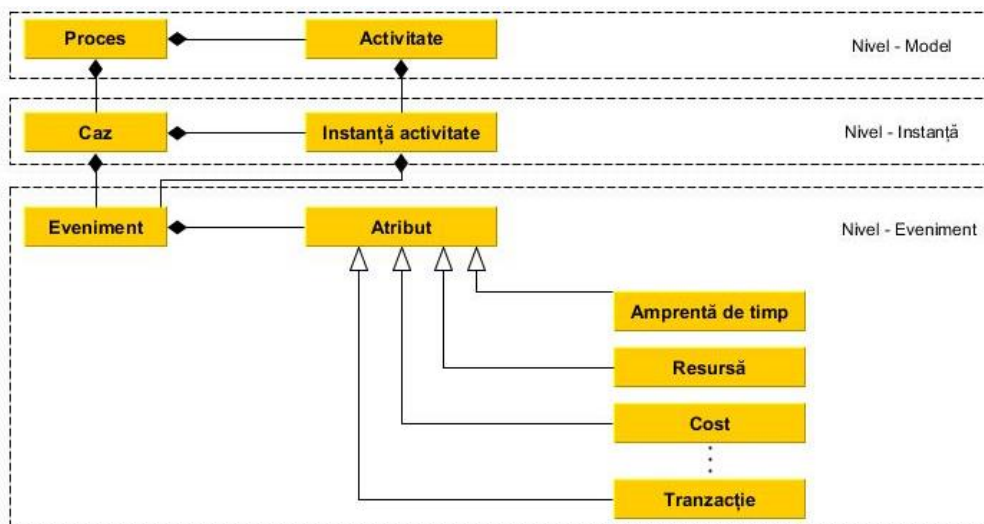


Fig. 3.14. Legătura dintre atributele de eveniment și proces

În concluzie, modelele de proces sunt precum hărțile geografice; prin adăugarea de elemente cartografice specifice pot fi îmbunătățite modelele de proces; adesea sunt necesare mai multe hărți, care trebuie conectate cu date reale. Prin analogie, sistemele informatice pentru conducerea proceselor se aseamănă cu instrumentele de navigare.

3.7.1. Mining-ul punctelor de decizie

Legat de îmbunătățirea proceselor, amintim două modalități de realizare a acestora [8]:

- *extensia* – adăugarea de perspective adiționale la model și folosirea datelor din evenimente;
- *repararea* – îmbunătățirea calității modelului folosind datele din evenimente.

Pentru îmbunătățirea proceselor prin Mining-ul punctelor de decizie este nevoie de jurnalul de evenimente și modelul de proces. Demersul pornește de la presupunerea că jurnalul de evenimente și modelul de proces sunt aliniate. În acest caz, se dorește descoperirea de tipare pentru prezicerea căii de urmat în punctele de decizie. În diagramele BPMN, există următoarele puncte de decizie: decizia exclusivă (XOR-split gateway), decizia inclusivă (OR-split gateway), bifurcarea paralelă (AND-split gateway) și condiția complexă. Punctele de decizie se mai numesc și portaluri¹⁴. Trebuie remarcat că bifurcarea paralelă nu este considerată punct de decizie, deoarece activitățile următoare se vor executa amândouă, doar ordinea lor de execuție fiind cea care diferă. De asemenea, trebuie precizat că rețelele Petri nu pot explica decizia inclusivă (XOR-split), dar această decizie este suportată de terminologia BPMN sau YAWL. În cazul deciziei inclusive se consideră trei variante de ieșire și anume: activitatea A este urmată de B, sau de C, sau de B și C (Fig. 3.15).

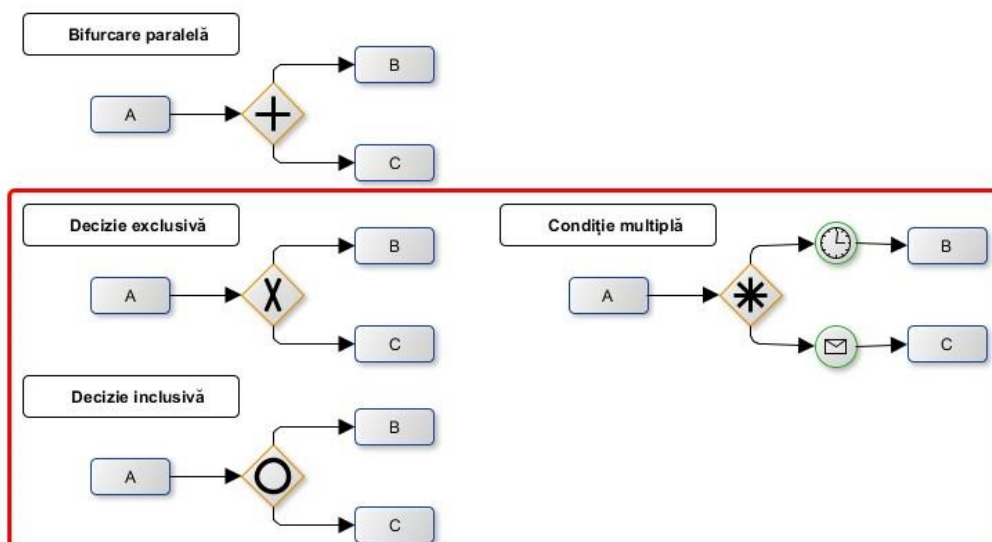


Fig. 3.15. Puncte de decizie în BPMN

¹⁴ Terminologia BPMN a fost adaptată în română în cadrul proiectului CertiBPM [63].

Mining-ul punctelor de decizie sau Mining-ul din perspectiva cazurilor se concentrează pe proprietățile cazurilor și are ca scop găsirea regulilor ce pot explica aceste decizii [144, 145]. Sunt folosite tehnici de clasificare precum învățarea bazată pe arbori de decizie (prezentată în 2.2.2.4). De reținut că intrarea pentru învățarea bazată pe arbori de decizie este un tabel în care fiecare rând definește o variabilă categorială de răspuns (de exemplu, activitatea aleasă) și mai multe variabile predictor (de exemplu, proprietăți ale clientului). Arborele de decizie își propune să explice variabila de răspuns din punctul de vedere al variabilelor predictor. În analiza deciziilor, sunt considerate atribute ale evenimentelor anterioare și ulterioare. Reluarea este folosită aici pentru identificarea rezultatului punctului de decizie și pentru a calcula probabilitatea fiecărei ramificații.

Dezavantajul Mining-ului decizional (după cum este susținut în [145]) este acela că nu sunt descoperite condițiile asociate cu decizii exclusive și cu multe bucle, iar jurnalul de evenimente este obligatoriu să fie conform cu modelul descoperit, ceea ce în realitate se întâmplă relativ rar.

Progresele recente în verificarea conformității pot fi utilizate pentru a alinia un jurnal de evenimente cu date și un model de proces cu punctele de decizie. Aceste alinieri pot fi folosite pentru a genera o problemă de clasificare bine definită per punct de decizie [104]. În acest fel, pot fi descoperite fluxul de date și gărzi ce se pot adăuga la modelul procesului. Această tehnică [104] este independentă de orice notație folosită (BPMN, EPC), iar algoritmul care stă la bază este implementat cu rețele Petri de date. În termeni Petri, garda reprezintă o condiție care trebuie să fie adevărată pentru ca o tranziție să poată fi permisă. Descoperirea gărzilor tranzițiilor asociate cu un punct de decizie poate fi tradus în problema de a găsi cel mai bun estimator pentru o funcție, folosindu-se algoritmul C4.5 specific Mining-ului de date și învățării automate. Dezavantajele algoritmului ar fi că gărzile descoperite descriu mai degrabă ceea ce s-a întâmplat (adică sunt descriptive), decât ceea ce trebuia să se întâmple (adică nu sunt predictive).

Variabilele predictor pot fi raportate la diferite contexte [14], cum ar fi:

- contextul instanței de proces (de exemplu, mărimea comenzii, tipul de client);
- contextul procesului (de exemplu, numărul de resurse alocate procesului, numărul de cazuri aflate în curs de execuție);
- contextul social (de exemplu, rețele sociale, competiție internă);
- contextul extern (de exemplu, vremea, climatul economic, schimbări în legislație).

În concluzie, se poate afirma că Mining-ul punctelor de decizie este o îmbinare fructuoasă dintre *Data Mining* și *Process Mining*.

3.7.2. Mining-ul operațional

Mining-ul operațional sau demersul de Process Mining din perspectiva timpului se ocupă atât cu detaliile temporale ale evenimentelor, cât și cu frecvența lor. Prezența amprentelor de timp în jurnalul de evenimente permite descoperirea ștrangulărilor (locurilor înguste sau *bottlenecks*) analiza duratei activităților, monitorizarea utilizării resurselor, precum și predicția timpului de finalizare pentru cazurile în curs de execuție. Pentru ca Mining-ul operațional să aducă cât mai multă valoare este necesar ca pentru fiecare activitate să se cunoască momentul de început, respectiv momentul de finalizare. Ca demers de analiză se va utiliza reluarea jurnalului de eveniment ținându-se cont de amprentele de timp [8], iar statisticile sunt culese pe măsura progresului reluării evenimentelor asupra modelului. În [8], au

fost identificate următoarele metrice folositoare din punctul de vedere al perspectivei temporale:

- *Durata unui caz* (diferența dintre timpul de final al ultimei activități și timpul de start al primei activități) cu variantele: durată minimă, durată maximă și durată medie;
- *Durata unei activități* (diferența dintre timpul de final și timpul de începere a unei activități) cu variantele: durată minimă, durată maximă și durată medie. De asemenea, se poate calcula și durata totală ce reprezintă suma tuturor activităților de un anumit tip prezente în jurnalul de activități;
- *Durata de așteptare dintre 2 activități succesive* (diferența dintre timpul de începere al unei activități și timpul de final al activității precedente) cu variantele: durată minimă, durată maximă și durată medie. Această metrică este foarte utilă în determinarea strangulărilor, a elementelor de blocaj al unui proces;
- *Utilizarea resurselor* (distribuția în timp a activităților executate de anumite persoane), punând în raport duratele în care sunt executate activități cu duratele timpilor de așteptare dintre activități;
- *Durata și frecvența ciclurilor* (anumite activități sunt executate de mai multe ori în același caz).

Activitățile care sunt executate în timpul unui proces pot avea caracter tranzacțional, acest lucru însemnând că ele se pot supune ciclului de viață al unei tranzacții. Astfel, pe lângă evenimentele de start și finalizare, o tranzacție poate avea următoarele evenimente: suspendare, continuare, terminare, alocare ș.a. Pentru fiecare astfel de eveniment, în cadrul aceleiași activități, se pot crea noi metrice, cum ar fi durata de suspendare (diferența dintre suspendare și continuare).

Dacă revenim la analogiile dintre modele de proces – hărți, BPMS – sisteme de navigație GPS, observăm că un GPS oferă timpul estimativ de ajungere la destinație, deci pentru un BPMS, un aspect important îl oferă estimarea timpului rămas pentru cazurile care sunt în curs de execuție [7]. Estimarea timpului rămas pentru executarea unui proces este un subiect foarte recent, abordări teoretice și practice fiind prezentate în [19, 30, 136, 139].

Studiul aprofundat al unui model de proces din perspectiva timpului poate conduce la identificarea de tipare temporale [101] grupate în patru categorii, după cum este sintetizat în Fig. 3.16. Aceste tipare sunt utile în momentul în care se analizează modelul de proces, pentru îmbogățirea lui cu date importante.

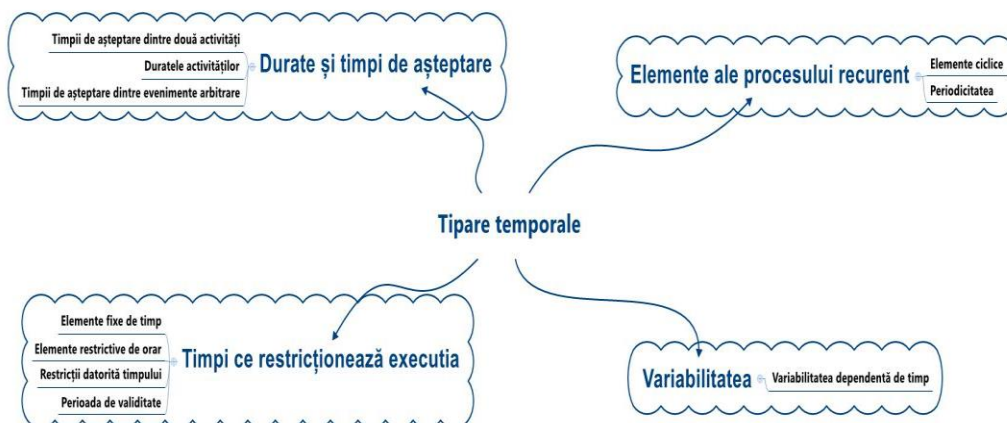


Fig. 3.16. Tipare temporale utile Mining-ului operațional

Process Mining din punctul de vedere al timpului este important pentru identificarea problemelor de performanță. Până acum procesul s-a analizat doar din punct de vedere funcțional și nu s-a pus problema performanței. Astfel, performanța procesului subliniază importanța acestei perspective de Mining.

3.7.3. Mining-ul rețelelor sociale

Alte două perspective interesante pentru *Process Mining* sunt Mining-ul rețelelor sociale și Mining-ul organizațional. Cele două perspective sunt foarte similare, amândouă procesând informațiile din jurnalele de evenimente referitoare la resursele care execută activități [16]. Diferența constă în faptul că în cadrul Mining-ului rețelelor sociale ne referim la resursa privată la nivel de individ, pe când în cadrul Mining-ului organizațional, resursele pot fi un grup de indivizi, o echipă, un departament, rolul sau funcția indivizilor ș.a.

Folosind informațiile disponibile despre resurse, există tehnici pentru a afla mai multe detalii despre oamenii implicați în proces, mașini, structuri organizatorice (roluri și departamente), distribuția muncii (alocarea de sarcini) și modele de lucru. Aceste rețele sociale sunt inspirate din sociometrie sau sociografie, științe care se ocupă cu prezentarea relațiilor dintre indivizi/grupuri sub formă grafică sau tabelară. Jacob Levy Moreno a folosit aceste metode în anii 1930 pentru a plasa studenții în cămine [8].

Nodurile într-o rețea socială corespund entităților organizaționale. Arcele reprezintă relațiile dintre entitățile organizaționale. Atât nodurile, cât și arcele pot avea ponderi, reprezentând importanța lor: cu cât ponderea este mai mare, cu atât nodul/arcul este mai important. De asemenea, pentru un arc, distanța este definită ca inversul importanței, astfel încât cu cât distanța este mai mică între noduri, cu atât nodurile sunt mai apropiate ca importanță. În Fig. 3.17 se prezintă exemplul unei rețele sociale cu anumite persoane (noduri) și relații dintre acestea (arce ponderate). Cu cât ponderile sunt mai mari, cu atât arcele sunt mai groase și nodurile mai mari, iar cu cât ponderile sunt mai mici, cu atât arcele sunt mai subțiri și nodurile mai mici. Folosind aceste reprezentări grafice putem observa ușor care persoană este mai importantă sau care relații sunt mai importante.

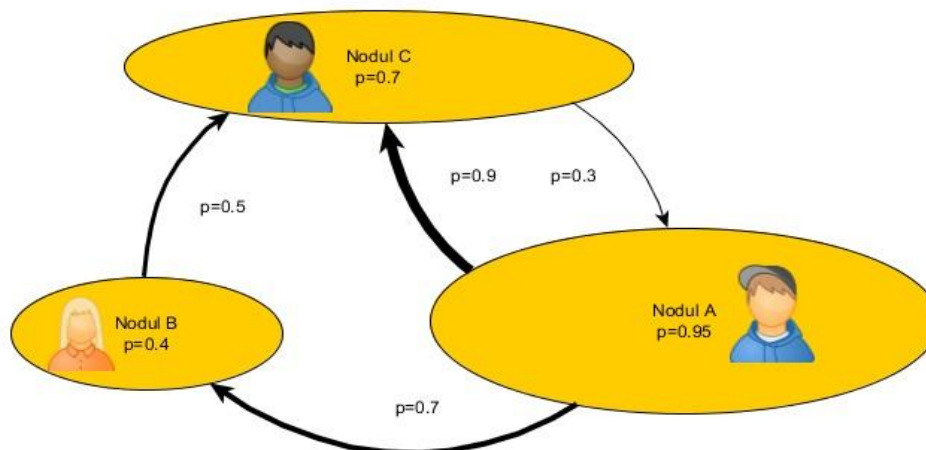


Fig. 3.17. Rețea socială cu noduri și arce ponderate

Pentru o rețea socială sunt importante anumite metrici, precum: centralitatea (suma tuturor căilor geodezice dintr-un graf raportată la suma tuturor căilor geodezice către și de la un anumit nod; gradul de centralitate reprezintă numărul de conexiuni pe care un anumit nod îl are), apropierea (inversul tuturor sumelor distanțelor geodezice către un anumit nod) și intermedierea (raportul unor căi geodezice ce vizitează un anumit nod) [79, 153].

Din analiza rețelelor sociale se pot identifica clicile (grupurile de indivizi ce relaționează împreună) și brokerii (indivizi care facilitează legăturile între clici).

Mining-ul în rețelele sociale se realizează prin crearea *matricei de resurse-activități* (exemplu prezentat în Tabelul 3.7), ce reprezintă media numărului de activități pe care o resursă o execută într-un caz [155]. Matricea de resurse-activități oferă informații despre "cine ce execută".

Se pot crea rețele sociale după transferul lucrului de la o persoană la alta. *Analiza transferului de lucru* se face cu ajutorul *matricei de transfer al lucrului* (exemplu prezentat în Tabelul 3.8). Celulele matricei reprezintă frecvența cu care lucrul este transferat de la o resursă către alta, ținând cont de media dintr-un caz. Dependențele cauzale în modelul de proces sunt folosite pentru a număra cazurile de transfer al lucrului între resurse din jurnalul de evenimente.

Din matricea de transfer al lucrului se poate crea *graful de transfer al lucrului*, în care se poate vedea cui îi dă de lucru o anumită persoană (exemplul prezentat în Fig. 3.18). Din graf se observă că unele arce sunt mai îngroșate, ceea ce arată că există o relație mai puternică între acele persoane, adică o anumită persoană are preferință să îi dea de lucru unei alte persoane în detrimentul celorlalte. De asemenea, și nodurile (persoanele) au diferite dimensiuni, arătând importanța lor în acea rețea socială (de exemplu, o persoană este mai importantă dacă are un număr mare de relații cu celelalte persoane).

Rețelele sociale sunt importante pentru organizație deoarece arată structurile de comunicare (formale și informale) în întreprinderi. Acest lucru poate fi folosit pentru a proiecta infrastructura de comunicații sau pentru realizarea planurilor de amplasare a birourilor [155].

Tabelul 3.7. Exemplu de matrice resurse-activități

	Activitatea A	Activitatea B	Activitatea C	Activitatea D
Ana	1	0.25	0	0
Bogdan	0	0	0.5	0.25
Corina	0	0	0.5	0.25
Daniel	0	0.5	0	0
Marcel	0	0.25	0	0

Tabelul 3.8. Exemplu de matrice de transfer al lucrului

	Ana	Bogdan	Corina	Daniel	Marcel
Ana	0	0.4	0	0	0
Bogdan	0.4	0	0.5	0.25	0
Corina	0	0	0.5	0.25	0.3
Daniel	0	0	0	0	0.9
Marcel	1	0.25	1.2	0	0

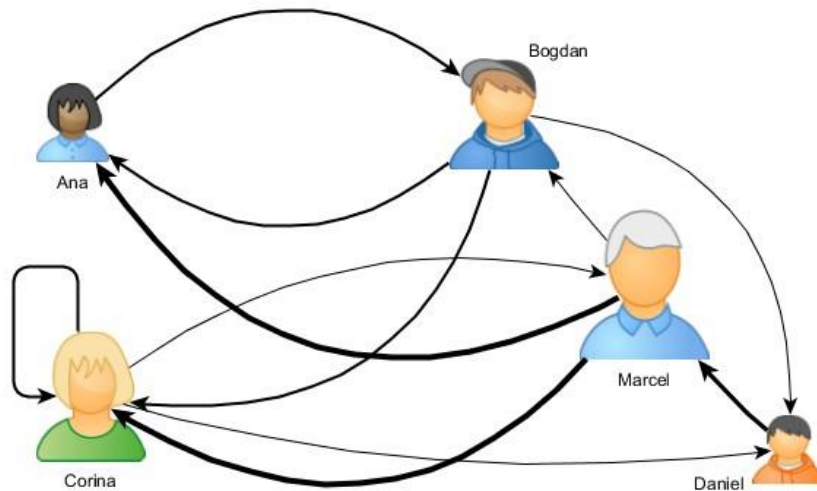


Fig. 3.18. Rețea socială reprezentând transferul de muncă

Grafurile de transfer al lucrului permit realizarea unor anumite analize:

- Se pot identifica persoanele cheie din cadrul procesului;
- După numărul de conexiuni și ponderea lor, se pot identifica rolurile acelor persoane implicate (manager, asistenți, personal executiv etc.) [16];
- Dacă grupăm persoanele cu același rol, putem obține un graf de grupuri;
- Se poate observa și volumul de lucru pe care o anumită persoană/categorie de persoane îl execută;
- Se pot identifica cliți (grupuri) și persoanele de legătură dintre acestea (brokeri);
- Se poate studia similaritatea dintre resurse, fiecărei resurse fiindu-i atașat un profil (fiecare rând din tabelul de transfer al lucrului poate fi considerat un vector cu valori caracteristice unei persoane).

3.7.4. Mining-ul organizațional

Aprofundând problema Mining-ului rețelelor sociale, apare întrebarea cum să se adauge perspectiva organizațională/de resurse la modele de proces folosind datele din jurnalele de evenimente?

Comportamentul unei resurse poate fi caracterizat printr-un profil, adică un vector care indică cât de des fiecare activitate a fost executată de respectiva resursă. Acest vector se poate obține din matricea de resurse-activități. Asupra profilului se pot aplica tehnici de *Data Mining*, de tip clustering (prezentat în capitolul 2.2.2.1). Altfel spus, aceste date se pot îmbogăți cu noi informații (de exemplu, vârstă, gen, nivel salarial), iar apoi se pot grupa folosind tehnicile de clustering. După ce grupurile s-au format, acestea pot fi relaționate cu activitățile din proces. Tehnica de tip clustering se poate folosi pentru a grupa cazurile similare (variantele de proces) sau pentru a grupa resursele similare (identificarea rolurilor).

Totodată, având creat câte un vector, cu ponderea de activități pentru fiecare resursă, se pot calcula distanțe între vectori (de exemplu, distanța Euclidiană, distanța Manhattan, distanța Minkowski și coeficientul de corelare al lui Pearson). Astfel, se

pot crea grafuri bazate pe similaritatea profilelor, resursele ce execută aceleași activități fiind înrudite.

Se pot pune și alte întrebări pentru Mining-ul resurselor, și anume: când sunt resursele disponibile (ținând seama de vacanțe, concedii medicale, muncă part-time, distribuția sau încărcarea resurselor între diferite procese sau proiecte), care resurse colaborează bine, care resurse se descurcă bine pe anumite activități (în termenii eficienței și/sau performanței) ș.a. În acest caz, se poate aplica legea de excitare a lui Yerkes-Dodson, potrivit căreia performanța poate crește proporțional cu stresul până la un punct de maxim. Dacă se mărește stresul după acest punct, performanța va scădea. În funcție de această lege se va alege nivelul de încărcare corespunzător unei resurse [108].

Așadar, modelul de proces poate fi extins (îmbunătățit) cu perspectiva organizațională prin adăugarea de informații referitoare la grupuri, departamente, rolurile grupurilor ș.a. Prin reluare jurnalului de evenimente asupra modelului se poate analiza fluxul de informație dintre grupuri și se poate analiza performanța organizațională [155].

3.7.5. Combinarea diferitelor perspective

Din cele prezentate asupra perspectivelor de *Process Mining*, se poate concluziona că atât procesele, cât și oamenii sau alte resurse sunt în legătura foarte strânsă (puternică), iar o analiză asupra proceselor trebuie realizată în manieră holistică. Problema aplicării eficiente a demersului de *Process Mining* este similară cu cea prezentată de poemul „*The Blind Men and the Elephant*” al lui John Godfrey Saxe, inspirat din povestea indiană „*Oamenii orbi și elefantul*” (Fig. 3.19), în care fiecărui orb i s-a dat să observe o parte a elefantului și fiecăruia i s-a părut că elefantul arată într-un anumit fel, dar niciunul nu l-a putut vedea în ansamblu.

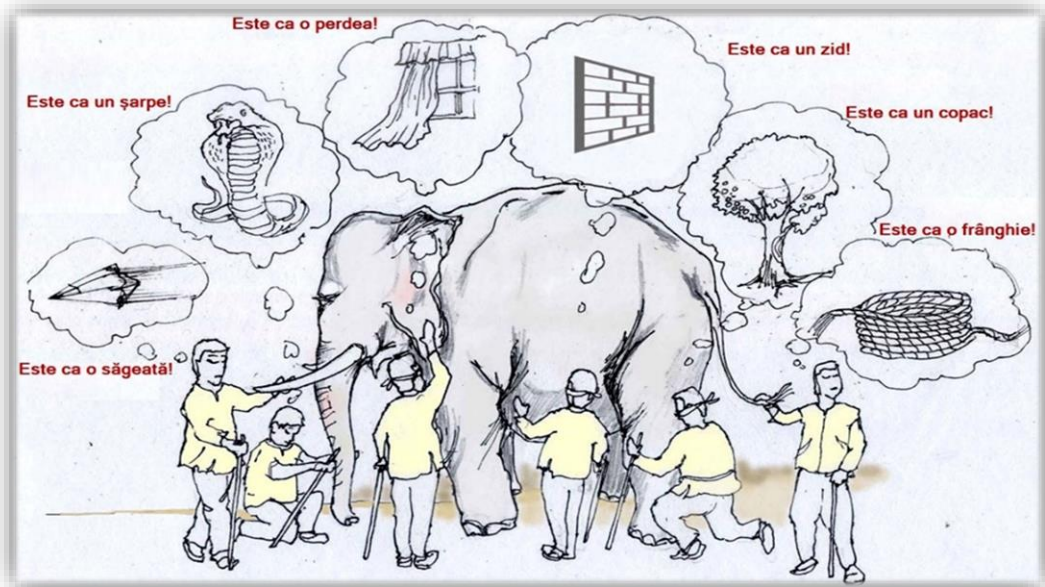


Fig. 3.19. Cum percep oamenii orbi un elefant

Concluzia poemului extrapolată la cazul *Process Mining* este că folosirea doar a unei perspective în abordarea unui sistem este echivalentă unui demers de studiu, analiză în izolare a acestuia, iar rezultatul obținut este foarte departe de realitate. Astfel, când se dorește studiul unui proces, perspectivele lui sunt fluxul de control, datele, resursele, timpul, riscurile ș.a., dar „motorul principal” pentru studierea procesului este fluxul de control.

Din studiul procesului actual (AS-IS, așa cum este) se dorește migrarea către procesul dorit (TO-BE, cum ar trebui să fie sau cum ar fi de dorit), iar modelul trebuie să fie adecvat, pretabil pentru simulare, adică asupra acestuia poate fi aplicată activitatea repetitivă de „play-out” pentru o mai bună înțelegere a sa.

Simularea poate fi folosită pentru a explora diferite alternative de design. Se pot studia efectele producerii unei concurențe mai mari a procesului sau efectele adăugării de resurse. Simularea are nevoie de modelul procesului și de mediul de execuție, dar *Process Mining* furnizează aceste intrări necesare simulării.

În Fig. 3.20 sunt descrise provocările unui proiect de simulare și riscurile aferente fiecărei etape. De asemenea, folosind demersul de *Process Mining*, simularea nu este similară cu abordarea tradițională, după cum este sintetizat în Fig. 3.21.

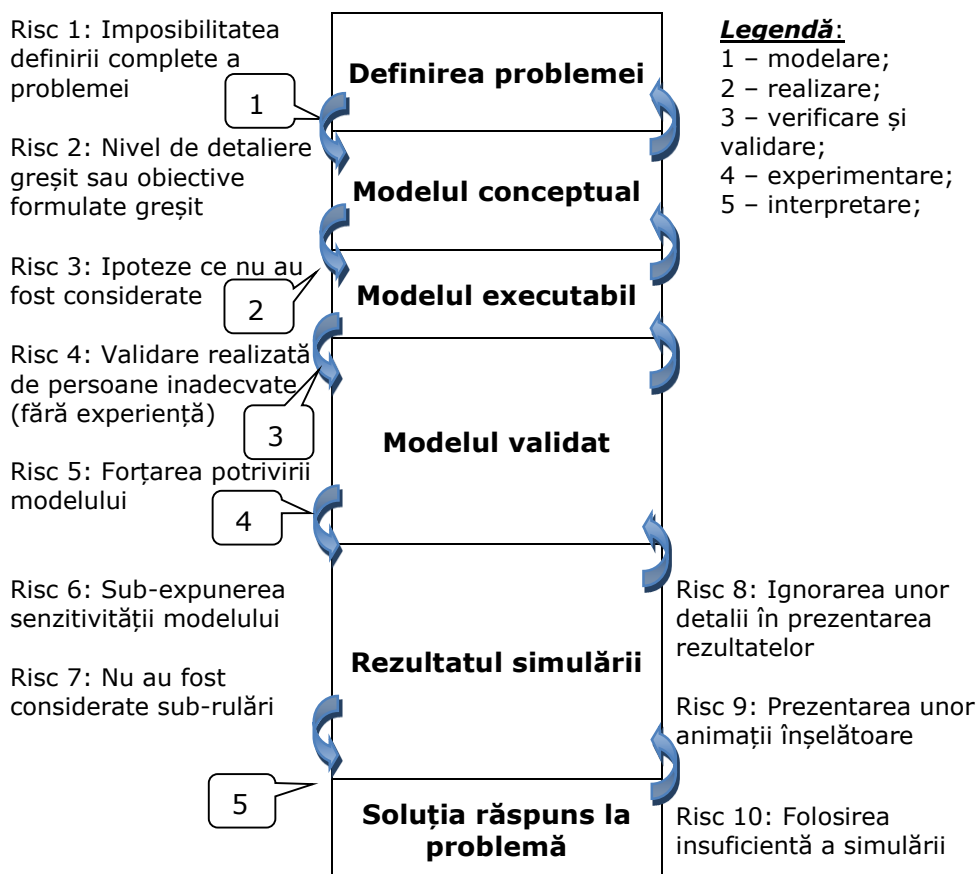


Fig. 3.20. Provocările unui proiect de simulare și riscurile asociate fiecărei etape

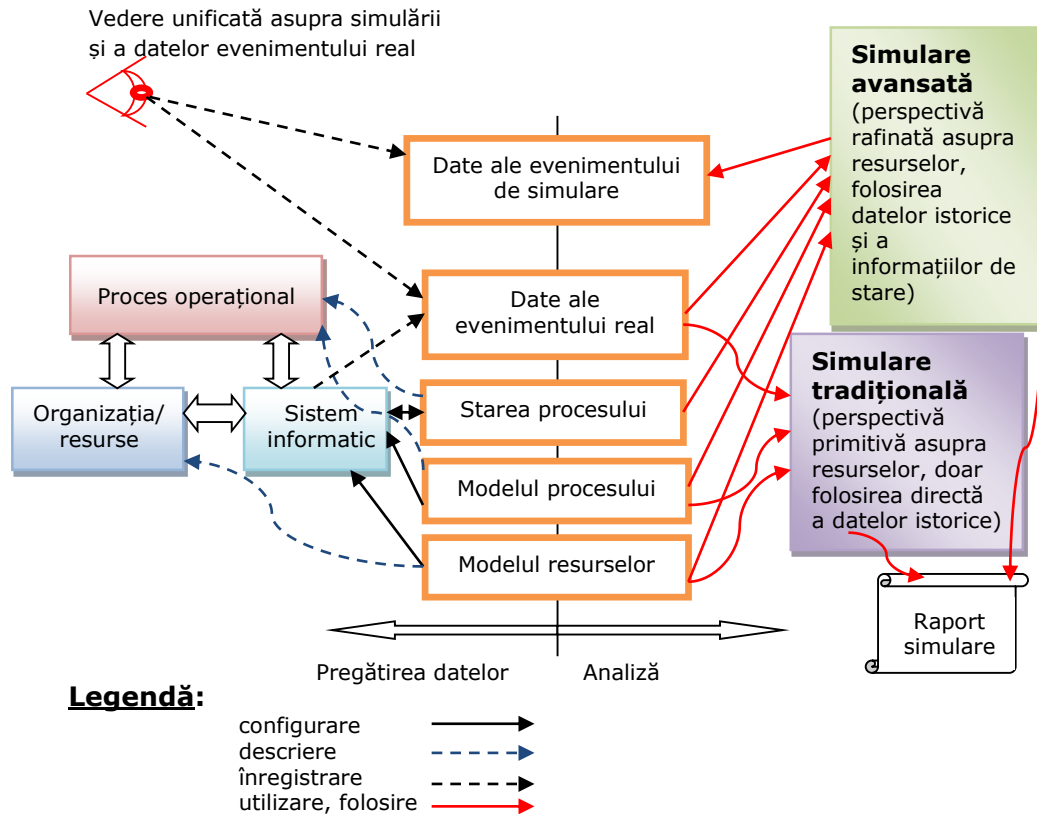


Fig. 3.21. Comparație între simularea tradițională și cea pe bază de Process Mining

3.8. Cadru de lucru pentru Process Mining rafinat

Obiectivul cadrului de lucru pentru *Process Mining* rafinat este de a furniza o panoramă pentru diferite activități de *Process Mining*. În Fig. 3.22 se prezintă alte direcții ce se pot aplica în *Process Mining*, urmând a explica în continuare modul de realizare a acestora pe baza cadrului de lucru pentru *Process Mining* rafinat (Fig. 3.23).

Datele de la care se începe *Process Mining* trebuie cunoscute din punct de vedere al provenienței; ele trebuie să fie de încredere, sigure și în stare să reproducă istoria proceselor derulate în cadrul organizației. Datele trebuie împărțite în pre-mortem și post-mortem (Fig. 3.23).

- *Evenimentele post-mortem* sunt acelea ale căror cazuri sunt complet încheiate (aceste date pot fi folosite pentru îmbunătățirea procesului și auditului, dar nu pot fi folosite pentru influențarea cazurilor la care se referă).
- *Evenimentele pre-mortem* se referă la acele cazuri care nu sunt încă încheiate, acestea încă rulează (se execută, derulează în organizație), iar informația din jurnal referitor la aceste cazuri poate fi exploatată pentru a asigura manipularea corectă și eficientă a cazurilor.



Fig. 3.22. Alte direcții aplicate în Process Mining

În partea de jos a modelului (Fig. 3.23) sunt prezentate modelele de proces care sunt extinse cu detalii obținute din mai multe perspective. Există două categorii de modele: „*de jure*” și „*de facto*”. Un model *de facto* este descriptiv, are rolul de a descrie realitatea așa cum este, fiind folosit la capturarea procesului „AS- IS”, util în cazurile de re-inginerie și suport operațional. Un model *de jure* este normativ, având rolul de a descrie realitatea cum ar trebui să fie. De exemplu, un model de proces folosit la configurarea unui BPMS este normativ, pentru că forțează oamenii să lucreze într-un anumit mod (stabilit prin BPMS). Cadru de lucru (Fig. 3.23) furnizează zece activități de *Process Mining*, grupate în trei categorii: cartografiere, audit și navigare.

În prima categorie, modelele de proces sunt văzute ca niște hărți. Cartografia conține activități de descoperire (extracția modelelor de proces), de îmbunătățire (când modelele de proces existente pot fi corelate cu jurnalul de evenimente, atunci este posibilă îmbunătățirea lor prin extindere și reparare) și de diagnoză (nu se folosește jurnalul de evenimente, ci se realizează prin analiza clasică bazată pe modele) (Fig. 3.23).

În cea de-a doua categorie (auditul) se realizează compararea modelului cu realitatea. Auditul conține activitățile de detectare (compară modelele *de jure* cu date curente pre-mortem; dacă o regulă este violată, o alertă este generată instantaneu, caracteristic pentru online), de verificare (indică deviațiile și cuantifică nivelul de conformitate, caracteristic pentru offline), de comparare (compară modelele *de facto* cu modelele *de jure* pentru a vedea în ce măsură realitatea deviază de la ce s-a planificat) și de promovare (elemente din modelul *de facto* vor fi promovate într-un model *de jure*) (Fig. 3.23).

Ultima categorie (navigarea) se referă la suportul și ghidajul execuției de proces. Navigarea conține activități de explorare (folosirea combinației dintre datele din evenimente și modele pentru a explora procesele de afaceri la execuție), de previzionare (folosirea combinației dintre informațiile despre cazurile ce rulează și modele fiind posibil să se facă previzionări pentru viitor) și de recomandare (informațiile folosite la previzionare pot fi folosite și pentru a recomanda acțiuni viitoare potrivite) (Fig. 3.23).

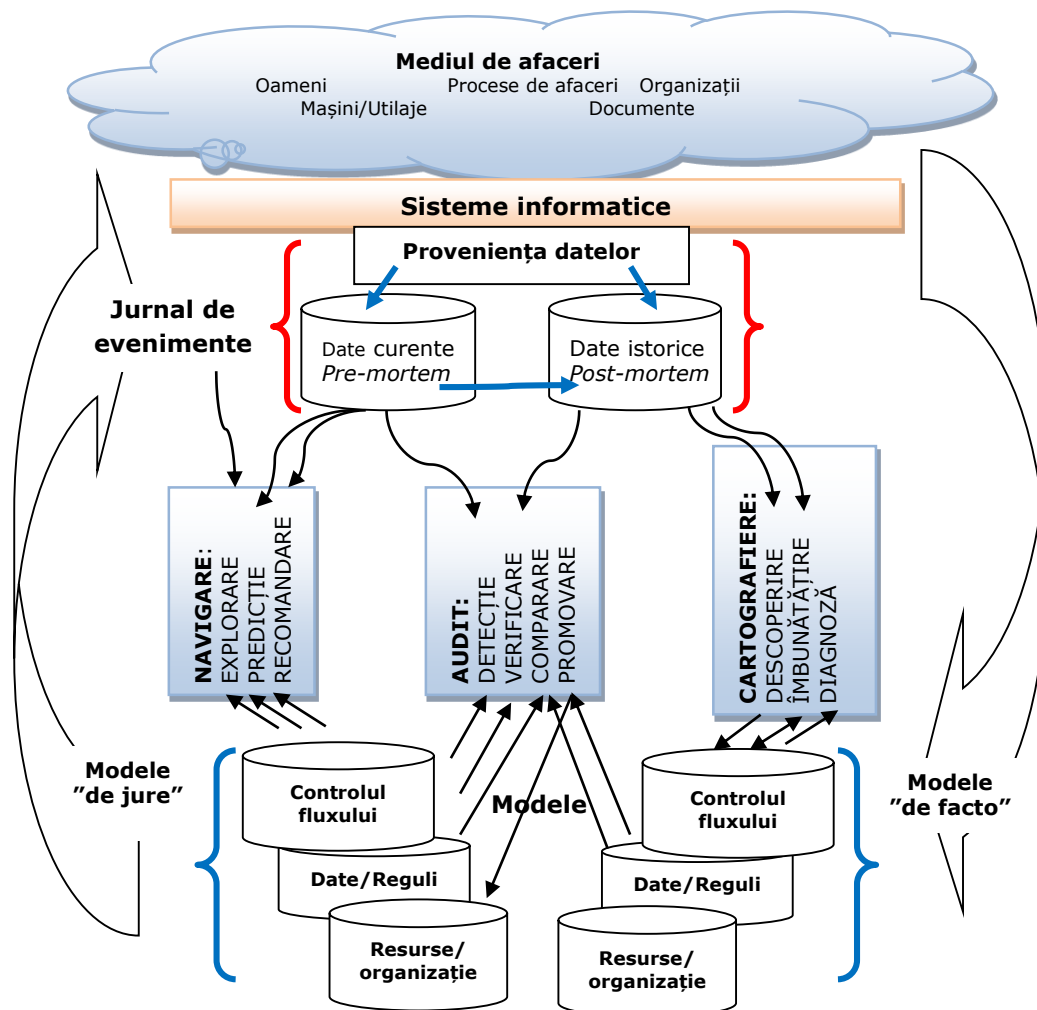


Fig. 3.23. Cadrul de lucru pentru Process Mining rafinat

3.9. Clasificarea modelelor obținute

Pentru clasificarea modelelor obținute se folosesc ca referințe procesele de tip Lasagna și de tip Spaghetti [8]. Procesele Lasagna sunt procesele structurate, cu comportament regulat, controlabile și repetitive. Procesele Spaghetti sunt procesele nestructurate, cu comportament neregulat, flexibile și foarte variabile. Aceste două categorii sunt extreme, de obicei un proces se află undeva pe acest spectru, variind între foarte nestructurat și foarte structurat.

3.9.1. Procese Lasagna

Un proces este numit Lasagna dacă este posibilă crearea cu efort limitat a unui model care are coeficientul de potrivire cel puțin 0.8, adică peste 80% din

evenimente se produc conform planului și părțile interesate confirmă validitatea modelului. Procesele Lasagna sunt foarte ușor de descoperit și este foarte puțin interesant pentru prezentarea procesului „real”, care este foarte apropiat de așteptări. Toate tehnicile (prezentate în capitolul 2.2) de *Data Mining* se pot aplica, inclusiv tehnici avansate de analiză bazate pe alinierea jurnalului cu modelul. Specifice pentru procesele Lasagna sunt următoarele acțiuni: detectarea și eliminarea găturilor, analiza cauzelor de bază ale devierilor, *Data Mining* comparat, suportul operațional ș.a.

3.9.2. Procese Spaghetti

În cazul proceselor Spaghetti nu se pot aplica toate tehnicile de *Data Mining*, dar anumite tehnici pot asigura îmbunătățiri majore. Ultimul stadiu din ciclul de viață, L^* (secțiunea 5.2.3) nu se poate executa pentru procesele Spaghetti, deoarece nu există un model de flux de control stabil. Primele stagii pot fi foarte provocatoare, dar și potențialul de îmbunătățire este mult mai mare. Procesele Spaghetti trebuie simplificate, operatorul concentrându-se pe un subset de activități (cele mai frecvente activități sau pe anumite regiuni din proces), pe un subset de cazuri (clustering, grupuri omogene de cazuri sau subcazuri naturale), pe un subset de căi (cele mai frecvente).

3.9.3. Clasificarea funcțiilor organizației după tipul de procese

Față de cele prezentate anterior, analiza se poate concentra și pe funcțiile unei organizații (producție, financiar, logistică, dezvoltare etc.). După cum este prezentat în Fig. 3.24, zonele cu albastru arată funcțiile unei organizații unde predomină procese Lasagna (achiziții, logistică, producție, financiar-contabil); în zonele cu mov predomină procesele Spaghetti (managementul resurselor, vânzări/relații cu clienții), iar în zonele cu violet sunt întâlnite atât procese Spaghetti, cât și Lasagna (Servicii, Dezvoltare de produs).

Specifice pentru procesele Spaghetti sunt următoarele acțiuni: descoperirea procesului, evidențierea zonelor cu probleme, îmbunătățirea evidentă a proceselor.

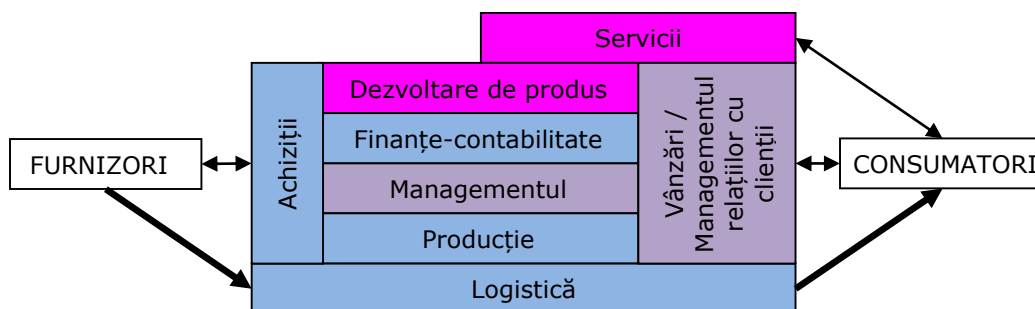


Fig. 3.24. Clasificarea funcțiilor organizației după tipul de procese

3.10. Concluzii

În cadrul capitolului au fost prezentate rezultate ale cercetărilor teoretice cu privire la modul de derulare al proiectelor de *Process Mining* în cadrul organizațiilor. După descrierea generală a modului de derulare a proiectelor de *Process Mining* (cinci etape, Fig. 3.2) s-a realizat panoramarea și explicarea modului de realizare al pregătirii datelor, descoperirii procesului, verificării conformității și al suportului operațional (detecția, predicția și recomandarea).

În finalul capitolului au fost prezentate alte perspective ale demersului de *Process Mining* (Mining-ul punctelor de decizie, operațional, al rețelelor sociale, organizațional, precum și modul de combinare al diferitelor perspective și avantajele obținute), cu implicații în susținerea procesului decizional la nivel organizațional.

Demersul de cercetare, adoptat ulterior și în cadrul cercetărilor aplicative, a fost descris prin prezentarea modului de realizare al conducerii unui proiect de *Process Mining* (caracterizând problematica sarcinilor de realizat în fiecare etapă), precum și al cadrului de lucru pentru *Process Mining* rafinat (Fig. 3.24).

Clasificarea modelelor obținute s-a realizat pe baza descrierii proceselor Lasagna și Spaghetti, dar și cu unele completări referitoare la clasificarea funcțiilor organizației după tipul de procese asociate acestora.

Pentru ca un analist de proces să înțeleagă tehnicile *Process Mining* trebuie să cunoască modele de reprezentare ale proceselor, precum și tehnici specifice *Data Mining*, care au fost definite în capitolul 2. Algoritmii de descoperire a proceselor au ca suport diferite reprezentări de modele de proces: rețele Petri, rețele Flux de Lucru, rețele cauzale tip C-nets, rețele euristice, arbori de proces, modele fuzzy ș.a. Există numeroși algoritmi de *descoperire*, în prezenta lucrare fiind însă documentați doar cei mai reprezentativi (algoritmul α , algoritmul HeuristicsMiner, algoritmul Fuzzy Miner, algoritmul Inductive Miner ș.a.). Tehnicile de *verificare a conformității* oferă metode de comparare a modelelor de proces descoperite pentru a-l alege pe cel mai adecvat, dar și metode de investigare și reparare a modelelor de proces.

Pentru o mai bună înțelegere a tehnicilor, au fost oferite exemple practice. În viziunea autorului, cercetările teoretice descrise în prezentul capitol sunt bagajul minim de cunoștințe pe care un analist de proces trebuie să îl dețină pentru a porni pe drumul implementării *Process Mining* în organizații.

În vederea definirii mai clare a metodologiei de realizare și de conducere a proiectelor de *Process Mining* în organizații, este nevoie de o cercetare asupra mijloacelor software (disponibile) care pot susține etapele și acțiunile specifice și care pot contribui la creșterea performanțelor proceselor de afaceri. Acest lucru va fi realizat în capitolul următor al tezei de doctorat.

Proiecte de *Process Mining* au fost implementate în diferite domenii (medical, producție, public, bancar), existând deja câteva metodologii pentru implementarea unor astfel de proiecte. Este nevoie de o cercetare teoretică a acestor metodologii, cercetare ce va fi realizată în capitolul 5. După cum s-a arătat în capitolul 1, deși relativ nou în aria academică, domeniul *Process Mining* abia acum este adoptat și acceptat de către industrie (Fig. 1.9). **Implementările de tip *Process Mining* sunt încă la început și nu sunt realizate în contextul global BPM din cadrul unei organizații.** Training-ul *CertiBPM* prezintă complet modul de implementare a unui proiect BPM, iar experiența acumulată în *CertiBPM* și tehnicile de *Process Mining* vor pune bazele unei metodologii de implementare în viziune proprie a proiectelor BPM.

4. MIJLOACE SOFTWARE INTELIGENTE CARE SUSȚIN CREȘTEREA PERFORMANȚELOR PROCESELOR DE AFACERI

În cadrul capitolului se vor prezenta mijloace software utilizate în aplicarea demersurilor de *Process Mining* și *Data Mining*. Prin analiza critică a acestora vor fi stabilite mijloacele asociate metodologiei propuse pentru derularea proiectelor de *Process Mining* (ceea ce este echivalent cu definirea completă a acestui demers). Astfel, se vor crea premisele pentru cercetările aplicative.

Obiectivul operațional aferent cercetărilor din acest capitol a fost: **OP4** Cercetări teoretico-aplicative pentru identificarea și caracterizarea mijloacelor software asociate metodologiei pentru derularea proiectelor de *Process Mining* în organizații.

4.1. Poziționarea mijloacelor software inteligente pentru analiza proceselor

Mijloacele software inteligente de analiză a proceselor se pot grupa în două categorii: cele care sunt destinate strict demersului de *Process Mining* și cele ce se constituie ca suport pentru *Data Mining*. În continuare, vor fi prezentate câteva mijloace software inteligente pentru analiza proceselor, insistându-se asupra acelor aspecte care aduc valoare analizei.

Există un spectru larg de aplicații software, care se pot clasifica în două dimensiuni: ușurința de utilizare și specificitate, la un capăt poziționând *Process Mining*, iar la celălalt *Data Mining*. În Fig. 4.1 este prezentată poziționarea anumitor aplicații software ce sporesc informațiile pentru proces.

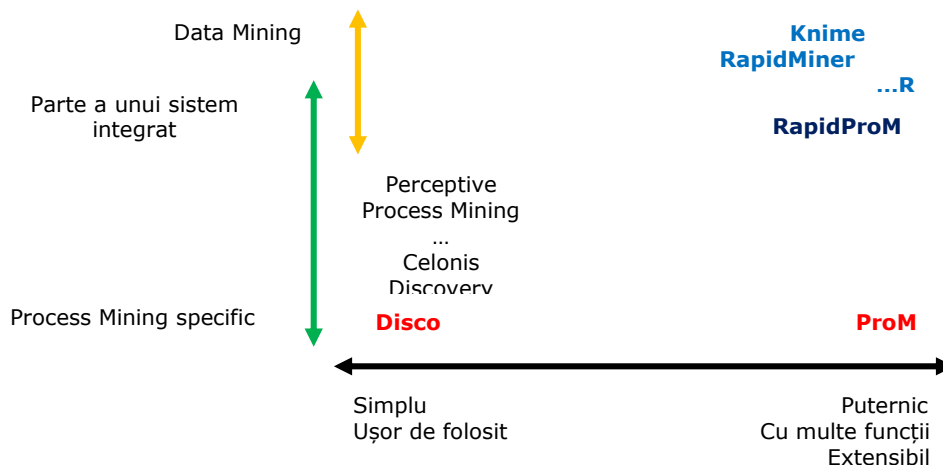


Fig. 4.1. Spectru de mijloace software pentru *Process Mining* și *Data Mining*

4.2. Mijloace software pentru Process Mining

Ca și mijloc util managementului organizației, un sistem de BPM ar trebui să conțină setul întreg de soluții pentru corelarea cu sistemele externe printr-o varietate de protocoale [72] definite în cadrul organizației.

Mijloacele software pentru *Process Mining* se adresează atât operatorilor de business, cât și cercetătorilor. Mijloacele pentru *Process Mining* au o valoare însemnată pentru lumea afacerilor. În Fig. 4.2 este prezentată poziționarea instrumentelor software pentru *Process Mining* în dimensiunea utilitate, respectiv în dimensiunea utilizatorilor. Putem concluziona, că mijloacele software pentru *Process Mining* aduc valoare mare (comparativ cu alte mijloace software utilizate în afaceri, cum este *Excel*), iar pentru a le utiliza nu sunt necesare cunoștințe foarte avansate (comparativ cu alte mijloace ce necesită abilități de programare, cum este *Hadoop*).

Process Mining a apărut ca și concept în cadrul Universității din Eindhoven, Olanda, sub îndrumarea profesorului Wil van der Aalst. Tot aici s-a dezvoltat software-ul open source *ProM* (<http://www.promtools.org>), care este o platformă ce permite instalarea multor extensii (plug-ins) utile pentru *Process Mining* [55]. Acest software a apărut în mediul academic, este foarte complex și este util cercetătorilor din domeniul *Process Mining*. Doi doctoranzi ai profesorului van der Aalst, Anne Rozinat și Cristian Günther au fost implicați în numeroase proiecte de *Process Mining* în cadrul unor companii multinaționale, și pornind de la nevoile identificate au creat un software comercial denumit *Disco* (<https://fluxicon.com/disco/>). *Disco* este o versiune adaptată a *ProM* pentru a răspunde nevoilor unei companii pentru *Process Mining*, această aplicație software fiind mult mai ușor de utilizat decât *ProM*.

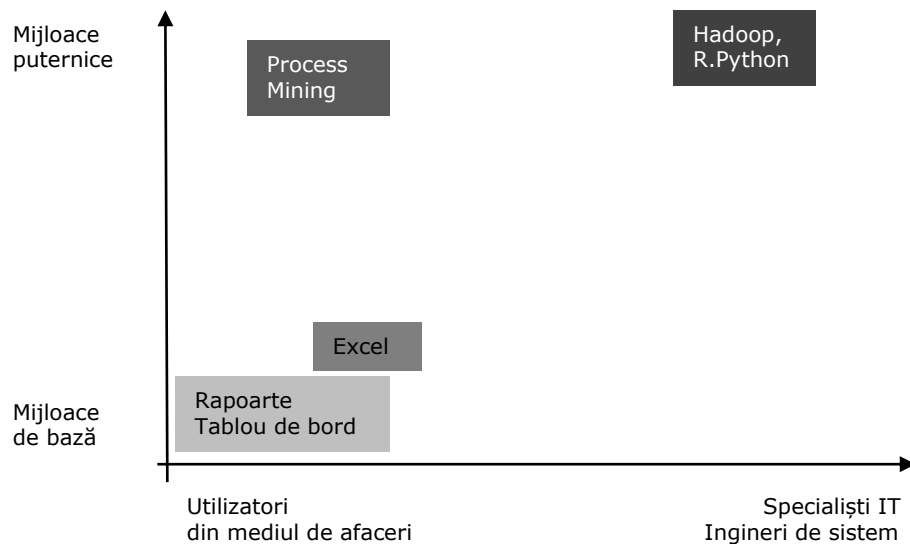


Fig. 4.2. Poziționarea mijloacelor software pentru *Process Mining* în funcție de tipul de utilizator și beneficii

4.2.1. Platforma ProM

ProM tinde să acopere tot spectrul de *Process Mining*: cartografiere, navigare și audit. *ProM* este un cadru extensibil care suportă o mare varietate de tehnici de *Process Mining* sub forma de plug-in-uri [160]. Fiind implementat în Java, *ProM* este independent de platforma pe care rulează, deci are o portabilitate mare. *ProM* este un proiect open-source, deci poate fi folosit gratuit. *ProM* s-a pornit ca un proiect open-source pentru a ajunge cât mai ușor la cercetători și dezvoltatori, care sunt încurajați să contribuie cu aplicații practice, sub forma unor noi plug-in-uri. În prezent există un număr mare de plug-in-uri, aproximativ 600. Multe din plug-in-uri sunt experimentale, și din acest punct de vedere nu sunt foarte intuitive, nici ușor de folosit. Un avantaj mare al platformei *ProM* este facilitarea interacțiunii dintre plug-in-uri. Există plug-in-uri de import, de export, de conversie, de analiză și de Mining.

ProM suportă multe notații: rețele Petri, BPMN, rețele C, modele fuzzy, sisteme de tranziții ș.a. *ProM* permite suportul operațional și poate verifica conformitatea modelelor descoperite.

Versiunea actuală de *ProM* (August 2015) este 6.5. Fișierele principale de intrare pentru aplicația *ProM* sunt jurnalele de evenimente. Formatul principal utilizat pentru jurnalele de evenimente pentru *ProM 6* este XES [159]. Denumirea formatului XES provine de la *eXtensible Event Stream* (Flux Extensibil de Evenimente). Versiunea anterioară (*ProM 5.2*) folosea ca import formatul MXML. Formatul XES prezintă câteva avantaje importante față de formatul anterior MXML și anume:

- Este folosită altă structură: în XES elementele de jurnal, trasee și evenimente definesc numai structura documentului și nu conțin nicio informație. Informația este regăsită ulterior în atribute. Cu ajutorul atributelor se pot stoca mai multe informații.
- În XES clasificatoarele de eveniment pot fi specificate în elementul de jurnal care atribuie o identitate fiecărui eveniment. Acest lucru facilitează compararea evenimentelor între ele prin identitatea lor atribuită.
- În XES faptul că anumite atribute au valori bine definite pentru fiecare traseu și/sau eveniment din jurnal poate fi specificat de către un element global. În cazul în care un traseu sau un eveniment nu au atribute definite, atunci ele vor fi luate din valorile globale.

Există trei aplicații care se instalează în pachetul *ProM*: *ProM 6*, *ProM 6 Package Manager* și *XESame*.

4.2.1.1. ProM 6

Fig. 4.3 prezintă interfața cu utilizatorul pentru *ProM 6*. În continuare se vor explica caracteristicile de bază ale acestei interfețe cu utilizatorul prin prezentarea diferitelor obiecte în această interfață. Pentru aceasta, elementelor din această interfață le sunt atribuite câte o cifră (de la 1 la 3), urmând ca apoi să se explice fiecare obiect în parte. Cifrele 1, 2 și 3 arată principalele puncte de vedere astfel: 1 – Spațiul de lucru, 2 – Secțiunea Acțiuni și 3 – Secțiunea Observare.

1. Spațiul de lucru

Spațiul de lucru arată toate resursele (cum ar fi jurnalele de evenimente și rețele Petri), care fie au fost importate în *ProM 6* sau sunt un rezultat al executării unei acțiuni asupra altor resurse. Din această listă de resurse, cel mult una poate fi selectată, pentru care sunt afișate detalii suplimentare. În Fig. 4.4 este prezentată o captură a spațiului de lucru.

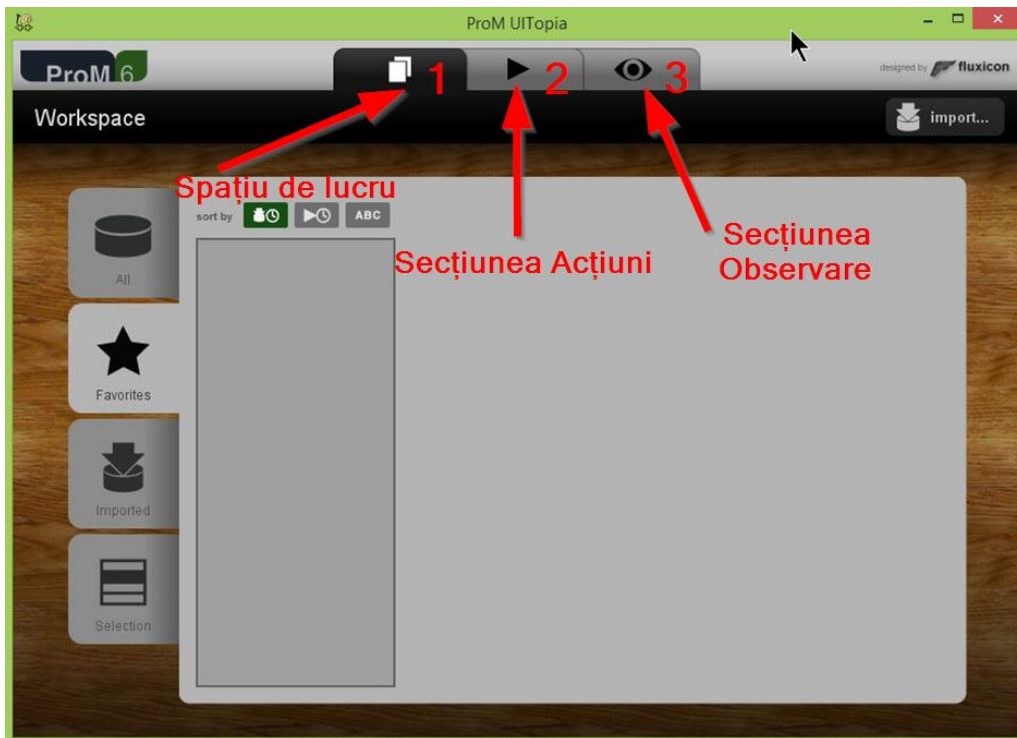


Fig. 4.3. ProM – interfața inițială de utilizare (captură ecran)

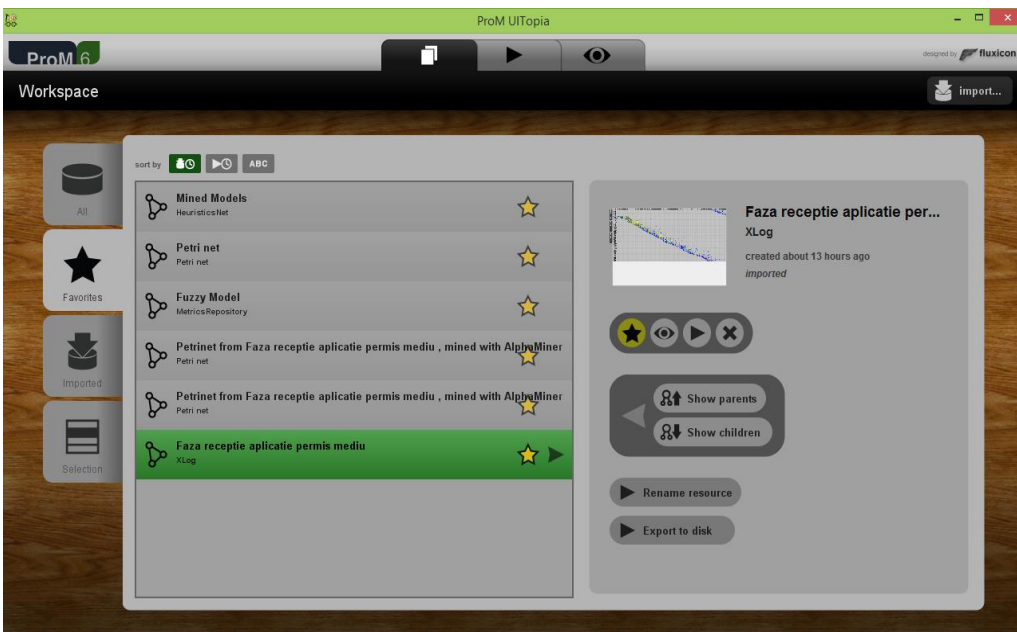


Fig. 4.4. ProM – Spațiul de lucru (captură ecran)

2. Secțiunea Acțiuni

În această secțiune sunt prezentate toate acțiunile care pot lua anumite resurse ca intrări, care pot duce la anumite tipuri ca ieșiri, și care corespund anumitor criterii de filtrare. Utilizând secțiunea Acțiuni, este simplă selecția acțiunilor adecvate pentru resurse. Interfața ajută utilizatorul să vizualizeze și să selecteze acțiunea, dorită prin filtrarea acțiunilor care au ca intrare resursa curentă. Dacă această resursă nu este selectată dinainte, utilizatorul va trebui să aleagă acțiunea, iar acțiunea aleasă sugerează utilizatorului atât tipul resursei de intrare, cât și tipul resursei de ieșire. În Fig 4.5 este prezentată o captură a secțiunii Acțiuni.

3. Secțiunea Observare

Această secțiune facilitează observarea resurselor (Fig. 4.6). Practic, se poate vedea ce conține o resursă sau o imagine de ansamblu a tuturor resurselor pentru care există o vedere.

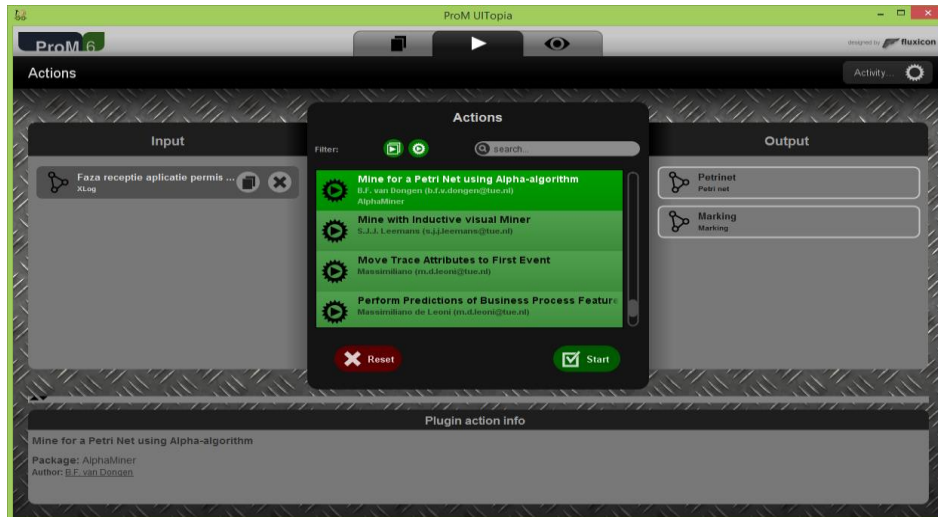


Fig. 4.5. ProM – Secțiunea Acțiuni (captură ecran)



Fig. 4.6. ProM – Secțiunea Observare (captură ecran)

4.2.1.2. ProM 6 Package Manager

Această interfață este folosită pentru configurarea *ProM 6*. Această interfață suportă: instalarea și deinstalarea plug-in-urilor, configurarea alocării spațiului de memorie pentru *ProM 6*. Toate modificările de setări implică restartarea aplicației *ProM 6*. Această interfață este prezentată în Fig. 4.7.

4.2.1.3. XESame

Prima etapă din *Process Mining* o reprezintă pregătirea jurnalelor de evenimente. De multe ori, aceste jurnale de evenimente trebuie create din numeroase surse. De asemenea, jurnalele de evenimente pot fi îmbogățite cu informații despre alte perspective gen: resurse, timpi de începere și de încheiere a activităților, costul activităților ș.a. Aplicația XESame ne ajută în crearea acestor jurnale de evenimente, mai exact în extragerea acestora. Ca orice proiect de Business Intelligence, prima etapă o reprezintă extragerea datelor. Aplicația XESame a fost dezvoltată de J.C.A.M. (Joos) Buijs în cadrul lucrării sale de disertație [42]. Aplicația poate crea jurnale de evenimente în formatul XES și MXML.

Acest instrument se poate conecta la orice bază de date ODBC, și permite expertului în domeniu să adauge detalii cu privire la extragerea dorită într-un mod simplu. Scopul acestei aplicații este obținerea unui jurnal de evenimente XES, iar de la acest jurnal se va începe *Process Mining*. Versiunea actuală de XESame este 1.5 (August 2015). Aplicația XESame are trei secțiuni și anume: Configurare, Mapare și Acțiune (Fig 4.8).

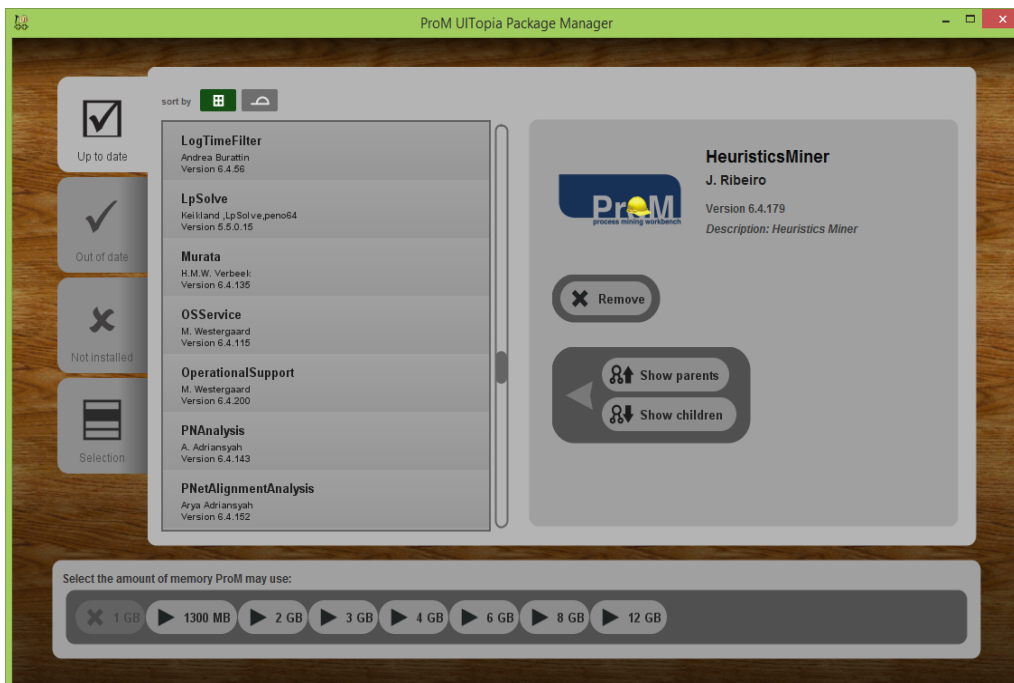


Fig. 4.7. ProM 6 Package Manager (captură ecran)

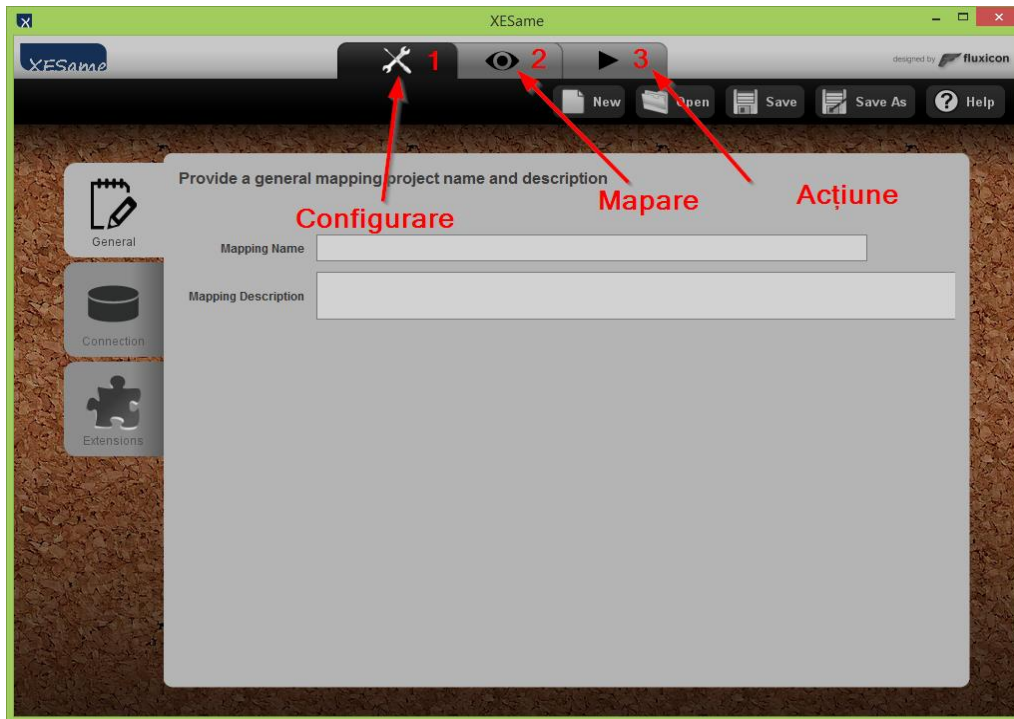


Fig. 4.8. XESame – interfața inițială de utilizare (captură ecran)

1. Secțiunea Configurare

În secțiunea Configurare sunt definite numele mapării și conexiunea la baza de date, și se stabilesc ce extensii vor fi folosite în cadrul conversiei. De asemenea, se pot crea, modifica și salva mapări.

2. Secțiunea Mapare

Această secțiune permite definirea mapării, și anume stabilirea atributelor XES (la nivel de jurnal, traseu, eveniment, sau atribut) ce urmează a fi incluse în jurnalul XES și cum sunt captate valorile corespunzătoare acestor atribute din baza de date sursă (Fig. 4.9). De asemenea, această secțiune permite vizualizarea într-un mod grafic a legăturilor dintre atributele XES și baza de date sursă. Definirea atributelor este importantă dacă se dorește realizarea unui demers de Mining în mai multe dimensiuni.

3. Secțiunea Acțiune

Această secțiune permite controlul și monitorizarea execuției mapării (Fig. 4.10). Are ca rezultat producerea jurnalului de evenimente în format XES din baza de date sursă folosind configurația de cartografiere și definirea specificată în cele două secțiuni anterioare.

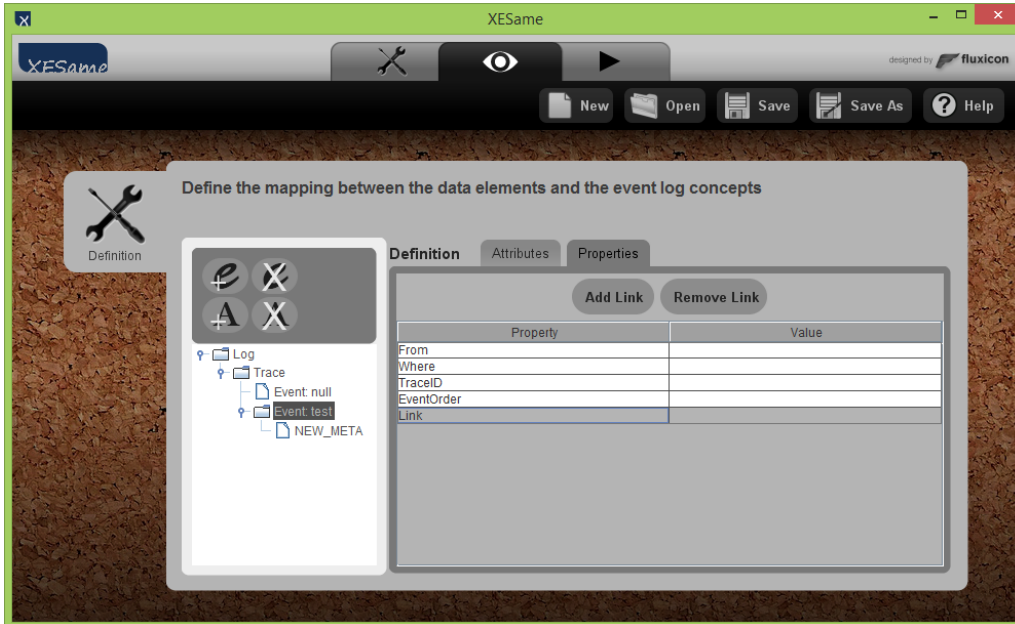


Fig. 4.9. XESame – Secțiunea Mapare (captură ecran)

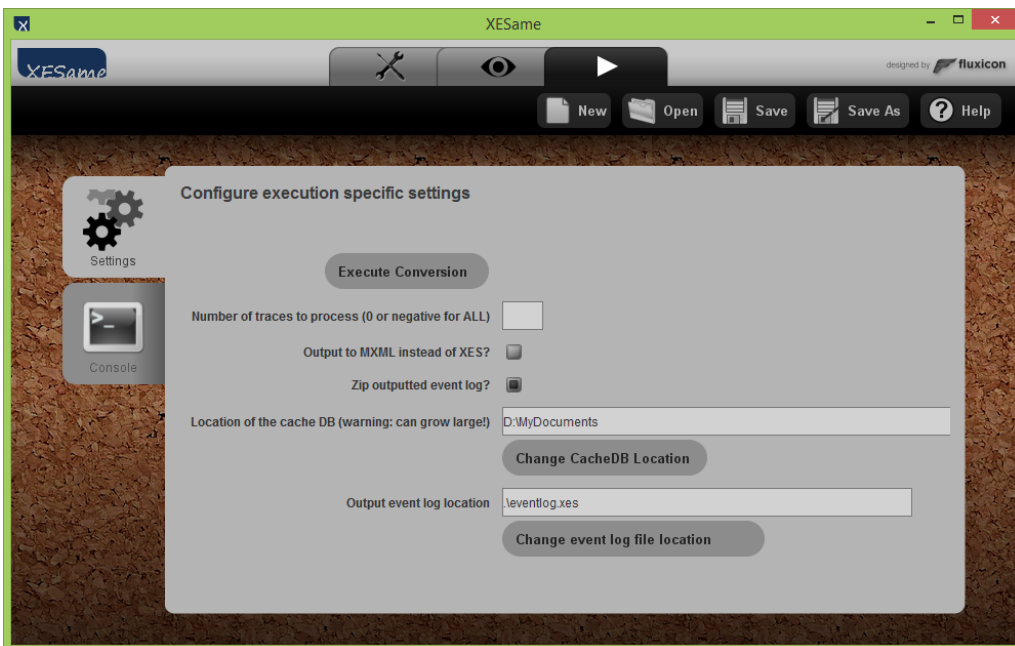


Fig. 4.10. XESame – Secțiunea Acțiune (captură ecran)

4.2.2. Disco

După cum am menționat mai sus, o versiune simplificată a *ProM* este *Disco*. *Disco* este un software comercial, dar poate fi dobândit cu licență academică datorită faptului că autorii au fost la rândul lor implicați în cercetare. *Disco* este axat pe descoperire și analiza performanței de proces. Ca și implementare pentru descoperirea proceselor, el folosește o variantă de modele fuzzy. *Disco* nu are suport pentru verificarea conformității și nu permite suportul operațional. Dacă ne referim la poziționarea *Disco* în cadrul *Process Mining* rafinat, putem concluda că *Disco* are suport doar pentru cartografiere, iar nu pentru navigare și audit. *Disco* este foarte ușor de utilizat și are performanțe foarte bune. *Disco* dispune și de capabilități de filtrare, astfel se poate executa *Process Mining* doar pe anumite părți din jurnalul de evenimente.

Ca și funcționalități, acesta are trei secțiuni: Hartă, Statistici și Cazuri. La secțiunea Hartă se poate vedea procesul descoperit. Există controale de tip „slider” prin care se poate ajusta gradul de acoperire a căilor și a activităților în funcție de frecvența de apariție din jurnalul de evenimente. Procesul modelat poate fi urmărit din două perspective: din perspectiva frecvenței cazurilor și din perspectiva performanței. La secțiunea Statistici, sunt prezente statistici referitoare la evenimentele din jurnal precum: câte evenimente sunt prezente per caz, care ar fi variantele de cazuri cele mai folosite, activitatea cazurilor în timp, durata cazurilor ș.a. În secțiunea Cazuri, există informații referitoare la fiecare caz în parte. Cazurile sunt sortate și se pot obține informații despre activitățile ce sunt conținute și despre durata de viață a cazului. În Fig. 4.11 sunt prezentate capturi de ecran aferente programului *Disco*.

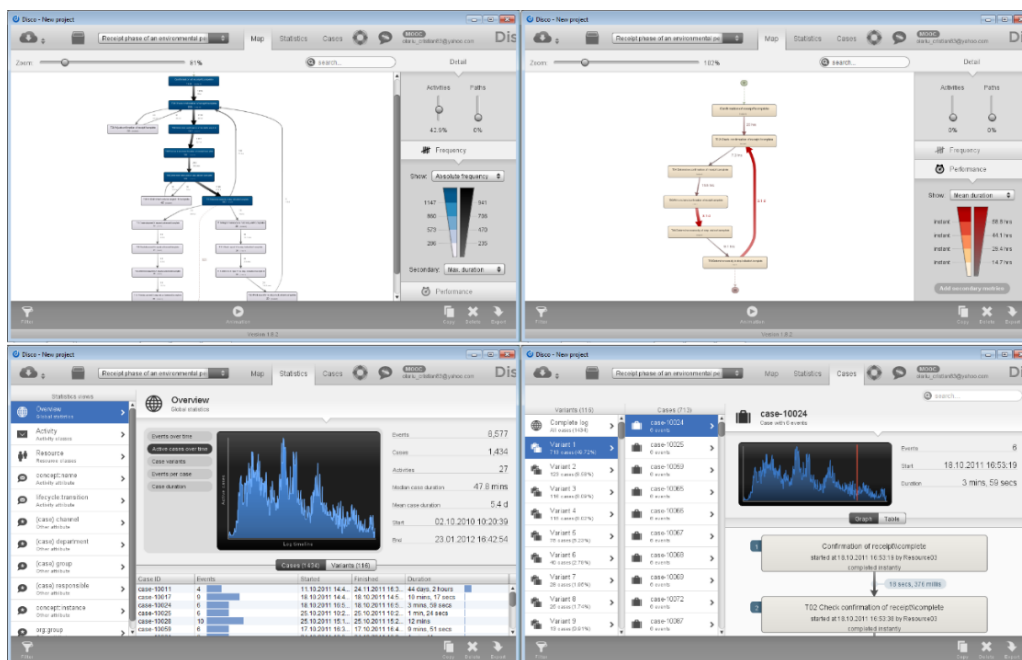


Fig. 4.11. Disco – capturi ecran ale diferitelor interfețe

4.2.3. Alte aplicații software pentru Process Mining

În momentul de față, sunt foarte puține BPMS care suportă descoperirea de proces și verificarea conformității. În același timp, nu există nici software-uri de *Data Mining* sau de învățare automată care să suporte *Process Mining*. Același lucru se poate spune și despre software-urile BI, niciunul din aceste mijloace nu suportă *Process Mining*. Câteva software-uri BI sunt: *IBM Cognos Business Intelligence (IBM)*, *Oracle Business Intelligence (Oracle)*, *SAP BusinessObjects (SAP)*, *WebFOCUS (Information Builders)*, *MS SQL Server (Microsoft)*, *MicroStrategy (MicroStrategy)*, *NovaView (Panorama Software)*, *QlikView (QlikTech)*, *SAS Enterprise Business Intelligence (SAS)*, *TIBCO Spotfire Analytics (TIBCO)*, *Jaspersoft (Jaspersoft)* și *Pentaho BI Suite (Pentaho)*. Produsele BI sunt capabile de a realiza ETL (*Extract, Transform, Load*), interogări ad-hoc, rapoarte, tablouri de bord interactive, generare de alerte. Toate aceste produse nu sunt orientate pe proces, iar tipul lor de analiză implementat este adesea considerat simplist.

Process Mining încă este o disciplină tânără, fiind implementată doar în câteva produse software. Există BPMS-uri care suportă *Process Mining*. Aceste software-uri au funcționalități numeroase pentru implementarea de procese de afaceri, sunt produse cu o lungă tradiție BPM.

În urma cercetărilor recente în domeniul *Process Mining*, multe dintre BPMS-urile cu tradiție încep să ofere această funcționalitate. *Process Mining* sporește informațiile pentru procese, de aceea această funcționalitate este una importantă care a dus la transformarea BPMS, în iBMPS (*Intelligent Business Process Management Software/Suites*) sau SIMPA (Software Inteligent pentru Managementul Proceselor de Afaceri) [130].

Printre programele comerciale sau mijloacele software care suportă *Process Mining*, amintim: *Perceptive Process Mining* (anterior *Futura Reflect și BPM|one*, produs de Perceptive Software), *ARIS Process Performance Manager* (Software AG), *QPR Process Analyzer (QPR)*, *Celonis Discovery (Celonis)*, *Interstage Process Discovery (Fujitsu)*, *Discovery Analyst (StereoLOGIC)*, *XMAnalyzer (XMPro)* ș.a. Majoritatea acestor produse software sunt similare cu *Disco*, de multe ori mai greoaie în utilizare, având o interfață cu utilizatorul mai puțin prietenoasă. În cele mai frecvente cazuri, aceste produse software acoperă următoarele secvențe ale demersului de *Process Mining*: descoperirea de proces, evaluarea performanței în funcție de timp și anumite tehnici de îmbunătățire.

4.3. Programe software pentru Data Mining

Mijloacele software pentru *Data Mining* fac parte din sfera Business Intelligence. Există atât soluții comerciale, cât și open-source. O analiză cuprinzătoare a acestor soluții software specifice pentru *Data Mining* este prezentată în [121]. Acest studiu grupează software-urile după: grupul de utilizatori, structurile de date, atribuții și metode, interacțiune și vizualizare, importul și exportul de date și modele, platforme și licențe.

Printre cele mai utilizate software-uri comerciale remarcăm: *ADAPA (Zementis)*, *CART (Salford Systems)*, *IBM SPSS Modeler*, *IBM SPSS Statistics*, *KXEN (SAP)*, *Matlab (MathWorks)*, *Oracle Data Mining*, *SAP Netweaver Business Warehouse*, *SAS Enterprise Miner*, *SQL Server Analysis Services (Microsoft)*, *Statistica (StatSoft)*, *Teradata Database (Teradata)*, *TIBCO Spotfire (Tibco)*.

Există multe soluții software open-source care reprezintă o bună alternativă față de cele comerciale. Cele mai răspândite soluții software de tip open-source (pentru *Data Mining*) sunt: *ImageJ* (<http://rsbweb.nih.gov/ij/>), *ITK* (www.itk.org), *KNIME* (<http://www.knime.org/>), *Orange* (<http://orange.biolab.si/>), *Pentaho* (<http://sourceforge.net/projects/pentaho/>), *R* (www.r-project.org), *RapidMiner* (<https://rapidminer.com/>) și *WEKA* (<http://sourceforge.net/projects/weka/>).

Una dintre cele mai de succes soluții software open-source este *RapidMiner*, care va fi descrisă în cele ce urmează. De asemenea, există și o combinație între *ProM* și *RapidMiner*, astfel luând naștere soluția software *RapidProM* (<http://www.rapidprom.org/>), care este de fapt este un plug-in pentru *RapidMiner*.

4.3.1. RapidMiner

RapidMiner a fost dezvoltat de cercetători din domeniul Inteligenței artificiale din cadrul Universității Tehnice din Dortmund. Inițial s-a numit *YALE (Yet Another Learning Environment)*, iar proiectul a început în 2001 [120]. În 2007 numele i s-a schimbat în *RapidMiner*, iar dezvoltarea produsului a fost continuată de către autorii inițiali care au pornit o afacere cu numele *RapidMiner*. În 2014, acest software s-a clasificat ca lider pe piață, ca platformă de analiză avansată *AAP (Advanced Analytics Platforms)*, fiind plasat în cadranul „magic” al companiei de consultanță Gartner [87]. Acest produs este o platformă ce are o interfață care nu necesită scrierea de cod pentru a obține diverse analize de date. Produsul este disponibil atât în cloud, cât și ca platformă client/server. Există și o versiune cu plată ce permite conectarea mai multor surse de date și prelucrează date mult mai mari. Platforma *RapidMiner* este extensibilă prin diferite limbaje, biblioteci analitice, platforme BI și aplicații web. *RapidMiner* este soluția cea mai frecvent aleasă de către clienți, datorită ușurinței în utilizare, implementare, precum și datorită costurilor mici de licență.

RapidMiner oferă suport pentru *Data Mining* și pentru învățare automată: extragerea, transformarea și încărcarea datelor (*Extract, Transform, Load, ETL*); preprocesarea și vizualizarea datelor; analiza predictivă și modelarea statistică, evaluare și implementare. *RapidMiner* este scris în limbajul de programare Java. *RapidMiner* oferă o interfață grafică pentru a proiecta și executa fluxuri de lucru analitice. Aceste fluxuri de lucru sunt numite „proces” în *RapidMiner* și conțin o multitudine de „operatori”. Fiecare operator efectuează o singură sarcină în cadrul procesului și ieșirea fiecărui operator face intrarea celui următor. Alternativ, motorul poate fi apelat de către alte programe sau folosit ca o Interfață pentru Programarea de Aplicații, IAP (Application Programming Interface - API). Funcțiile individuale pot fi apelate din linia de comandă.

Pe platforma *RapidMiner* se pot rula algoritmi: învățarea bazată pe arbori de decizie, clustering-ul *k-means* și învățarea bazată pe reguli asociative, algoritmi care au fost explicați în capitolul 2.2. O captură a *RapidMiner* este ilustrată în Fig. 4.12.

4.3.2. Extensia RapidProM

Extensia *RapidProM* este principalul motiv pentru care s-a ales prezentarea software-ului *RapidMiner*. *RapidProM* se poate instala ca plug-in peste platforma *RapidMiner*. După instalare, se pot observa în lista cu operatori din *RapidMiner*, operatori specifici pentru *ProM* grupați pe categorii: context, import, Mining, analiză, export, filtrare, conversie și paralelizare (Fig. 4.13).

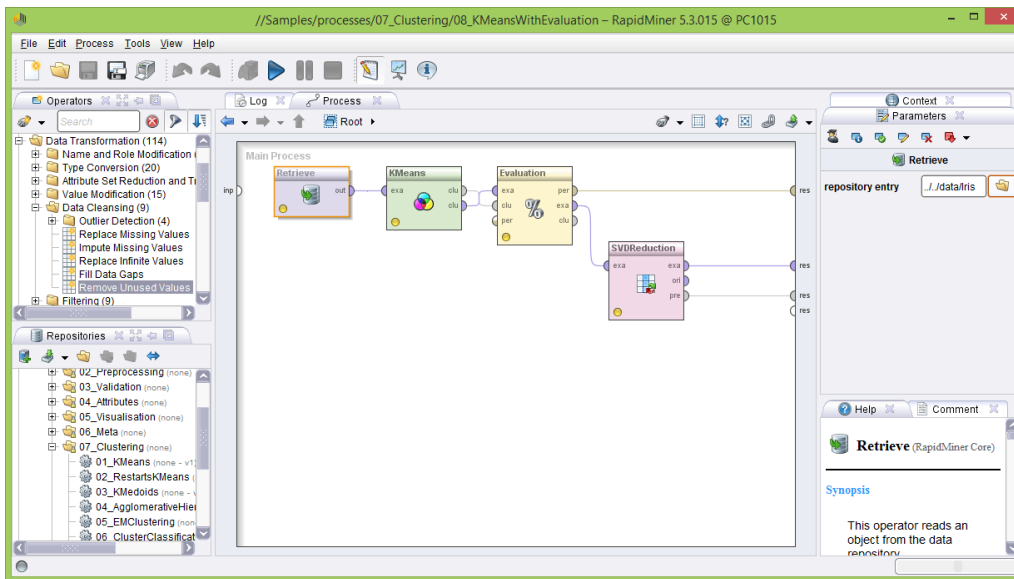


Fig. 4.12. RapidMiner (captură ecran)

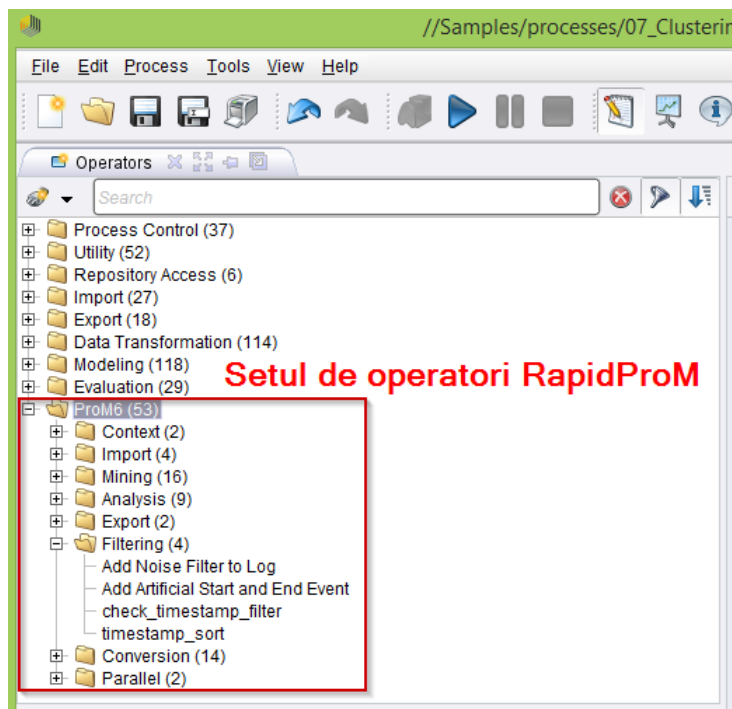


Fig. 4.13. RapidProm – extensia de legătură dintre platformele ProM și RapidMiner (captură ecran)

Cu ajutorul *RapidProM* se pot executa fluxuri de lucru pentru *Process Mining* [113]. Avantajele folosirii unui flux de lucru pentru *Process Mining* sunt:

- Analizele prin tehnicile de *Process Mining* pot fi ușor repetate. Avem de a face cu o automatizare a fluxului de lucru, astfel că un cercetător care execută anumite activități repetitive în *ProM*, le poate executa mult mai repede în *RapidProM*;
- Un experiment executat anterior poate fi ușor repetat;
- În cazul în care se studiază un algoritm la care se variază parametrii de intrare, din punct de vedere experimental, *RapidProM* reduce mult timpul de execuție;
- Anumite prelucrări de date, analize, vizualizări și explorări de date pot fi corelate cu *Process Mining*, obținându-se rezultate mai interesante decât prin simpla rulare a *ProM*.

4.4. Concluzii

În cadrul acestui capitol au fost prezentate două categorii de mijloace software: pentru *Process Mining* (platforma *ProM*, *Disco* și alte soluții existente pe piață) și pentru *Data Mining* (*RapidMiner* și extensia *RapidProM*). De fiecare dată, s-a insistat pe caracterizarea acelor considerente care stau la baza alegerii soluțiilor software, mai ales pe caracteristicile privind funcționalitățile deținute (atât în cazul *Process Mining*, cât și în cazul *Data Mining*) și pe aspecte legate de ușurința în exploatare, prin descrierea interfeței cu utilizatorul (facilități de utilizare/exploatare). Mijloacele software prezentate au fost alese deoarece sunt open-source și sunt create chiar de către pionierii *Process Mining*. Datorită faptului că sunt open-source (excepție *Disco*), aceste mijloace pot fi luate în considerare mai ales de către firmele mici și mijlocii, care nu își permit un buget IT ridicat. Alt motiv pentru care au fost considerate aceste soluții software sunt plasarea lor atât în domeniul academic, cât și în domeniul practic. Astfel, folosind *ProM*, se poate avea acces la ultimele noutăți din domeniul academic. Platforma *ProM* este folosită în toate metodologiile de *Process Mining* care vor fi exemplificate în capitolul următor, fiind un mijloc software validat în proiecte practice. Platforma *ProM* are suport pentru toate etapele unui proiect de *Process Mining*, prezentat în Fig. 3.2. Etapa de extragere a datelor este suportată de aplicația *XESame*, iar algoritmii de *descoperire* de proces și de *verificare a conformității* sunt integrați în platforma *ProM* sub formă de plug-in-uri. Pentru a familiariza un viitor analist de proces cu platforma *ProM*, modul de utilizare al acesteia a fost exemplificat. Programul *Disco* a fost ales întrucât el este derivat din *ProM* și este gândit special pentru a obține o investigație rapidă a procesului. În plus, un alt criteriu pentru alegerea acestui program îl reprezintă faptul că este ușor de folosit și intuitiv.

Pe măsura ce vor apărea mijloace software superioare celor exemplificate în acest capitol, se poate apela la o metodă de alegere prin comparare a celui mai indicat mijloc software. Un exemplu este Metoda ELECTRE [26], care reprezintă o metodologie de luare a deciziei în condiții de certitudine.

Ca urmare a analizei soluțiilor software realizate și corelate cu metodologia demersului de *Process Mining* (capitolul 5), există creat cadrul metodologic pentru realizarea cercetărilor aplicative. Acestea vor evidenția mai bine utilitatea demersului de *Process Mining* în BPM și în sprijinul deciziei manageriale din organizații.

5. METODOLOGIA DE IMPLEMENTARE A UNUI PROIECT BPM CU AJUTORUL PROCESS MINING

În cadrul capitolului se vor prezenta metodologiile de implementare a unui proiect BPM. Se va porni printr-o abordare clasică prezentată în materialele *CertiBPM*. Se vor prezenta și metodologiile de implementare a proiectelor de *Process Mining*, care sunt în număr foarte redus. Proiectele *Process Mining* se inspiră în principal din proiecte de *Data Mining*, dar ele nu sunt abordate holistic în cadrul unei implementări BPM.

Obiectivul operațional aferent cercetărilor din acest capitol a fost: **OP5** Cercetări teoretice pentru conceperea metodologiei *CertiBPM-PM*, propuse pentru derularea proiectelor de *Process Mining* în organizații.

5.1. Implementarea unui proiect BPM în viziunea *CertiBPM*

Managementul unui proiect BPM în viziunea cunoașterii dobândite prin proiectul *CertiBPM* a fost descris în elementul de studiu *BPM.U1.E2* intitulat *Managementul proiectelor BPM* (Tabelul 1.2) aparținând materialelor necesare certificării *CertiBPM*. Managementul unui proiect BPM este prezentat schematic în Fig. 5.1.

Etapele reprezentative ale abordării *CertiBPM* sunt explicate în continuare în detaliu:

1. **Justificarea și definirea obiectivelor proiectului BPM**

În această etapă se definesc obiectivele. Obiectivele trebuie să fie măsurabile, deoarece nu se poate îmbunătăți ceva ce nu se poate măsura. Obiectivele ar putea fi:

- documentarea pentru obținerea unei certificări ISO 9001:2008;
- automatizarea unor procese;
- specificarea cerințelor pentru un nou sistem informatic;
- îmbunătățirea timpului de răspuns către clienți;
- reducerea costurilor;
- procese transparente, ș.a.

2. **Obținerea sprijinului tuturor părților interesate**

Punctul de start este convingerea managementului sau, dacă de la management pornește inițiativa, atunci punctul de start este convingerea managementului superior ierarhic și a angajaților. Dacă nu se obține sprijinul tuturor părților interesate, atunci angajații vor crede că managementul implementează un nou sistem de tip „big-brother”.

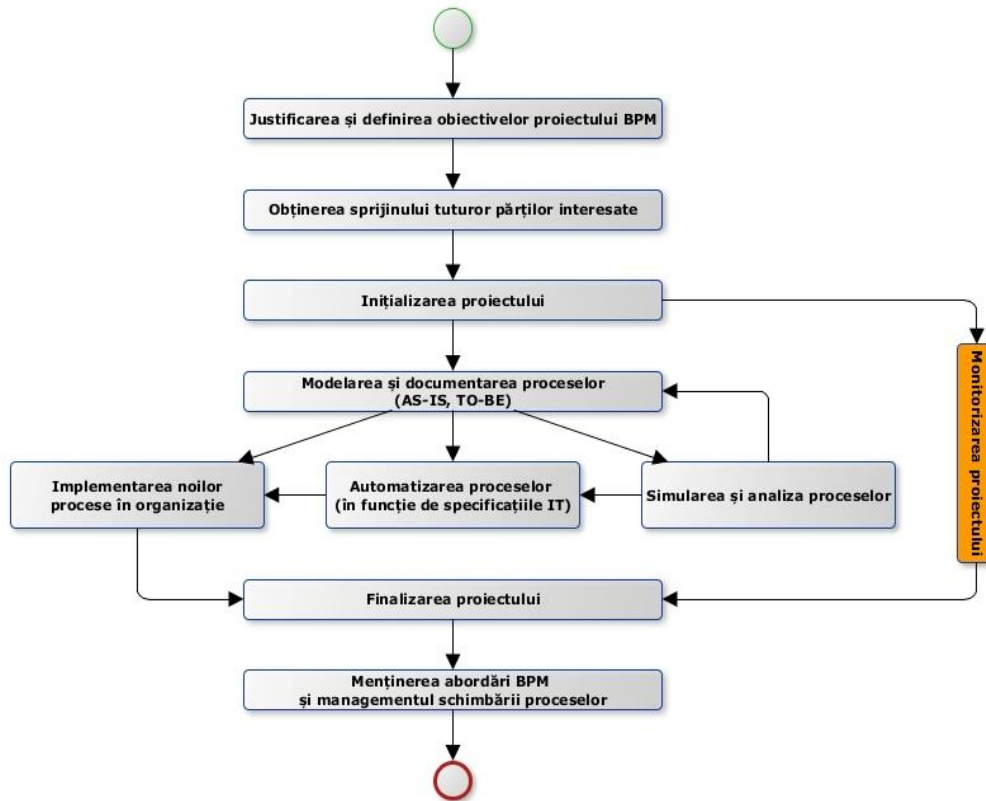


Fig. 5.1. Metodologia de implementarea a unui proiect BPM în viziunea CertiBPM

3. Inițializarea proiectului

Se pregătește manualul de proiect, adică setul de documente importante destinate definirii și impulsionalării proiectului. Manualul de proiect se va distribui tuturor părților implicate în proiect, pentru obținerea semnăturilor de confirmare. Manualul de proiect conține:

- Justificarea și obiectivele proiectului (măsurabile!);
- Descrierea echipei de proiect (management, supraveghetori, analiști, echipa tehnică, furnizorii de informații, mediatori externi și experți);
- Planul de proiect (repere, evenimente, sarcini) cu planul de monitorizare (necesar activității de *Monitorizare proiect*, Fig. 5.1);
- Calculul costurilor proiectului;
- Listă de procese cheie, care vor fi acoperite;
- Planul de risc (și compensații);
- Planul de comunicare;
- Planul de gestionare a configurației (versionare);
- Planul de calitate.

După finalizarea manualului de proiect, se vor executa următoarele activități:

- Membrii proiectului vor participa la cursuri de BPM;
- Lista de procese se va rafina/discuta;
- Se va obține documentația existentă (ISO, descrieri sumare de procese);

- Se vor stabili workshop-uri interactive cu proprietarii de proces și liderii de departamente;
- Se va obține confirmarea scrisă din partea managementului pentru fiecare proces ce urmează a fi analizat.

4. Modelarea și documentarea proceselor (AS-IS, TO-BE)

Această etapă începe cu identificarea listei proceselor (AS-IS). Nu toate procesele trebuie analizate și/sau descrise în detaliu. Este recomandat ca procesele identificate să fie prioritizate. Identificarea proceselor pornește din definirea obiectivelor. De exemplu, dacă obiectivul este obținerea unei certificări ISO, atunci se vor alege doar procesele cheie. Procesele sunt identificate de către managementul intern și consultanții externi.

Pentru identificarea modelului de proces vor avea loc interviuri și workshop-uri. Participanții la un workshop vor fi managerul de departament (sau înlocuitor), proprietarul procesului, executantul procesului, analistul de proces și mediatorul (externi). În timpul workshop-ului se va modela procesul cu implicarea tuturor participanților. Primul draft ar trebui să fie obținut în aproximativ 4 ore. Modelul obținut trebuie să aibă sens și toate ambiguitățile lui să fie rezolvate. Pentru modelare se va folosi BPMN. Modele astfel obținute sunt trimise spre evaluare managementului, urmând a fi rafinate în funcție de comentariile obținute. Faza de documentare a procesului AS-IS se încheie cu confirmarea modelului de proces și obținerea de semnături privind luarea la cunoștință.

Modelarea procesului are o durată ridicată. Efortul pentru elaborarea unui proces de către un analist de proces este estimat între 40 și 70 de ore. Pentru 10 procese cheie analizate se estimează un efort de 4-6 luni. Durata e ridicată pentru că analistul de proces trebuie să observe procesul din foarte multe perspective. *Durata ridicată a etapei modelării procesului este marele dezavantaj al acestei metode.*

Urmează identificarea problemelor și propunerea noului model de proces (TO-BE) cu aceste probleme rezolvate. Din nou, pentru elaborarea noului model de proces se va ține cont de obiectivele definite la început.

5. Simularea și analiza proceselor

Procesele propuse (TO-BE) sunt supuse simulărilor. Rezultatele obținute în urma simulărilor sunt analizate. Același proces se poate simula modificând parametrii de intrare (numărul de resurse, costurile, etc.). Procesul se va optimiza în funcție de rezultatele analizei. Această etapă este una iterativă, deoarece rezultatele ei influențează procesul TO-BE.

6. Automatizarea proceselor (în funcție de specificațiile IT)

Este faza în care modelul de proces este transformat în aplicație software. Acest lucru se poate face prin intermediul unui BPMS sau iBPMS [130] sau prin dezvoltare clasică de software.

7. Implementarea noilor procese în organizație

Este considerată cea mai dificilă fază din ciclul BPM, din cauza rezistenței la schimbare a angajaților. Pentru a avea succes, este recomandat a se folosi următorii pași:

- a. Implementarea trebuie planificată;
- b. Implementarea trebuie comunicată, pentru că este o schimbare în organizație. Se poate comunica printr-o scrisoare oficială a managementului sau printr-o ședință cu angajații;

- c. Resursele trebuie pregătite din timp (noi documente, mijloace software etc.);
- d. Trebuie neapărat ținute cursuri de training;
- e. Trebuie stabilită data oficială de start a noului proces.
- f. Trebuie observat cât de ușor este acceptat noul proces. Acest lucru se poate sonda folosind chestionare sau prin interviuri cu cei care execută procesul. De obicei, prima versiune de implementare a noului proces nu e cea mai bună.
- g. Procesul trebuie rafinat conform reacției părților implicate.
- h. Procesul trebuie reimplementat, după caz.

8. Finalizarea proiectului

Reprezintă etapa în care implementarea s-a terminat. Această etapă cuprinde activități de arhivare a tuturor artefactelor obținute (documente, software, mesaje, instrucțiuni). Arhivarea va avea loc prin respectarea trasabilității. Mai sunt întâlnite în această etapă discuții despre ce a mers bine sau nu. Discuțiile sunt importante pentru că pot duce la îmbunătățiri pentru proiectele BPM viitoare.

9. Menținerea abordării BPM și managementul schimbării proceselor

După cum se observă din ciclul de viață BPM, abordarea BPM trebuie să fie continuă. Într-o organizație se vor desemna acele persoane care să supravegheze abordarea BPM iar, după caz, să semnaleze (re)începerea unui proiect BPM.

Dezavantajele metodologiilor de identificare a modelului de proces ce au loc prin interviuri și workshop-uri [138], printre care și metodologia *CertiBPM*, sunt: etapele acestora durează foarte mult pentru că implică discuții lungi cu muncitorii (utilizatorii și realizatorii proceselor), implică un volum mare de muncă de analiză, observare atentă a participanților, precum și un volum mare de muncă de antrenare a acestora. Există diferențe între procesele actuale și modul în care acesta este percepute și descrise de utilizatori. Oamenii nu percep procesele holistice, mai ales dacă acestea sunt realizate ad-hoc sau dacă ele se schimbă continuu.

5.2. Metodologii de implementare Process Mining existente

Elementele cheie ce trebuie urmărite în cadrul unui proiect de *Process Mining* au fost descrise în capitolul 3. Deși s-au studiat intens algoritmi și s-au făcut alte cercetări teoretice în domeniul *Process Mining*, literatura are foarte puține referințe referitoare la metodologia de implementare a unui proiect de *Process Mining*. Metodologiile existente sunt următoarele: *PDM – Process Diagnostics Methodology (Metodologia de Diagnosticare a Procesului – MDP)*, *PDM în domeniul medical (PDM-HealthCare)*, *L** și *PM²*.

5.2.1. Metodologia PDM

Metodologia PDM a apărut în 2009 [35] și s-a dorit a fi o metodă rapidă de investigare a unui proces din trei perspective: perspectiva fluxului de control, perspectiva performanței și perspectiva organizațională. Metodologia ajută la a răspunde rapid la întrebările: „Oare noi deținem controlul?” sau „Oare sistemul de informații utilizat reflectă într-adevăr starea de concretă a procesului de afaceri?”. Metodologia PDM este redată în termeni BPMN în Fig. 5.2.

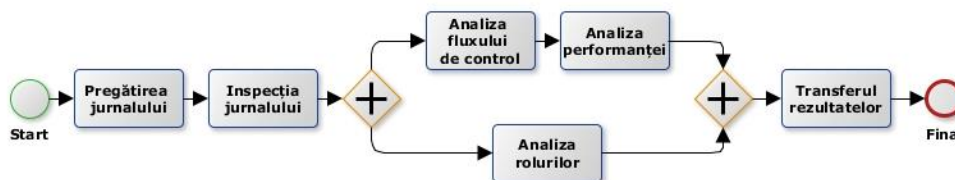


Fig. 5.2. Metodologia PDM

Metodologia PDM realizează investigarea unui proces folosind tehnici de *Process Mining* și are următoarele etape:

1. *Pregătirea jurnalului*: În acest pas sunt colectate datele care formează jurnalul de evenimente și sunt selectați candidații ce definesc un caz și evenimentele cazului;
2. *Inspectia jurnalului*: Este pasul în care analistul de proces se familiarizează cu procesul. Sunt colectate statisticile despre jurnal (număr cazuri, număr evenimente, număr mediu eveniment per caz, evenimente de start/final etc.);
3. *Analiza fluxului de control*: Este pasul în care este descoperit modelul de proces din jurnalul de evenimente. Acesta este comparat cu procesul documentat pentru a se vedea conformitatea cu acesta;
4. *Analiza performanței*: Pe baza modelului descoperit de proces se observă duratele activităților și a timpilor de așteptare, iar apoi se încearcă identificarea gâtuirilor;
5. *Analiza rolurilor*: Este o analiză organizațională de bază, în care sunt identificate rolurile resurselor și se poate observa dacă o resursă este generalistă sau specializată;
6. *Transferul rezultatelor*: Rezultatele sunt discutate cu clientul, pentru ca acesta să înțeleagă ce se întâmplă în realitate cu procesul.

Rezultatul PDM acoperă perspectiva controlului de flux (cum arată în realitate modelul procesului), perspectiva performanței (cât de eficient este sistemul) și perspectiva organizațională (cine este implicat în proces și cum). Rezultatele sunt de multe ori mai bine interpretabile dacă organizația este capabilă să interacționeze cu modelele găsite, să vizualizeze excepțiile din proces, și, în general, să aibă o mai bună percepție asupra procesului.

5.2.2. Metodologia PDM în domeniul medical

Procesele din domeniul medical sunt foarte dinamice, complexe, ad-hoc și implică multi-disciplinaritate. Procesele medicale sunt foarte variate și se pot grupa după tipul de îngrijire: primară, secundară, terțiară și de urgență [112]. Din punct de vedere al *Process Mining*, procesele medicale sunt grupate în categoriile:

- Îngrijire non-electivă cu sub-categoriile: procese urgente și procese imediate;
- Îngrijire electivă cu sub-categoriile: procese standard, procese de rutină și procese de non-rutină.

Metodologia PDM pentru domeniul medical (PDM-Healthcare) [138] a fost propusă în 2012 și aplicată la spitalul Sao Sebastiao, din nordul Portugaliei. S-a pornit de la metodologia PDM, și, după cum se vede în Fig. 5.3, este similară PDM, având în plus o activitate și anume *Analiza secvențelor de clustering*. Această activitate este un sub-proces prezentat în Fig. 5.4.

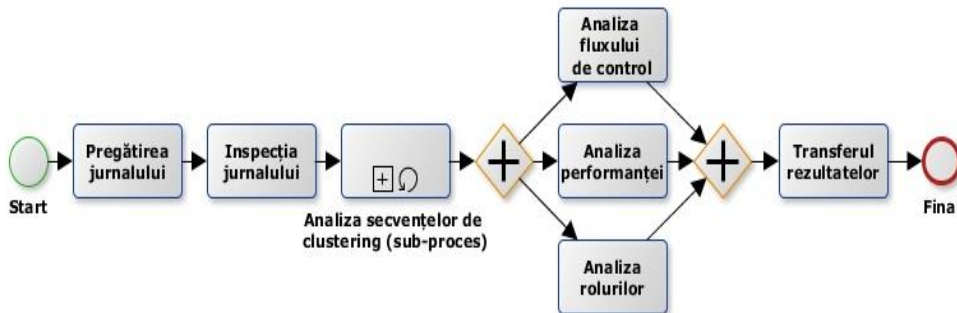


Fig. 5.3. Metodologia PDM-Healthcare

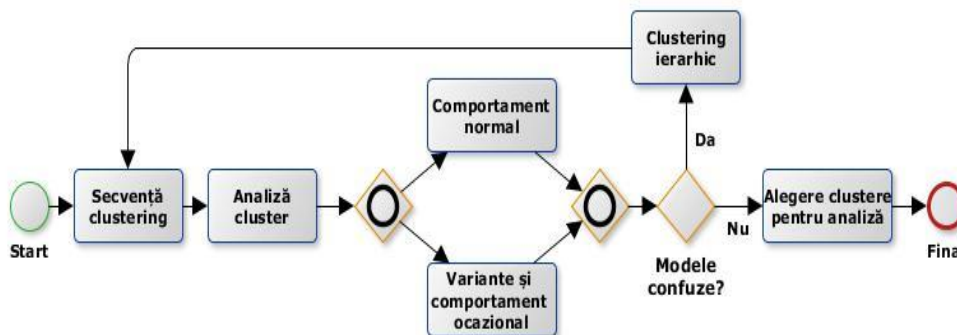


Fig. 5.4. Sub-procesul „Analiza secvențelor de clustering” din Metodologia PDM-Healthcare

Metodologia PDM nu a funcționat în domeniul medical pentru că nu dă rezultate bune pentru procesele complexe și ad-hoc. Problemele întâmpinate au fost: (i) jurnalele de evenimente utilizate sunt incomplete și conțin zgomote; (ii) există o multitudine de variante de proces; (iii) comportamentul ocazional nu poate fi ignorat, deoarece unele abateri de la regulament pot salva vieți.

Pentru domeniul medical, este importantă studierea comportamentului și a variantelor de proces ocazionale. Pentru a face acest lucru s-a extins cu o nouă activitate în urma inspekției jurnalului, *Analiză secvență clustering*. De fapt, acest nou pas este o sub-metodologie care include un set de tehnici pentru a grupa în clustere jurnalul și a pre-analiza procesul. Scopul nu este numai de a produce modele mai simple pentru etapele următoare, ca și acela de a sistematiza analiza variantelor de proces și comportament ocazional.

5.2.3. Modelul L*

Pentru a lua în considerare definirea și implementarea unui proiect de *Process Mining*, ar trebui în primul rând identificate cazurile de folosire a *Process Mining* în cadrul unei organizații. Orice caz de folosire are un obiectiv și un set specific de acțiuni pentru a îndeplini acest obiectiv (Fig. 5.5).

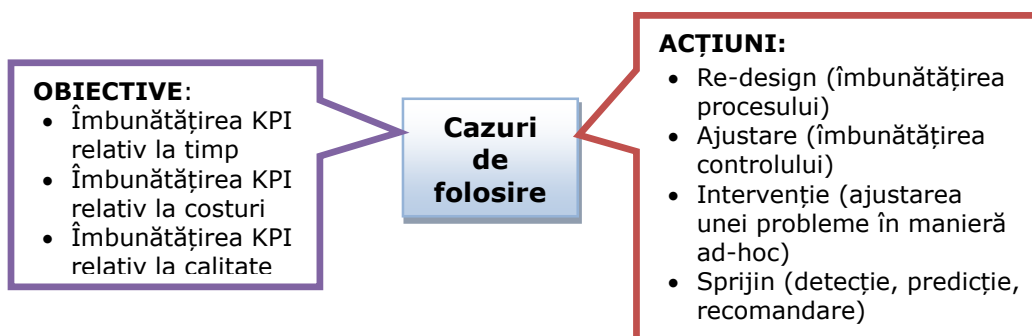


Fig. 5.5. Obiective și acțiuni în cazurile de folosire a Process Mining în organizație

Exemple de cazuri de folosire a *Process Mining* în organizații ar fi:

- Identificarea blocajelor pentru a declanșa un proces de re-design care reduce timpul de executare cu 30%;
- Identificarea problemelor de conformitate utilizând verificarea conformității. Unele dintre problemele de conformitate sunt rezultatul intervențiilor ad-hoc, în timp ce altele implică ajustări ale parametrilor utilizați pentru distribuirea muncii;
- Armonizarea a două procese după o fuziune și compararea cu procesele actuale. Scopul unei astfel de armonizări este acela de a reduce costurile;
- Prezicerea timpului de executare rămas pentru a îmbunătăți serviciul clienți;
- Furnizarea de recomandări pentru alocarea resurselor cu scopul utilizării echilibrate a lucrătorilor;
- Identificarea cazurilor excepționale care generează prea mult lucru suplimentar. Învățând profilul pentru astfel de cazuri, ele pot fi tratate separat, pentru a reduce timpul general de execuție;
- Vizualizarea celor mai complicate sau consumatoare de timp 10 cazuri pentru a identifica riscurile potențiale.

S-a introdus ciclul de viață L^* [2] pentru *Process Mining* care este asociat cu parcurgerea a patru etape corespunzătoare managementului unui proiect, după cum este prezentat în Fig. 5.6.

Acest model de ciclu de viață este similar cu modelele de referință ce descriu un proiect de *Data Mining* sau de *Business Intelligence*. Există două cicluri de viață mai reprezentative, și anume: Cross Industry Standard Process for Data Mining, CRISP-DM și Sample, Explore, Modify, Model and Assess, SEMMA – Specificare, Explorare, Modificare, Modelare și Apreciere (Fig. 5.7).

Etapele definite de ciclul de viață L^* sunt (Fig. 5.6):

Etapa 0 – Planificarea și justificarea ține seama de tipul proiectului (condus de date, condus de întrebări, condus de obiective). În această etapă se planifică proiectul și se justifică acțiunile ce vor avea loc;

Etapa 1 – Extragerea presupune localizarea, extragerea și transformarea datelor de evenimente. Tot în această etapă se colectează modele, artefacte, obiective, și întrebări și are loc explorarea domeniului actual de cunoștințe;

Etapa 2 – Crearea modelului pentru fluxul de control și conectarea la jurnalul de evenimente presupune definirea fluxului de control, care este fundamentul pentru orice proces. Astfel, este vital a se crea un model de flux de control foarte bine conectat la datele disponibile de evenimente; verificarea conformității și stabilirea alinierii sunt elemente cheie. Activitățile acestei etape au un grad foarte mare de iterativitate;

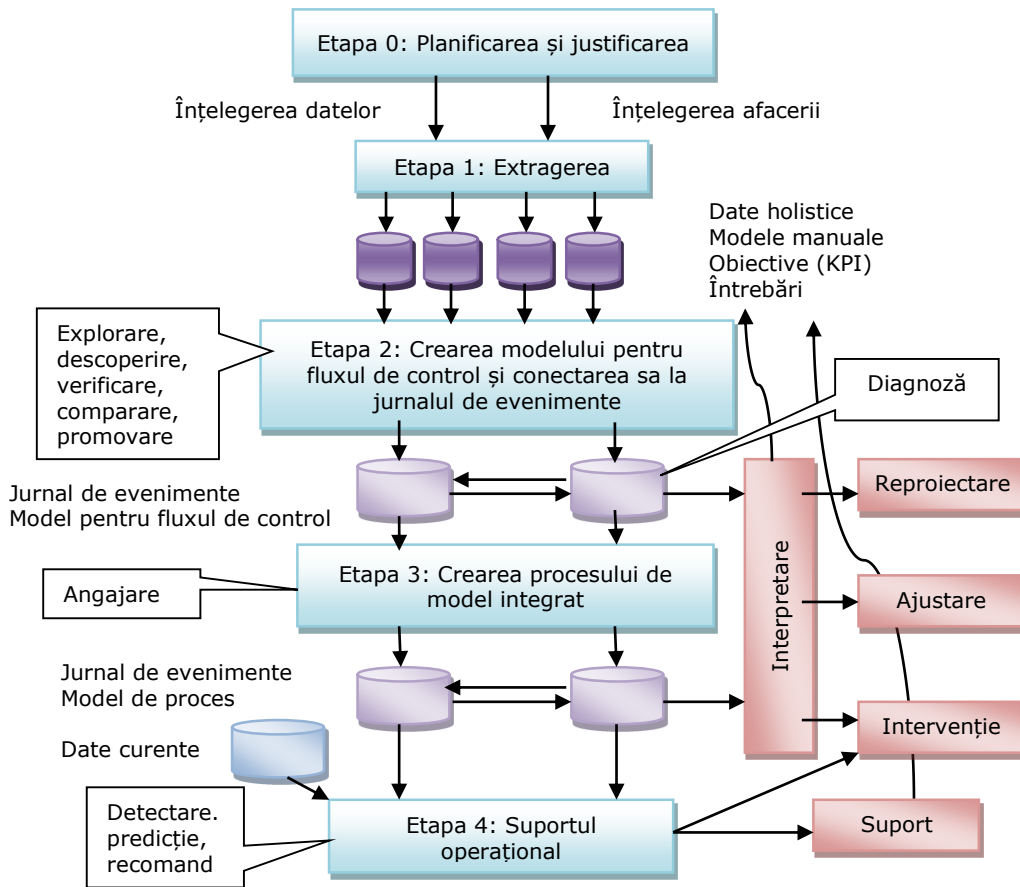


Fig. 5.6. Ciclul de viață L^* pentru managementul proiectelor de Process Mining

Etapa 3 – Crearea procesului de model integrat se referă la reluarea jurnalului de evenimente pentru a studia alte perspective: timp, date, resurse ș.a. Se recomandă fuzionarea mai multor modele, rezultând doar unul în care sunt reprezentate mai multe perspective;

Etapa 4 – Suportul operațional presupune folosirea datelor curente (pre-mortem) pentru detectarea deviațiilor, pentru predicții și recomandări (este posibil doar pentru procesele *Lasagna*).

Relația dintre ciclul de viață L^* și cazurile de folosire se observă în etapa 1 când sunt stabilite obiectivele, iar *Process Mining* conduce la informații de acțiune care se pot relaționa cu acțiunile cazurilor de folosire, după cum este prezentat în Fig. 5.8.

Ciclul de viață L^* se poate relaționa și cu cadrul de lucru pentru *Process Mining* rafinat: etapa 2 cuprinde activități precum: explorare, verificare, comparare, promovare, descoperire; etapa a 3-a cuprinde activități precum îmbunătățirea, iar etapa 4 cuprinde activități precum: predicția, recomandarea și detecția.

Procesele se situează între două limite, într-o parte având procesele structurate (numite și procese *Lasagna*), iar în cealaltă parte procese nestructurate (numite și procese *Spaghetti*). Modelul L^* a fost creat să trateze în special procesele structurate.

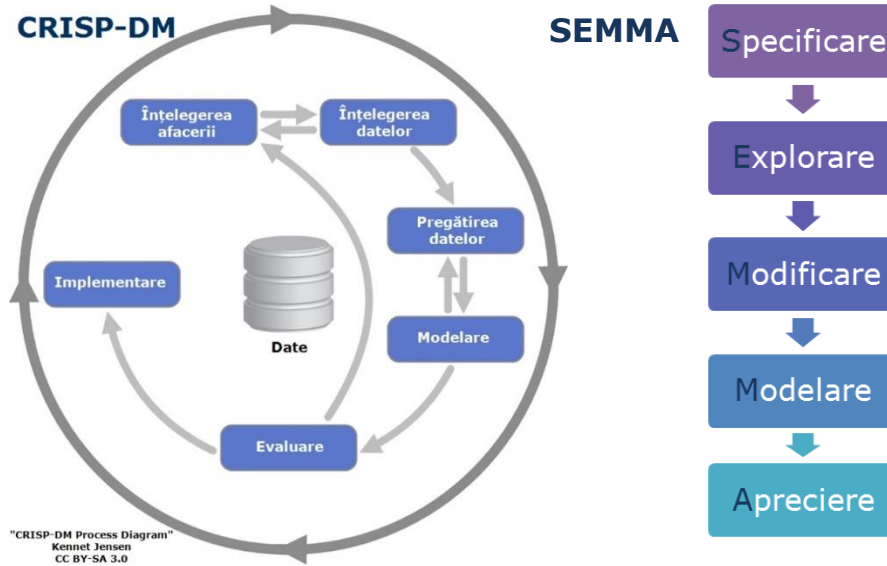


Fig. 5.7. Cicluri de viață reprezentative: CRISP-DM și SEMMA

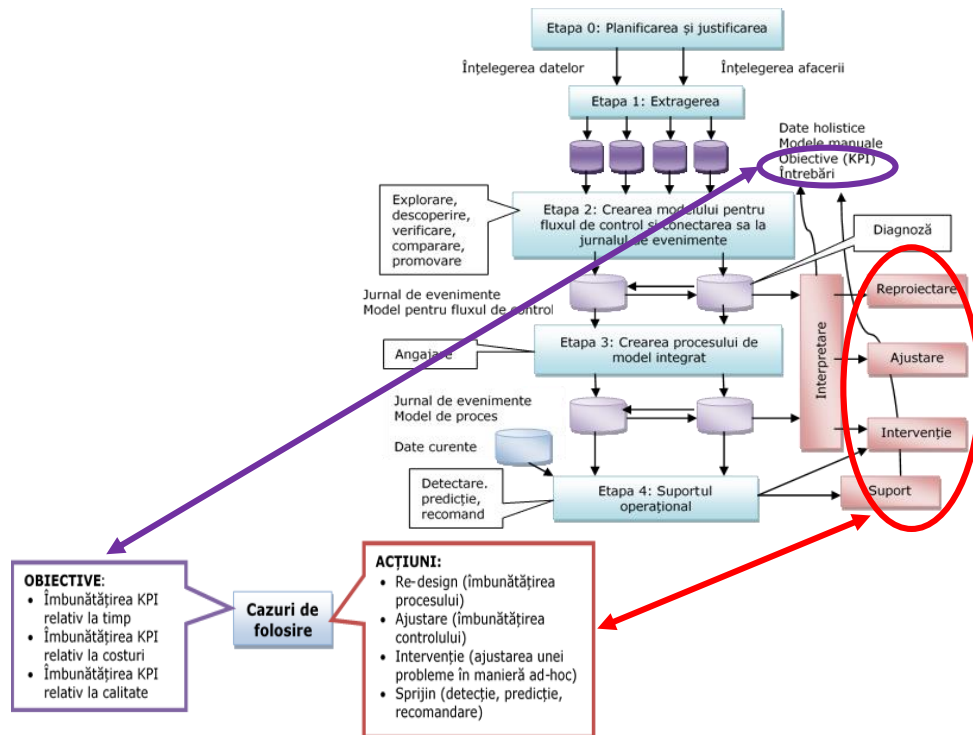


Fig. 5.8. Relația dintre ciclul de viață L* și cazurile de folosire

5.2.4. Metodologia PM²

PM² – *Process Mining Project Methodology (Metodologia de Proiect pentru Process Mining – MPPM)* este cea mai nouă metodologie apărută (iunie 2015) [68] și este clădită pe celelalte metodologii. S-a observat că metodologiile anterioare nu sunt potrivite pentru orice proiect. Sfera de acțiune a metodologiei PDM este foarte redusă, și nu este potrivită pentru proiecte complexe. L* acoperă mai multe tehnici, dar a fost conceput în primul rând pentru analiza proceselor structurate și are drept scop descoperirea un singur model integrat de proces. Nici L*, nici PDM nu încurajează în mod explicit de analiza iterativă.

Metodologia PM² are ca scop acoperirea analizei atât a proceselor structurate, cât și a celor nestructurate. Pentru fiecare stagi, PM² are definite intrările și ieșirile în mod clar, iar pașii ce trebuie executați sunt concreți. Metodologia PM² este prezentată schematic în Fig. 5.9. Metodologia are 6 etape care au ca intrări și ieșiri obiecte legate de scop, obiecte de date și modele. Obiectele legate de scop sunt: (i) întrebări de cercetare, (ii) constatări de conformitate, (iii) constatări de performanță și (iv) idei de îmbunătățire. Obiectele de date sunt (i) sisteme informatice, (ii) date de evenimente și (iii) jurnale de evenimente. Modele pot fi de proces sau analitice (de exemplu arbori de decizie). Cele 6 etape sunt:

1. *Planificarea*: are ca obiectiv planificarea proiectului (alegerea proceselor și formarea echipei) și determinarea întrebărilor de cercetare.
2. *Extracția*: conține determinarea dimensiunii extracției de date (perioadă, attribute etc.), extragerea datelor evenimentelor și transferul cunoștințelor de proces.

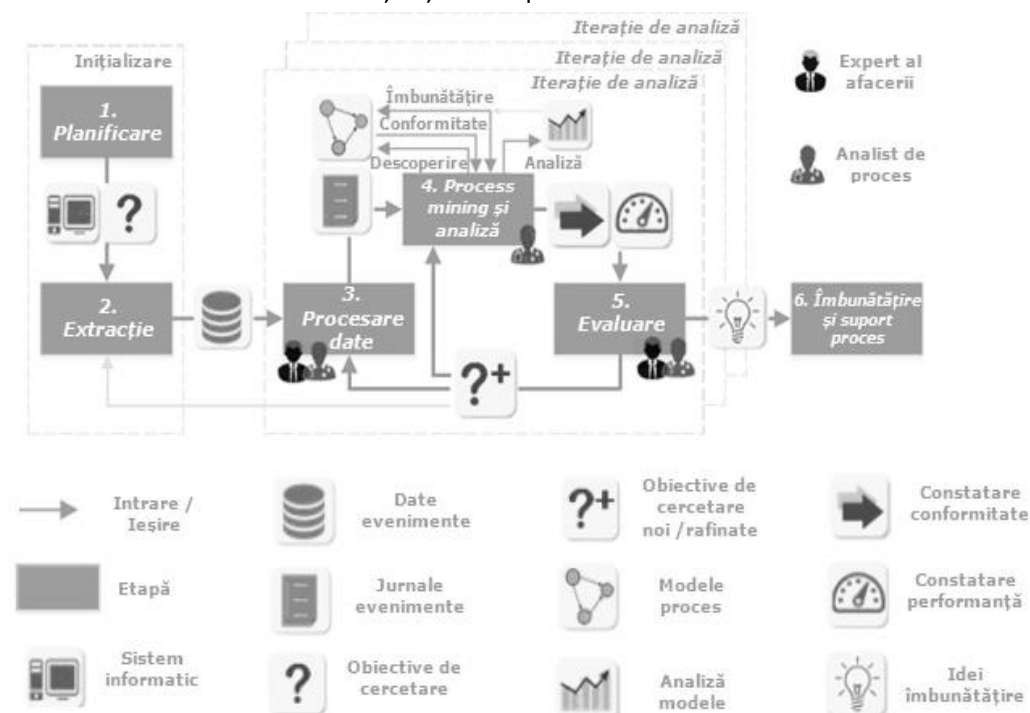


Fig. 5.9. Metodologia PM² [68]

3. *Procesarea datelor*: conține activități de creare a jurnalelor de evenimente, de agregare a evenimentelor, de îmbogățire și filtrarea jurnalelor.
4. *Process Mining și analiză*: conține cele trei stadii de *Process Mining* (descoperire, verificarea conformității, îmbunătățire) și un stadiu de analiză de proces specific *Data Mining*.
5. *Evaluarea*: conține activități de diagnoză, verificare și validare generatoare de idei de îmbunătățire.
6. *Îmbunătățire și suport proces*: au loc activități de implementare a sugestiilor de îmbunătățire și se oferă suport operațional pentru aceste implementări.

5.3. Metodologia CertiBPM îmbunătățită cu elemente de Process Mining

Pornind de la cele de mai sus metodologia propusă în cadrul prezentei teze de doctorat se constituie ca o combinație a implementării unui proiect BPM în versiunea *CertiBPM* cu îmbinarea simplistă de conducere a unui proiect de *Process Mining* din *PDM*, la care se adaugă conceptul iterativ din *PM²*. Metodologia a fost denumită *CertiBPM-PM*. Pentru proiecte cu procese complexe în care jurnalul de evenimente se creează dificil se poate recurge la metodologia *PDM* din domeniul medical, dar acest pas nu a fost introdus în modelul propus. După cum spuneam, analiza manuală a unui proces AS-IS durează între 40 și 70 de ore. Cu metodologia *CertiBPM-PM* această analiza va dura mult mai puțin, dacă jurnalul de evenimente este disponibil și prezintă un grad ridicat de calitate. Alt beneficiu este faptul că procesul descoperit este mult mai exact, adică este bazat pe evenimente concrete. Metodologia *CertiBPM* pentru implementarea unui proiect BPM este foarte detaliată și ea conține elemente de resurse umane, de psihologie organizațională, de planificare a schimbărilor, de comunicare etc.

Modelul propus este prezentat în Fig. 5.10. Acesta se aseamănă cu cel din Figura 5.1, doar că activitatea *Modelare și documentare a proceselor (AS-IS, TO-BE)* a fost descompusă astfel încât, în funcție de disponibilitatea jurnalului de evenimente, se poate alege ca modelarea proceselor (AS-IS) să se facă manual și/sau cu ajutorul tehnicilor de *Process Mining*. Pentru a implementa un proiect de BPM în manieră *Process Mining* trebuie parcurse următoarele etape:

1. *Justificarea și definirea obiectivelor proiectului BPM* (identic *CertiBPM*)
2. *Obținerea sprijinului tuturor părților interesate* (identic *CertiBPM*)
3. *Inițializare proiect* (identic *CertiBPM*)
Pentru fiecare proces au fost considerate următoarele activități:
4. *Modelarea manuală a procesului a procesului - AS-IS*: Activitatea inițială *Modelarea și documentarea proceselor (AS-IS, TO-BE)* a fost împărțită în trei activități și anume: *Modelarea manuală a proceselor AS-IS*, *Documentarea proceselor (AS-IS)* și *Modelarea și documentarea proceselor (TO-BE)*. Această activitate se poate executa ca mai înainte sau poate fi înlocuită de abordarea specifică *Process Mining*. Rezultate mai interesante pot rezulta din compararea abordării clasice cu abordarea specifică *Process Mining*.
5. *Descoperirea procesului și crearea unui model îmbunătățit cu ajutorul tehnicilor specifice Process Mining*:

- a. *Obținere jurnal* cu activități de extragere date (identic PDM)
 - b. *Inspecție jurnal* cu activități de corecția datelor și filtrare (identic PDM)
 - c. *Descoperirea procesului, analiza performanței lui și analiza organizațională*. Activități ce se pot desfășura paralel (identic PDM-Healthcare)
 - d. *Întocmirea raportului de analiză*: etapă în care sunt centralizate toate informațiile obținute în pașii precedenți sub forma unui raport ce urmează a fi prezentat. Etapa este utilă pentru pregătirea rezultatelor în vederea transferării lor.
 - e. *Transferul rezultatelor*: etapă în care sunt puse față în față rezultatele găsite și prezentate oamenilor implicați în aceste procese. Pentru a trage concluzii calitative se recomandă și compararea cu rezultatele modelării manuale a procesului sau cu procesul documentat în manualele de management al calității. (identic PDM)
Dacă modelul obținut trebuie ajustat, atunci pasul 5 se va relua, cu încercarea de a folosi alți algoritmi și de a aplica alte filtre asupra jurnalului de eveniment. Această abordare iterativă este specifică PM².
6. *Documentarea proceselor AS-IS*: este una din activitățile rezultate în urma divizarea activității *Modelarea și documentarea proceselor (AS-IS, TO-BE)*. În această activitate sunt documentate atât modelele obținute manual, cât și cele obținute prin abordarea Process Mining.
 7. *Modelarea și documentarea proceselor (TO-BE)*: Activitatea acesta este o activitate de propunere a modelelor TO-BE și poate avea un caracter iterativ, depinzând de simularea și analiza efectuată în etapa următoare.
 8. *Simularea și analiza proceselor* (identic CertiBPM)
 9. *Automatizarea proceselor (în funcție de specificațiile IT)* – identic CertiBPM
 10. *Implementarea noilor procese în organizație* (identic CertiBPM)
 11. *Finalizarea proiectului* (identic CertiBPM). De menționat că pe tot parcursul proiectului este importantă monitorizarea lui.
 12. *Menținerea abordării BPM și managementul schimbării proceselor* (identic CertiBPM).

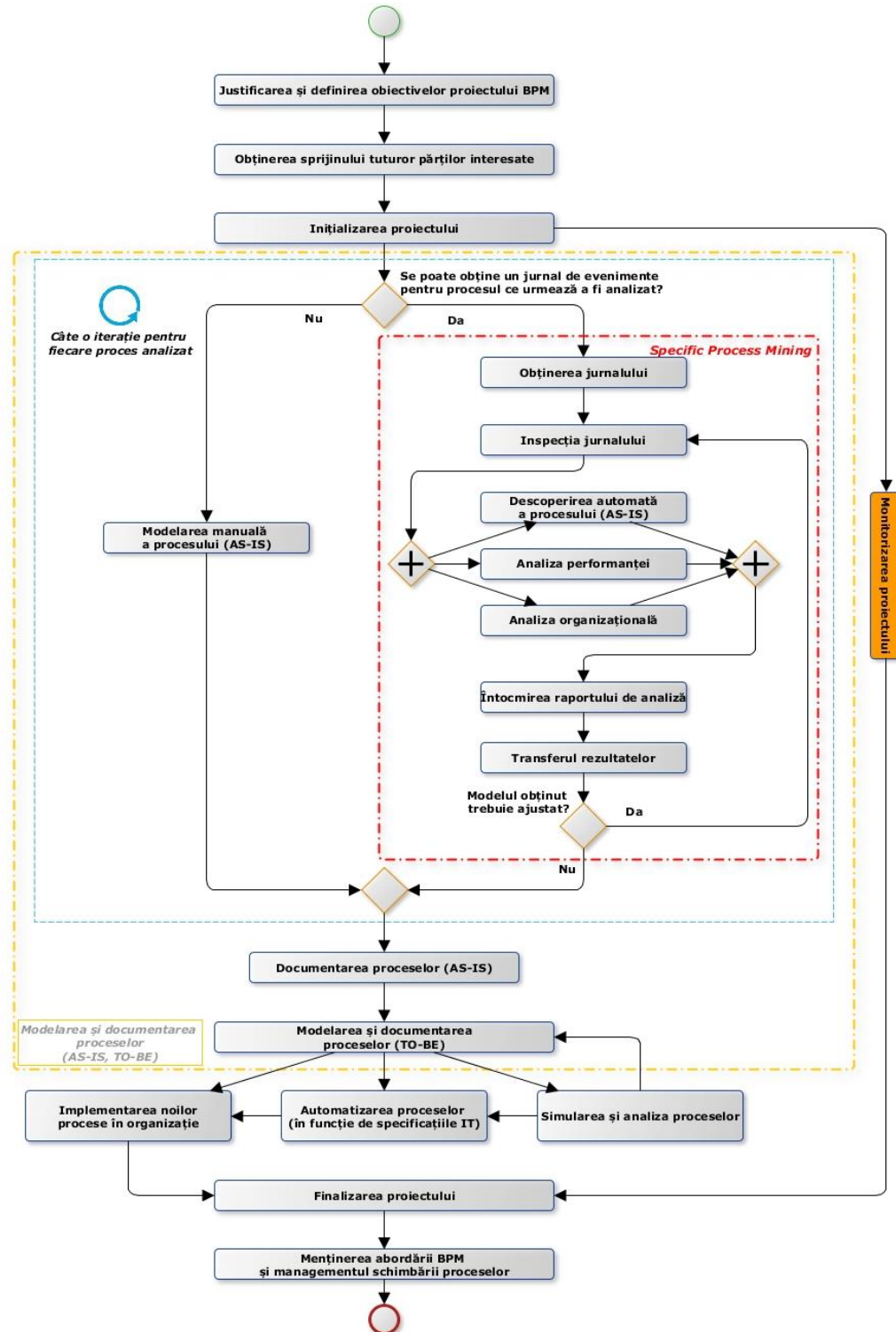


Fig. 5.10. Metodologia CertiBPM-PM

5.4. Concluzii

În prezentul capitol a fost stabilită și detaliată metodologia propusă pentru implementarea unui proiect BPM, îmbunătățită cu elemente specifice *Process Mining* pentru a conferi informații pentru afaceri (BI). Premisele metodologiei propuse, detaliile de aplicare a demersului de *Process Mining*, precum și mijloacele software disponibile pe piață la ora actuală au fost descrise detaliat anterior în capitolele 3 și 4.

Contextul de elaborare a metodologiei este determinat de modul de operare descris pentru implementarea proiectelor de BPM în organizații, în viziunea materialelor de formare aferente proiectului *CertiBPM*. Precizăm că, viziunea cunoașterii în contextul proiectului *CertiBPM* este una sintetică și integrată, datorită numeroaselor activități iterative, evolutive și corective derulate în procesul de transfer de inovare între specialiștii parteneri (și a echipele lor) în proiect. Astfel, cunoașterea și soluțiile inovative în domeniul implementării și derulării proiectelor de BPM, deținute de experți din Slovenia și Austria, au fost integrate și adaptate pentru situația organizațiilor din România (specificul proiectului a fost de transfer de inovare, dar acesta a facilitat și crearea de cunoaștere).

Evoluția domeniului BPM, mai ales în ceea ce privește metodologiile de conectare a sa cu demersul de *Process Mining*, sunt încă la început sau în faze incipiente de consolidare și generalizare în aplicare. Ca urmare, în prezentul capitol au fost panoramate unele metodologii existente de implementare a demersului de *Process Mining* menite să demonstreze nevoia organizațiilor pentru o mai bună percepție a proceselor sale și un mai bun management. Totodată, au fost precizate efectele, beneficiile obținute de organizații în cazul aplicării metodologiilor PDM, PDM-Healthcare, L* și PM², cum sunt: perspectiva controlului de flux, perspectiva performanței și perspectiva organizațională de execuție a proceselor.

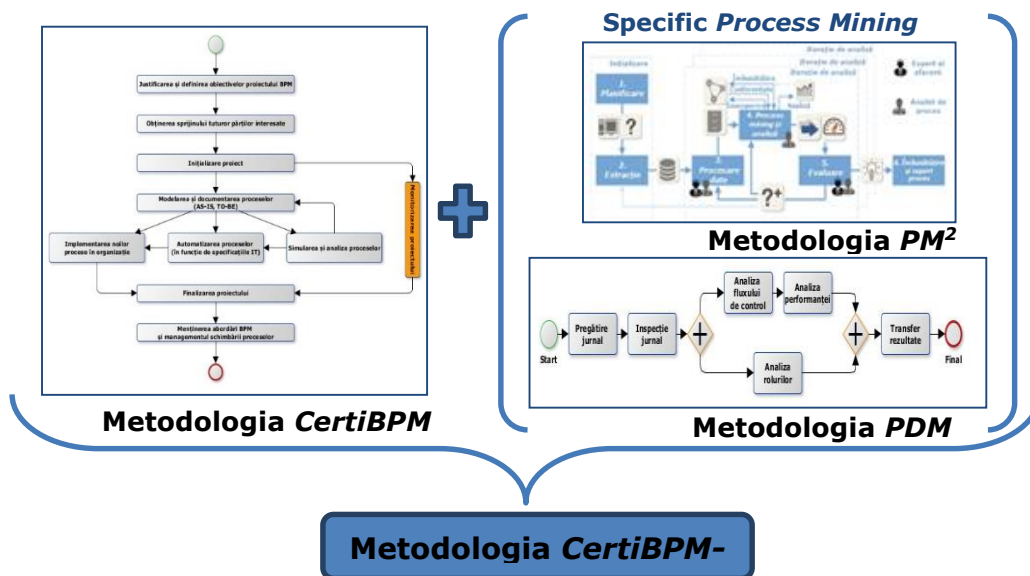


Fig. 5.11. Factorii determinanți care au contribuit la apariția metodologiei CertiBPM-PM

În finalul capitolului, s-a prezentat succint **metodologia CertiBPM-PM** (Fig. 5.10) concepută în cadrul cercetărilor și care se constituie ca o combinație a modului de implementare unui proiect BPM în versiunea *CertiBPM*, cu modul de conducere a unui proiect de *Process Mining* așa cum este el descris în metodologia PDM (Fig. 5.11). La această combinație s-a adăugat conceptul iterativ aferent metodologiei PM^2 (reprezentarea metodologiei *CertiBPM-PM* a fost realizată folosind mijloace de modelare grafică a proceselor).

Pentru a demonstra și a confirma eficiența și eficacitatea metodologiei *CertiBPM-PM* este nevoie de cercetări aplicative, experimentale. Aceste activități vor constitui subiectul capitolului 6 al prezentei teze de doctorat.

6. CERCETĂRI APLICATIVE PRIVIND APLICAREA DEMERSULUI METODOLOGIC CERTIBPM-PM ÎN VEDEREA IDENTIFICĂRII UNOR CĂI DE CREȘTERE A PERFORMANTELOR PROCESELOR DE AFACERI

Prezentul capitol își propune să demonstreze validitatea și utilitatea demersului de *Process Mining - Data Mining* în cazul concret al unei cercetări aplicative.

Obiectivul operațional aferent cercetărilor din acest capitol a fost: **OP6** Cercetări aplicative pentru demonstrarea modului de susținere a deciziei manageriale strategice prin demersul de *Process Mining - Data Mining* în cazul implementărilor de BPM în organizații (testarea și validarea metodologiei *CertiBPM-PM* propuse).

6.1. Descrierea contextului de cercetare

Metodologia *CertiBPM-PM* descrisă în capitolul 5.3 trebuie validată prin implementări de proiecte BPM. Pentru aceasta este nevoie de a se identifica companii sau instituții care să-și dorească îmbunătățirea proceselor prin sistemul BPM și care să posede un sistem informatic din care să se poată extrage date pentru construirea de jurnale de evenimente. Datele unei companii sunt foarte sensibile, iar din această cauză se manifestă o reticență față de asemenea abordări.

Acesta este motivul pentru care nu s-a găsit o companie locală care să îndeplinească cele două criterii mai sus enunțate. Metodologia a fost validată însă în cazul unei organizații de tip instituție publică, din Olanda. Datele acesteia au fost colectate în urmă cu doi ani în cadrul proiectului CoSeLog¹⁵. Jurnalul de evenimente din cadrul acestui proiect a fost pus la dispoziție de către însuși creatorul lui, domnul dr. Joos Buijs, de la care s-au obținut permisiunile de utilizare a acestuia în scopul declarat al cercetărilor din programul doctoral. De menționat este faptul că proiectul CoSeLog a avut alt obiectiv, și anume crearea unei structuri de tip „cloud” între municipii care să ofere servicii prin procese configurabile. Datele analizate (și care sunt cuprinse în jurnalul de evenimente) nu au mai fost folosite pentru validarea unor metodologii, cum este cea propusă prin cercetările descrise anterior (metodologia *CertiBPM-PM*).

Astfel, s-a considerat că datele furnizate ca urmare a derulării proiectului CoSeLog sunt potrivite pentru validarea metodologiei. Jurnalul de evenimente poate fi obținut public [40] din centrul de date 3TU¹⁶.

¹⁵ <http://www.win.tue.nl/coselog/wiki/start>

¹⁶ <http://data.3tu.nl/>

6.2. Considerente privind modul de aplicare a metodologiei de cercetare

După cum se poate observa în Fig. 5.10, elementul decisiv ce poate porni un proiect specific *Process Mining* este disponibilitatea unui jurnal de evenimente. Jurnalul de evenimente studiat conține un proces referitor la faza de recepție a unei aplicații pentru autorizația de mediu. Datele folosite au fost anonimizate de către autorul jurnalului. Nu se cunosc nici instituția și nici persoanele implicate.

Pentru o mai bună înțelegere, asupra jurnalului au fost operate traduceri în limba română: asupra numelor de activități și asupra anumitor atribute (activități, evenimente, cazuri, resurse ș.a.). Tabelul cu cele 27 de activități ale jurnalului este prezentat în ANEXA 3. Activitățile vor fi referențiate în continuare prin codul asociat (de exemplu, *A01* pentru activitatea *A01 Confirmare de primire*). Jurnalul de evenimente conține un proces având 1434 cazuri (instanțieri ale procesului) și 8577 evenimente. Jurnalul este în format XES.

6.2.1. Obținerea jurnalului

Obținerea jurnalului este explicată schematic în Fig. 6.1. Pașii ce trebuie urmați pentru obținerea jurnalului sunt următorii:

1. Bazele de date ale sistemului informatic trebuie analizate rezultând o schemă de import. În acest pas se vor identifica tabelele ce conțin date relevante și legăturile dintre ele (chei primare).
2. Jurnalul se poate crea astfel:
 - a. Prin aplicația XESame (prezentată în 4.2.1.3), unde schema de import este folosită pentru mapări dintre coloanele tabelor și atributele XES.
 - b. Se poate crea un fișier CSV din schema de import și apoi va fi transformat în fișier XES cu ajutorul plug-in-ului *CSVtoXES* [159] din ProM. În ProM, în secțiunea *Acțiune* se alege plug-in-ul *Convert CSV to XES (CSVtoXES)*, având ca intrare fișierul CSV. Se mapează coloanele fișierului către atributele XES, dintre care cele mai importante sunt: *concept:name* (la nivel de traseu) cu coloana ce identifică cazul, *concept:name* (la nivel de eveniment) cu coloana ce identifică evenimentul, *time:timestamp* cu timpul de completare al evenimentului.

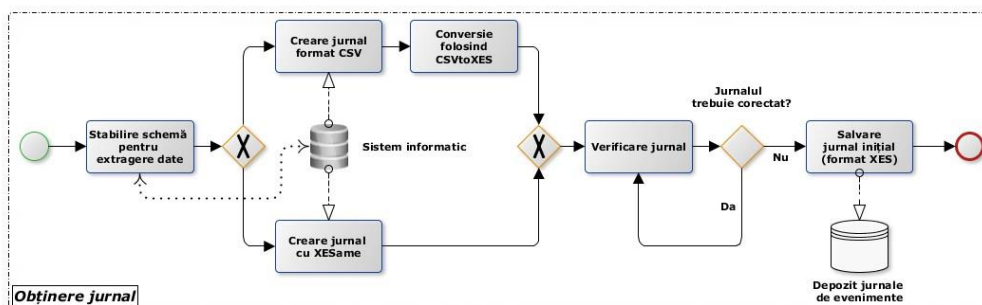


Fig. 6.1 Etapa „Obținerea jurnalului” explicată în detaliu (Metodologia CertiBPM-PM)

3. Se finalizează conversia, astfel încât jurnalul XES să fie disponibil pentru analiză în ProM și eventual să poată fi exportat.
4. Este indicată verificarea datelor de evenimente pentru erori folosind tehnici statistice și de inspecție vizuală, urmată de corectarea lor. Problemele ce pot apărea datorită datelor au fost detaliate în 3.3.4.
5. Jurnalul corectat care a fost creat se salvează într-un depozit de jurnale de evenimente în format XES. În continuare, acest jurnal va fi numit jurnal inițial pentru că el va constitui punctul de pornire al întregii analize.

În cazul jurnalului nostru, datorită mapării *time:timestamp* cu timpul de completare a evenimentului, activitățile vor avea sufixul „+complete”.

6.2.2. Inspecția jurnalului

Pentru inspecția jurnalului s-a folosit programul ProM descris în 4.2.1. Inspecția s-a desfășurat în 3 faze și este explicată schematic în Fig. 6.2.

1. Analiza de ansamblu folosind „Log Visualizer” (LV)

Jurnalul a fost importat în ProM, iar în secțiunea *Observare* > *Log Visualizer* (LV) acesta a fost inspectat. *Log Visualizer* se încarcă automat atunci când utilizatorul ajunge în secțiunea *Observare*. În paranteze vom nota de unde au fost culese datele, din ce zonă a ecranului. Următoarele informații au fost culese sau deduse:

- Un caz are între un eveniment și 25 evenimente (*Log Visualizer* > *Dashboard*), Fig. 6.3;
- Ca medie, avem 6 evenimente per caz. (*Log Visualizer* > *Dashboard*);
- Cazurile se întind pe perioada 2.10.2011 – 23.01.2013 (*Log Visualizer* > *Dashboard*), adică 478 zile calendaristice, sau 334 zile lucrătoare. S-au executat 1434 cazuri în cele 478 de zile calendaristice, deci avem în medie 3 cazuri pe zi. Dacă raportăm cazurile la zilele lucrătoare, avem în medie 4,28 cazuri pe zi;
- Există 27 de activități (*Log Visualizer* > *Dashboard*, notate cu „Event classes”). Un caz are între o activitate și 15 activități. În medie, un caz are 6 activități;

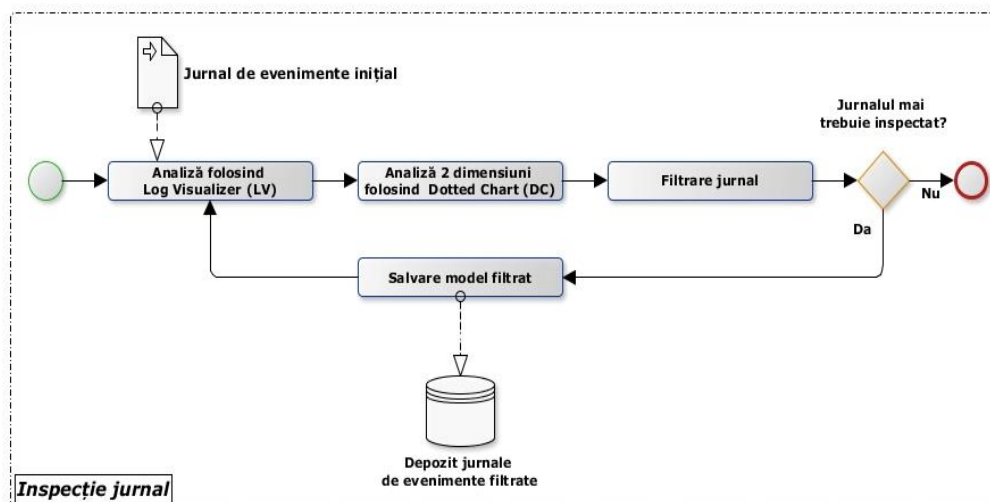


Fig. 6.2 Etapa „Inspecția jurnalului” explicată în detaliu (Metodologia CertiBPM-PM)



Fig. 6.3 Detalii despre proces din ProM > Log Visualizer > Dashboard (captură ecran)

- Se pot citi frecvențele fiecărei activități din cele 27 (*Log Visualizer > Summary > Activity classifier > All events*), Fig. 6.4. De exemplu, activitatea A05 apare de 1300 de ori, ceea ce reprezintă 15,157% din totalul de evenimente;
- Întotdeauna procesul începe cu activitatea A01 *Confirmare de primire*, în toate cele 1434 de cazuri (*Log Visualizer > Summary > Activity classifier > Start events*);
- Există 14 activități ce finalizează procesul (*Log Visualizer > Summary > Activity classifier > End events*). Activitățile cele mai frecvente pentru finalizarea procesului sunt A10 cu 828 apariții (57.741%) și A05 cu 400 apariții (27.894%);
- Există 116 cazuri netratate (*Log Visualizer > Summary > Activity classifier > End events*). Ele încep și se termină cu activitatea A01 *Confirmare de primire*;
- Resursa umană denumită *Resursa01* (angajat anonimizat) este cea mai ocupată, ea totalizând 1228 de evenimente finalizate, adică 14,317% din totalul de evenimente (*Log Visualizer > Summary > Resource classifier > All events*);
- Procesul este inițiat de 41 de persoane, între care *Resursa01* (de 195 ori – 13,598%) și *admin2* (de 114 ori – 7,95%) sunt inițiatorii principali (*Log Visualizer > Summary > Resource classifier > Start events*);
- Procesul este finalizat de 35 de persoane, între care *admin1* (de 249 ori – 17,364%) și *Resursa01* (de 195 ori – 13,598%) sunt finalizatorii principali (*Log Visualizer > Summary > Resource classifier > End events*);
- Grupul de resurse umane denumit *Grupul 1* (grup anonimizat) este cel mai ocupat, el totalizând 3152 de evenimente finalizate, adică 36,749% din totalul de evenimente (*Log Visualizer > Summary > Group classifier > All events*);
- Procesul este inițiat de persoane care nu aparțin unui grup anume, denumit generic *NEDEFINIT*. Frecvența este de 1136 ori – 79,219% (*Log Visualizer > Summary > Group classifier > Start events*);

- Procesul este finalizat de *Grupul 1* de 881 ori – 61,437% (*Log Visualizer > Summary > Group classifier > End events*).

Din captura ecran prezentată în Fig. 6.4 se poate observa că primele șase activități (ordonate descrescător după frecvență) sunt foarte frecvente și au valori aproximativ egale, între 14,959% și 16,719%. Se poate observa că frecvența activității *A10* este 14,959%, iar următoarea activitate *A03* are frecvența 0,641%, iar celelalte activități au frecvență nesemnificativă.

Primele șase activități însumează 94,53% din totalul de activități executate în acest proces. Restul de 21 activități însumează doar 5,47% din comportamentul procesului. Deci, se impune filtrarea jurnalului pentru a descoperi un model de proces cât mai simplu.

2. Analiza de ansamblu în 2 dimensiuni folosind „Dotted Chart” (DC)

Dotted chart (DC) [154] este un plug-in ProM prin care se pot observa multe detalii despre procesul de analizat. Acest plug-in este recomandat a se folosi inițial, deoarece oferă o imagine de ansamblu asupra evenimentelor analizate. Pornind de la acest chart, concluziile inițiale ne pot ajuta în determinarea analizei pe care o vom face. Plug-in-ul *Dotted chart* execută vizualizări asupra jurnalului în două dimensiuni, dintre care o dimensiune este întotdeauna timpul. A doua dimensiune poate fi (în paranteză este prezentată valoarea din ProM ce trebuie selectată pentru utilizare): cazul (*Cases*), resursa (*org:resource*), grupul (*org:group*), activitatea (*concept:name*) și altele, dacă acestea au fost definite. *Dotted Chart* este prezent în două versiuni: o versiune ca opțiune de vizualizare în secțiunea *Observare (XDottedChart)* și altă versiune ca plug-in în secțiunea *Acțiuni (Analyze using Dotted Chart)*, amândouă având funcționalități similare. *Dotted Chart* permite panoramarea și ordonarea a evenimentelor.

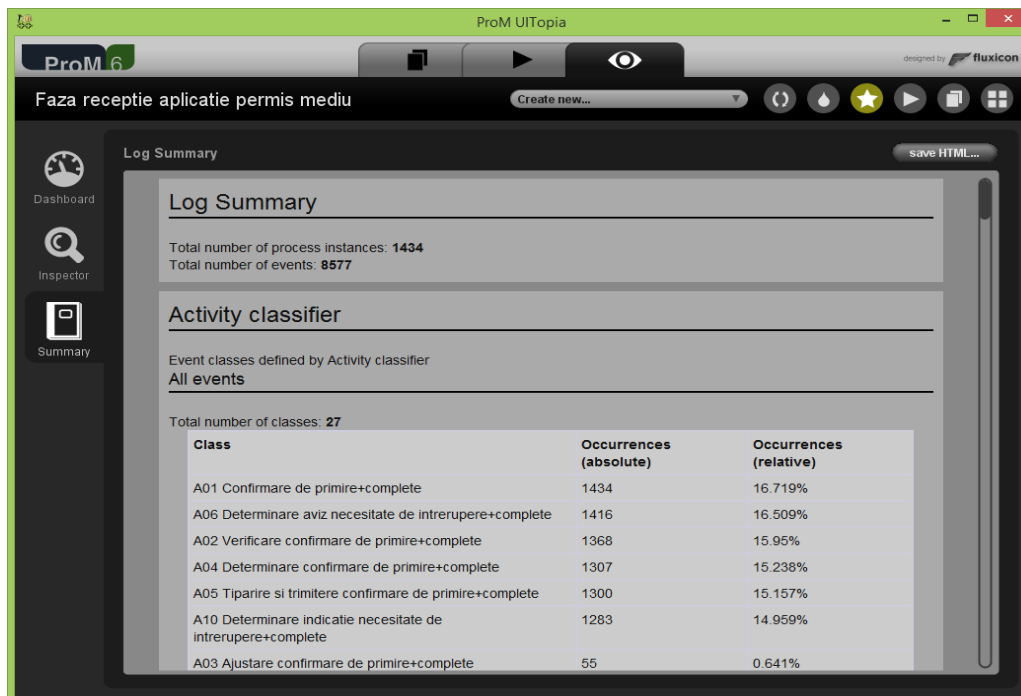


Fig. 6.4 Detalii despre proces din ProM > Log Visualizer > Summary (captură ecran)

În analiză s-a folosit XDottedChart și s-au reprezentat cazurile în raport cu timpul. Ordonarea cazurilor a fost crescătoare după prima apariție a unui eveniment în interiorul unui caz. S-a încercat să se descopere vizual anumite tipare de funcționare a procesului.

Dotted Chart s-a configurat în dimensiunile timp și caz (Cases). S-au formulat următoarele concluzii:

- Rata de sosire a cazurilor este relativ constantă. În Fig. 6.5 se poate observa că evenimentul de start A01 este reprezentat printr-un punct albastru. Dacă rata de sosire a cazurilor ar fi constantă, acest lucru s-ar reflecta prin prezența unei linii de puncte albastre din colțul stânga-sus către colțul dreapta-jos. Punctele albastre au fost încadrate între două linii drepte foarte apropiate între ele. De aici putem concluda că rata de sosire este relativ constantă. Dacă existau deviații, puteam concluda că procesul este sezonier. Rata de sosire a cazurilor este foarte importantă pentru a studia necesitatea de personal și resurse în momente cheie.
- Dacă se mărește zona din colțul stânga-sus, Fig. 6.6, se poate observa că sunt foarte puține cazuri care încep în acea lună (Octombrie 2011). Se pare că în momentul în care s-a pus procesul în funcțiune a existat o perioadă de acomodare cu foarte puține cazuri.
- Există cazuri în care mai multe activități sunt prelucrate în loturi (*batch-processed*). Acest tipar se poate observa din gruparea pe verticală a mai multor puncte în Fig. 6.7.
- Din punct de vedere global, se observă o modificare în timp a procesului din perspectiva duratei unui caz. Se observă că spre finalul perioadei de analiză procesele se termină mai repede, adică s-a ajuns la o fază de maturitate. Fiecare activitate este reprezentată printr-un punct. Un caz este o înșiruire orizontală de puncte. Cu cât ele sunt mai dispersate orizontal, cazul durează mai mult, iar cu cât sunt mai aglomerate orizontal, cazul durează mai puțin. Trendul se poate observa în Fig. 6.8 și putem considera luna Iulie 2012 ca moment al schimbării de trend.

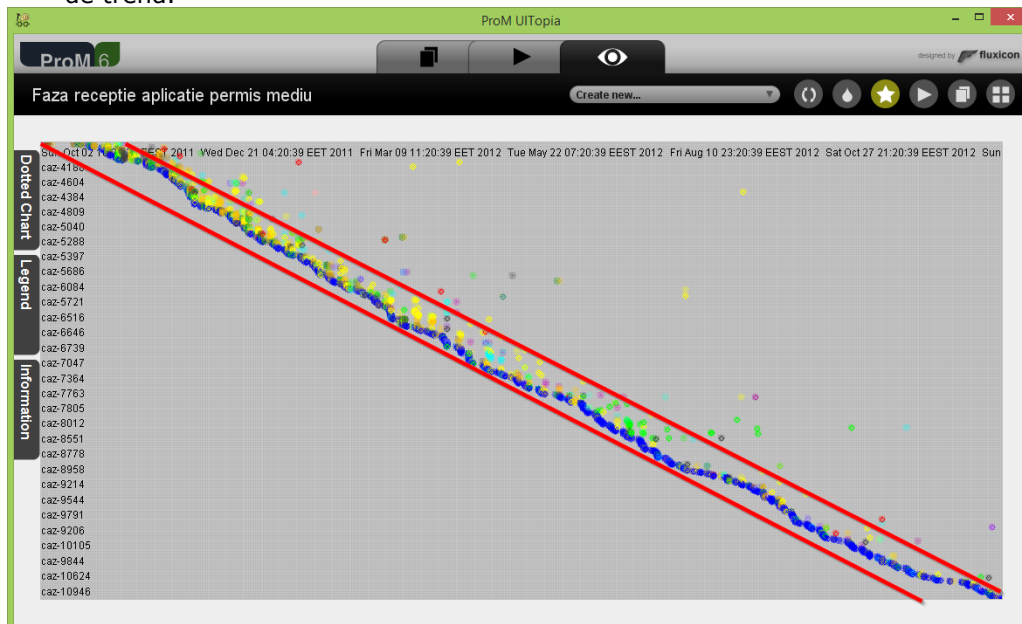


Fig. 6.5 Investigarea ratei de sosire a cazurilor cu ajutorul XDottedChart

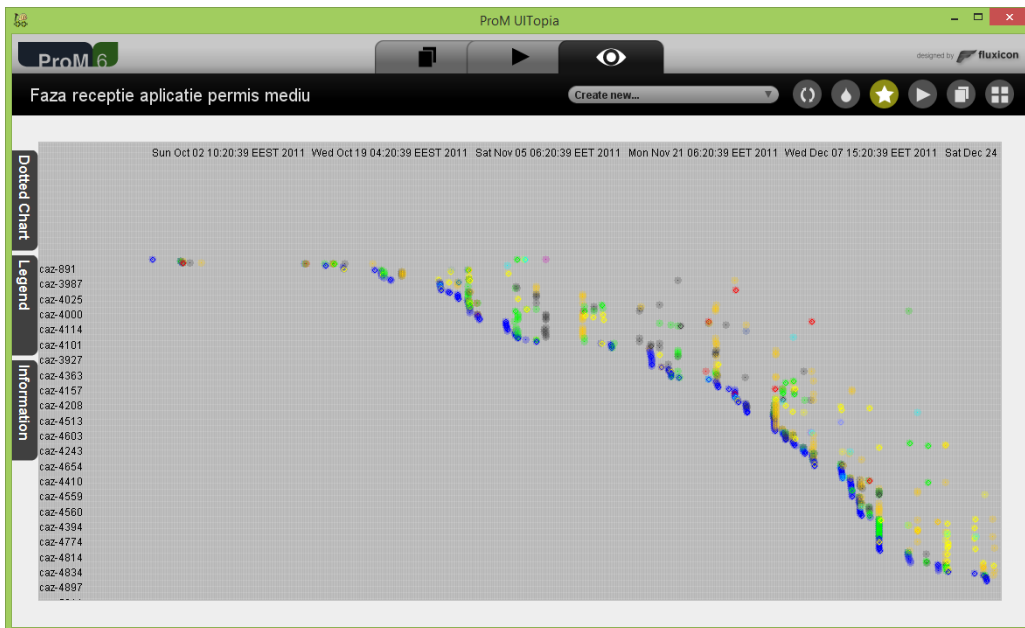


Fig. 6.6 Investigarea zonelor cu cazuri puține cu ajutorul XDottedChart



Fig. 6.7 Investigarea prelucrărilor activităților în loturi cu ajutorul XDottedChart

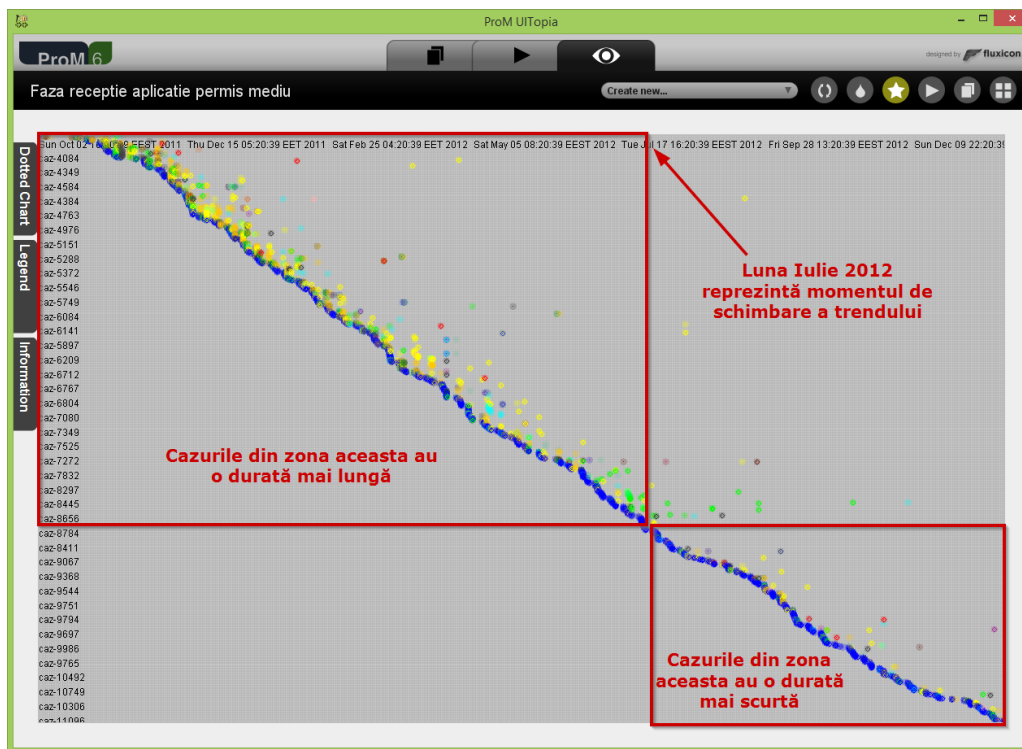


Fig. 6.8 Investigarea zonelor cu cazuri având durate diferite cu ajutorul XDottedChart

Dotted Chart s-a configurat în dimensiunile timp și resursă (*org:resource*). S-au putut formula următoarele concluzii:

- Cu cât resursa umană execută mai multe activități, cu atât această resursă are mai multe puncte (ce corespund activităților) repartizate pe un rând.
- Se poate observa atât perioada în care resursa a fost activă, cât și perioada în care resursa a fost inactivă.
- Există resurse specializate care execută doar anumite tipuri de activități. *Observația 1* din Fig. 6.9 arată că *Resursa02* execută doar activități marcate cu portocaliu (culoarea portocaliu corespunde activității *A05*). *Observația 2* din Fig. 6.9 arată că *Resursa24* execută doar activități marcate cu verde (culoarea verde corespunde activității *A02*). *Observația 3* din Fig. 6.9 arată că activitatea majoritară efectuată de *Resursa03* și *Resursa27* este *A01* (care este reprezentată printr-un punct albastru). *Observația 4* din Fig. 6.9 arată că resursele *Resursa30* și *Resursa19* execută doar două activități: *A02* (marcată cu verde) și *A04* (marcată cu gri). Aceste resurse prezentate mai sus sunt resurse specializate care execută doar anumite activități, iar restul sunt resurse generaliste care pot executa un număr ridicat de activități.
- Există resurse care nu efectuează activități de la începutul procesului și care se alătură mai târziu la executarea activităților procesului cum sunt: *Resursa22*, *Resursa14* și *Resursa42* (Fig. 6.10).



Fig. 6.9 Investigarea rolurilor resurselor cu ajutorul XDottedChart



Fig. 6.10 Investigarea momentelor de începere a lucrului cu ajutorul XDottedChart

- Datorită faptului că este înregistrat doar momentul în care se încheie activitatea, nu putem trage concluzii exacte despre cât de ocupată a fost o resursă. Dacă jurnalul de evenimente conținea și momentul de începere a unei activități, atunci puteam cunoaște exact datele când resursa era ocupată și când resursa stătea nefolosită.

Dotted Chart s-a configurat în dimensiunile timp și resursă (*org:resource*). S-a ajuns la concluzia că avem de a face cu grupuri specializate care execută doar un tip de activități (Fig. 6.11). În Tabelul 6.1 sunt expuse activitățile executate de grupuri.



Fig. 6.11 Investigarea rolurilor grupurilor cu ajutorul XDottedChart

Tabelul 6.1. Grupuri specializate pe anumite activități

Grup	Activitate	Culoare
Grupul 1	A06 Determinare aviz necesitate de intrerupere	Galben
Grupul 4	A02 Verificare confirmare de primire	Verde
NEDEFINIT	A01 Confirmare de primire	Albastru
Grupul 3	A04 Determinare confirmare de primire	Gri
Grupul 2	A05 Tiparire si trimitere confirmare de primire	Portocaliu
Grupul 13	A07-1 Draft intern aviz etapa 1	Verde închis

3. Filtrarea jurnalului de evenimente

Pentru filtrare se pot folosi următoarele plug-in-uri ProM: *Filter Log using Simple Heuristics (FLuSH)*, *Filter Log by Attributes (FLA)* și *Time Based Log Filter (TBLF)*. Caracteristicile plug-in-urilor de filtrare sunt prezentate în Tabelul 6.2.

Jurnalul inițial a fost supus filtrului *FLuSH* cu următorii parametri:

- Activități de start = {A01}, frecvență 100%
- Activități de final = {A01,A05,A10}, frecvență 93%
- Activități utilizate = {A01,A02,A04,A05,A06,A10}, frecvență 97% (Fig. 6.12)

Tabelul 6.2. Caracteristicile plug-in-urilor de filtrare (ProM)

Plug-in	Caracteristici filtrare
<i>FLuSH</i>	<ul style="list-style-type: none"> - Filtrează cazurile care au un anumit tip de tranzacție (de exemplu, filtrare cazuri care au fost suspendate) - Filtrează cazurile care încep/se termină cu anumite activități, ținând cont și de frecvența lor - Filtrează cazurile care conțin anumite activități, ținând cont și de frecvența lor
<i>FLA</i>	<ul style="list-style-type: none"> - Filtrează cazurile după atributele cazului: resursa, grup, alte atribute (de exemplu, filtrare cazuri atribuite unei anumite resurse) - Filtrează cazurile după atributele evenimentelor lor: resursa, grup, alte atribute (de exemplu, filtrare cazuri ce conțin evenimente executate de o anumită resursă) - Filtrează cazurile ce au o anumită lungime în funcție de numărul de evenimente ce le conțin
<i>TBLF</i>	<ul style="list-style-type: none"> - Filtrează cazurile din jurnal după duratele lor (de exemplu, filtrare cazuri care durează între 2 și 3 luni) - Filtrează cazurile din jurnal după datele lor de început (de exemplu, filtrare cazuri din perioada Iulie – Septembrie 2012)

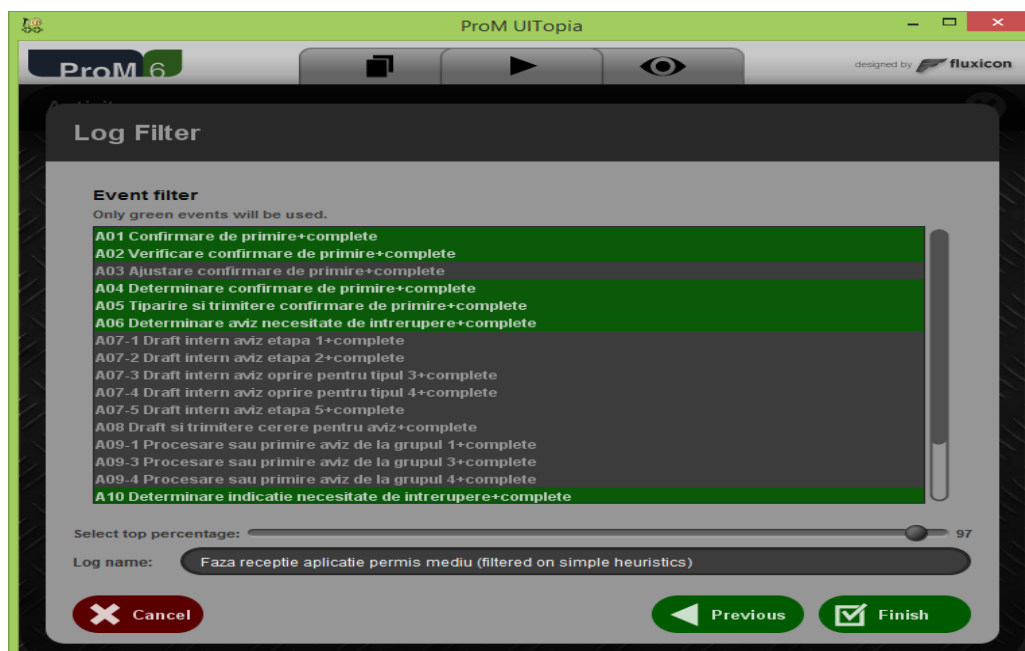


Fig. 6.12 Filtrarea jurnalului inițial cu plug-in-ul *FLuSH* (captură ecran)

A rezultat un jurnal ce conține doar activitățile {A01,A02,A04,A05,A06,A10} și care are 1344 cazuri (93,7% din cazurile jurnalului inițial) cu 7612 evenimente (88,7% din evenimentele jurnalului inițial).

Jurnalul inițial are un atribut la nivel de caz denumit *channel* și care reprezintă modul în care s-a pornit procesul. Fiind un proces din sfera serviciilor publice, el se poate porni în momentul primirii documentației de solicitare. Canalele prin care se poate depune cererea/solicitarea sunt: *Ghișeu*, *Internet*, *E-mail*, *Poștă* sau *Intern*. În Fig. 6.13 este reprezentată distribuția procentuală a canalelor de solicitare din totalul de cazuri.

Pornind de la acest atribut se ivesc noi întrebări de cercetare: cum arată procesul pentru cei care predau cererea la ghișeu? Cum arată procesul pentru cei care trimit cererea pe Internet? Care sunt diferențele din punct de vedere al procesului pentru cei recurg la Internet față de cei care predau la ghișeu? De ce cea mai mare rată de solicitare este prin Internet?

Pentru a răspunde acestor întrebări, avem nevoie să filtrăm jurnalul de evenimente folosind filtrul *FLA* (Fig. 6.14). Cu ajutorul *FLA*, jurnalul inițial a fost împărțit în două jurnale, primul care conține toate cazurile în care s-a solicitat de pe Internet, iar cel de-al doilea care conține toate cazurile în care s-a solicitat de la ghișeu.

Se pot obține și alte jurnale folosind filtrul *FLA*. De exemplu, se poate crea un jurnal filtrând după o anumită resursă. Modelul descoperit și îmbogățit cu detaliile de performanță poate fi comparat cu modelul rezultat din jurnalul inițial. Astfel, se poate observa performanța resursei în cadrul procesului.

FLuSH și *FLA* se găsesc în secțiunea *Acțiuni*. Spre deosebire de acestea, filtrul *TBLF* se găsește în secțiunea *Observare*. Cu acesta se pot crea jurnale pentru o anumită perioadă; de exemplu se pot crea jurnale pentru fiecare lună sau semestru. Modelele create pentru fiecare lună se pot compara, putându-se observa evoluția modelului în timp.

Toate jurnalele prelucrate se vor salva într-un depozit de jurnale, ele urmând a fi folosite pe măsură ce analiza progresează.

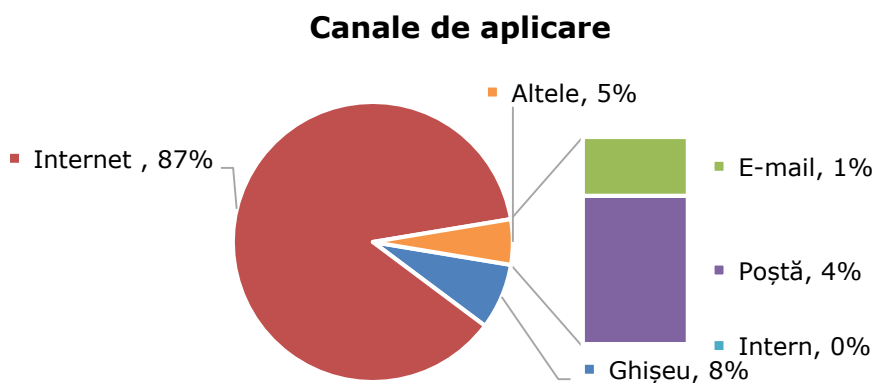


Fig. 6.13 Distribuția cazurilor în funcție de canalele de aplicare

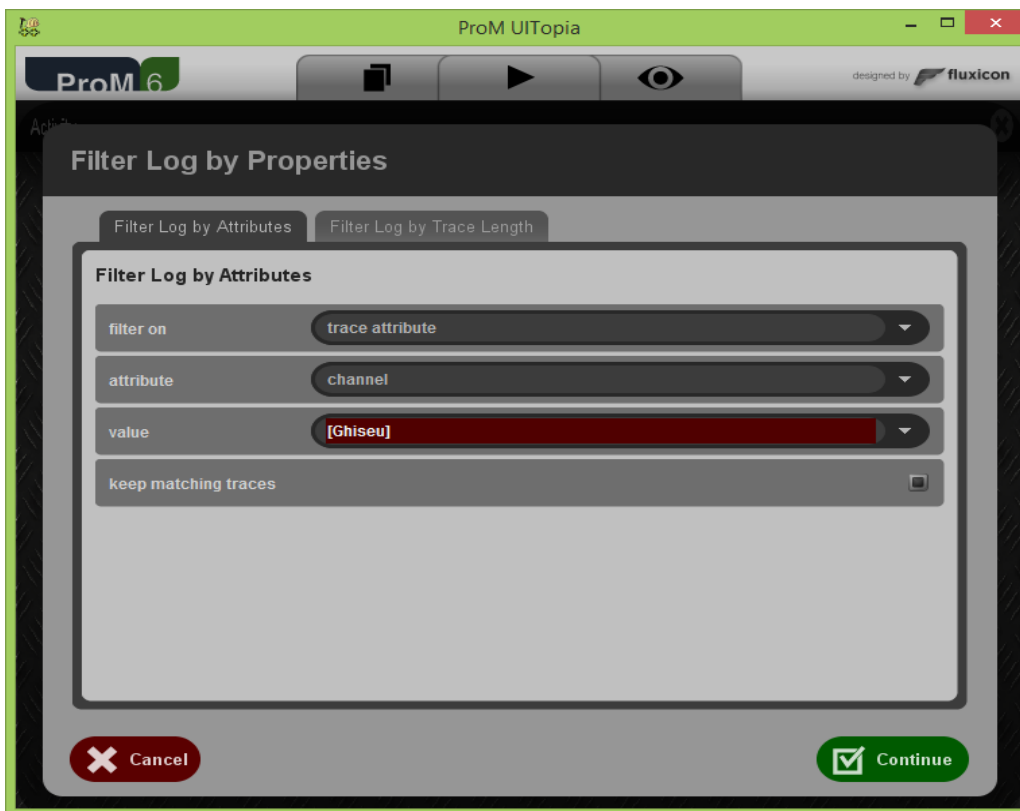


Fig. 6.14 Filtrarea jurnalului inițial cu plug-in-ul FLA (captură ecran)

6.2.3. Descoperirea procesului

După ce jurnalul a fost inspectat și filtrat, urmează să dezvoltăm modelul, analizând controlul fluxului. Modelele se vor crea atât din jurnalul inițial, cât și din jurnalul filtrat cu *FLuSH*. Modelele se vor crea folosind următorii algoritmi: algoritmul σ , algoritmul euristic, algoritmul fuzzy și algoritmul inductiv. Rezultatele vor fi convertite în rețele Petri, iar apoi în BPMN. La finalul acestei secțiuni, vom concluda cu alegerea celui mai bun model.

6.2.3.1. Algoritmul α

Algoritmul α (3.4.1) este implementat în platforma ProM prin plug-in-ul *Mine for a Petri Net using Alpha-algorithm (MPNAA)* [167]. Rezultatul acestui algoritm este o rețea Petri. Aceasta va fi convertită în format BPMN prin plug-in-ul *Convert Petri net to BPMN (CPNB)*.

Modelul descoperit cu algoritmul α pentru jurnalul inițial este prezentat în Fig. 6.15, iar modelul descoperit cu algoritmul α pentru jurnalul filtrat cu *FLuSH* este prezentat în Fig. 6.16.

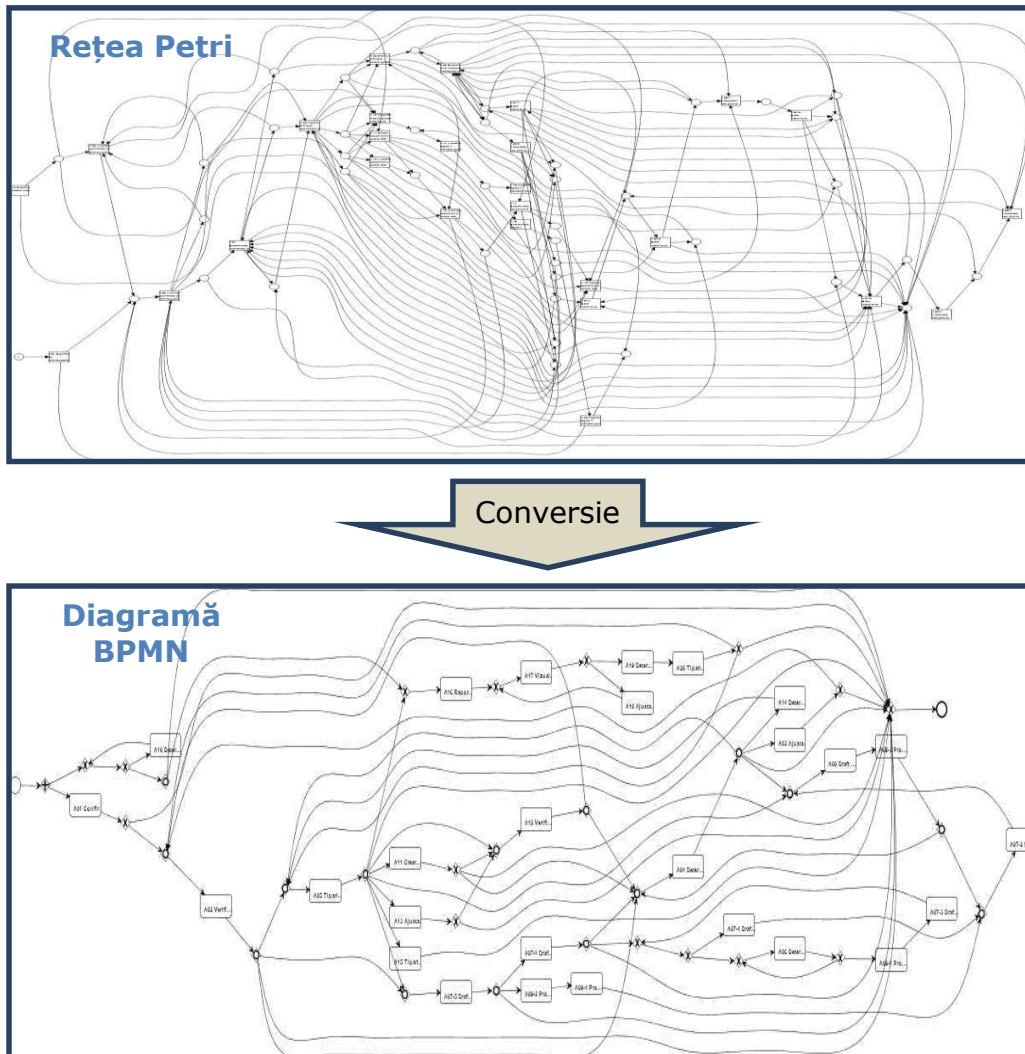


Fig. 6.15 Modele create din jurnalul inițial cu algoritmul α (rețea Petri și diagramă BPMN)

Se observă că modelul obținut din jurnalul inițial (numit în continuare Inițial-Alfa) este foarte complicat și el include cele 27 de activități și legăturile dintre ele. Celălalt model (numit în continuare *Filtrat-Alfa*) este prea simplu, conține doar cele mai frecvente 6 activități, dar ele sunt discontinue. Se poate observa că activitățile A02, A04, A06 și A10 se pot declanșa oricând, deci modelul are un comportament prea general. Modelul *Filtrat-Alfa* nu poate reprezenta o rețea flux de lucru, deci nu este pretabil pentru folosirea ulterioară.

Pentru jurnalul de evenimente studiat, algoritmul σ nu oferă soluții viabile. În descoperirea proceselor reale, algoritmul σ nu este recomandat. El a fost inclus în analiza curentă doar în scop demonstrativ.

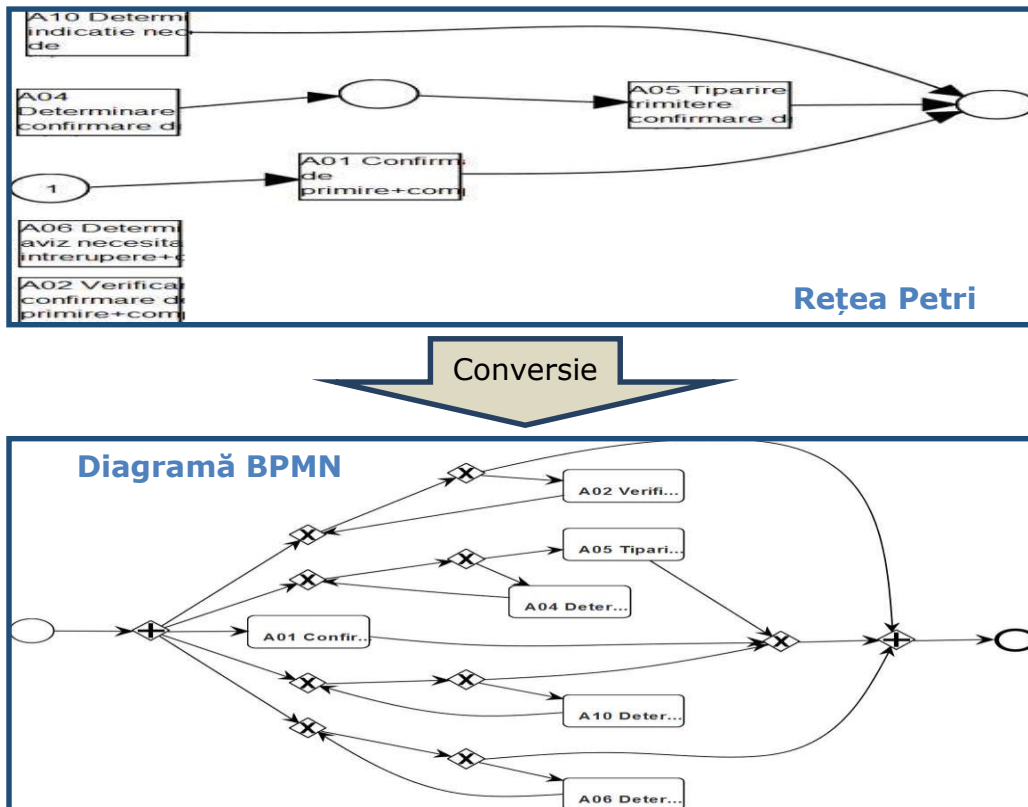


Fig. 6.16 Modele create din jurnalul filtrat cu algoritmul α (rețea Petri și diagramă BPMN)

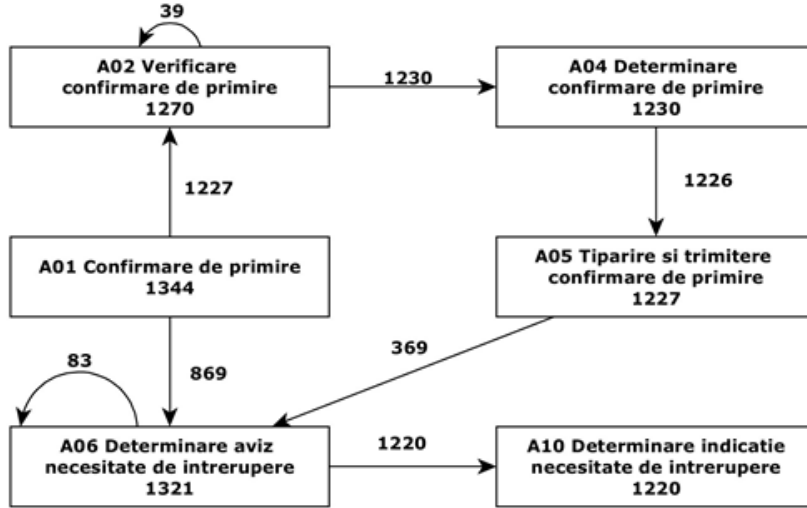
6.2.3.2. Algoritmul HeuristicsMiner

Algoritmul *HeuristicsMiner* (3.4.2.1) este implementat în platforma ProM prin plug-in-ul *Mine for a Heuristics Net using Heuristics Miner (MHNHM)* [164]. Rezultatul acestui algoritm este o rețea euristică. Această rețea va fi convertită într-o rețea Petri prin plug-in-ul *Convert Heuristics net into Petri net (CHNPN)*, iar apoi în format BPMN prin plug-in-ul *Convert Petri net to BPMN (CPNB)*.

Rețeaua euristică este strămoșul rețelei cauzale tip C-nets. În momentul de față, ProM nu are suport al *algoritmului pentru Mining euristic* pentru rețele cauzale. Dar rețeaua euristică este în mare măsură similară rețelelor cauzale tip C-nets, pentru că frecvențele activităților și a arcelor de legătură apar notate. Rețeaua euristică notată cu frecvențe se poate observa în Fig. 6.17.

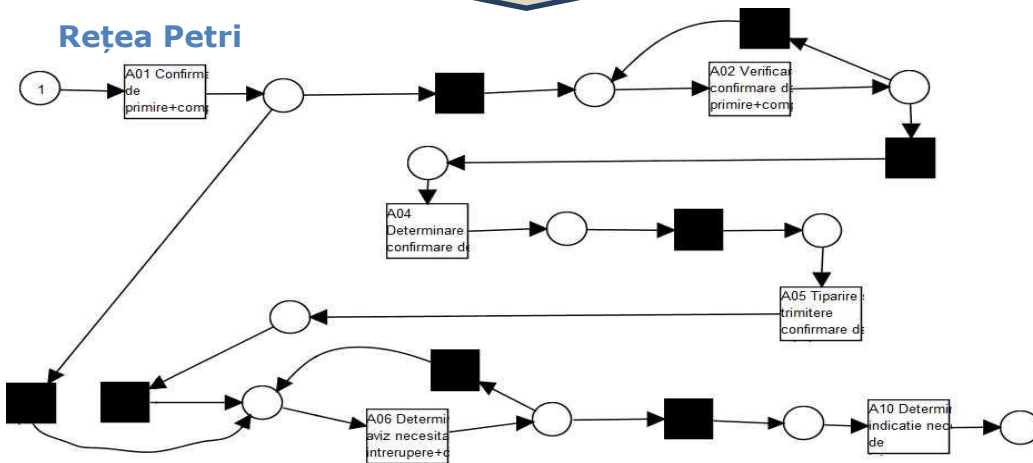
În Fig. 6.17 s-a reprezentat rezultatul BPMN obținut cu *algoritmul HeuristicsMiner* (cu conversiile aferente) aplicat asupra jurnalului de evenimente filtrat cu *FLuSH*. Diagrama BPMN este ușor de înțeles, este simplă și este un candidat la rezultatul final al modelului de proces. Am denumit *Filtrat-Euristic* modelul de proces reprezentat în rețea Petri și descoperit din jurnalul filtrat.

Rețea euristică



Conversie

Rețea Petri



Conversie

Diagramă BPMN

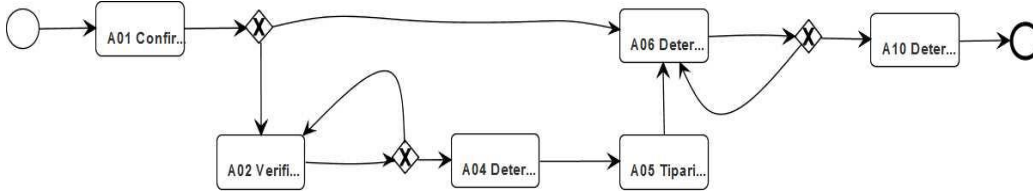


Fig. 6.17 Modele descoperite din jurnalul filtrat cu algoritmul HeuristicsMiner

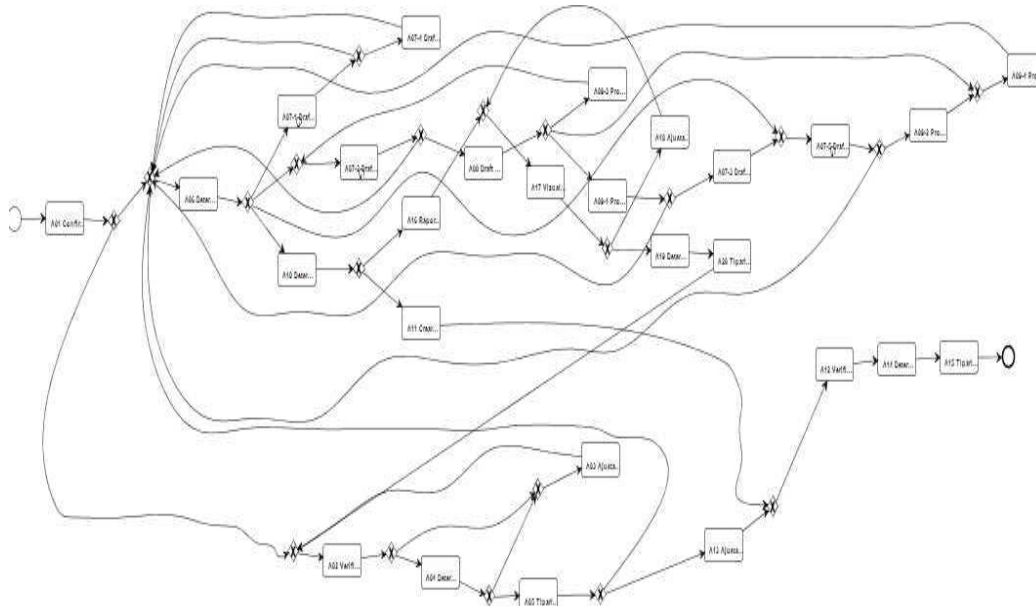


Fig. 6.18 Model descoperit din jurnalul inițial cu algoritmul HeuristicsMiner (ProM 5.2) – diagramă BPMN

Plug-in-ul *MHNHM* a fost aplicat și asupra jurnalului inițial și s-a obținut o rețea euristică, apoi o rețea Petri. Problema este că această rețea nu a putut fi convertită în ProM către o rețea BPMN, datorită complexității rețelei Petri. Dar s-a folosit o altă implementare mai veche a *algoritmului pentru Mining euristic*, anume plug-in-ul *Mine BPMN Using HeuristicsMiner of Prom5.2 (MBHMP5.2)*. Rezultatul obținut este prezentat în Fig. 6.18, este complex, dar utilizabil. De remarcat că plug-in-urile ProM obțin diagrame BPMN modelate doar cu portaluri exclusive. Am denumit *Inițial-Euristic* modelul de proces reprezentat ca rețea Petri și descoperit din jurnalul inițial.

6.2.3.3. Algoritmul Fuzzy Miner

Algoritmul Fuzzy Miner (3.4.2.2) este implementat în platforma ProM prin plug-in-ul *Mine for a Fuzzy Model (MFM)* [73]. Rezultatul acestui algoritm este un model fuzzy. Pentru analiza *algoritmului Fuzzy Miner* vom folosi programul *Disco* prezentat în 4.2.2. *Disco* folosește *algoritmul Fuzzy Miner* și are avantajul că este ușor de utilizat, spre deosebire de plug-in-ul *MFM* din ProM. *Algoritmul Fuzzy Miner* ține cont de frecvența de apariție a activităților și a arcelor. *Disco* prezintă două controale de tip *slider* prin care se pot ajusta frecvențele activităților și a arcelor. Pentru a vedea toate activitățile, valoarea folosită pentru frecvența activităților a fost 100%. Pentru a vedea doar arcele cu cele mai mari valori, valoarea folosită pentru frecvența arcelor a fost 0%.

Procesul descoperit din jurnalul inițial este prezentat în Fig. 6.19, iar Procesul descoperit din jurnalul filtrat cu *FLuSH* este prezentat în Fig. 6.20.

Dacă din jurnalul inițial se ajustează frecvența activităților la 0% și frecvența arcelor tot la 0%, atunci modelul de proces din jurnalul inițial coincide cu cel descoperit din jurnalul filtrat cu *FLuSH*.

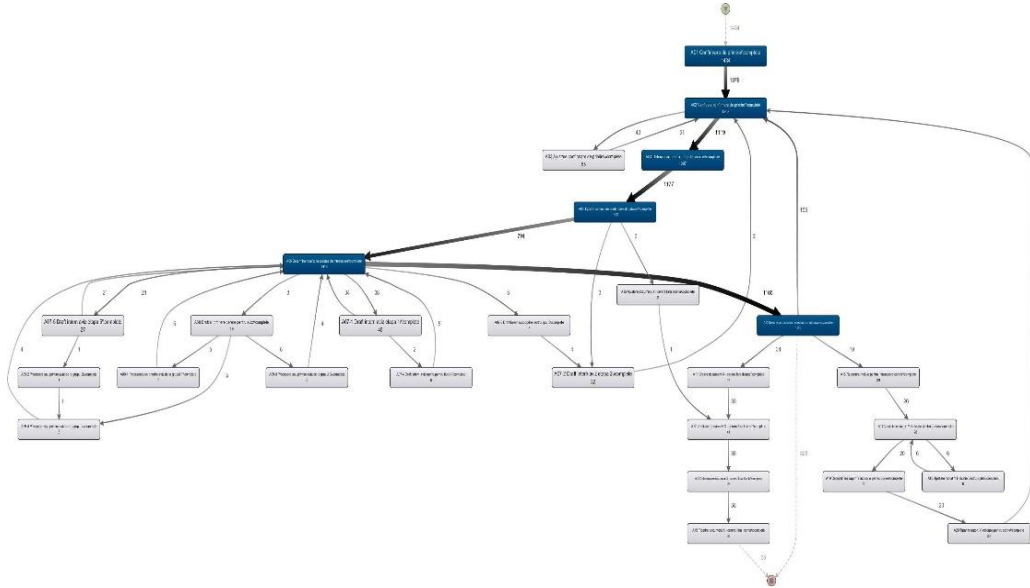


Fig. 6.19 Model descoperit din jurnalul inițial cu algoritmul Fuzzy Miner

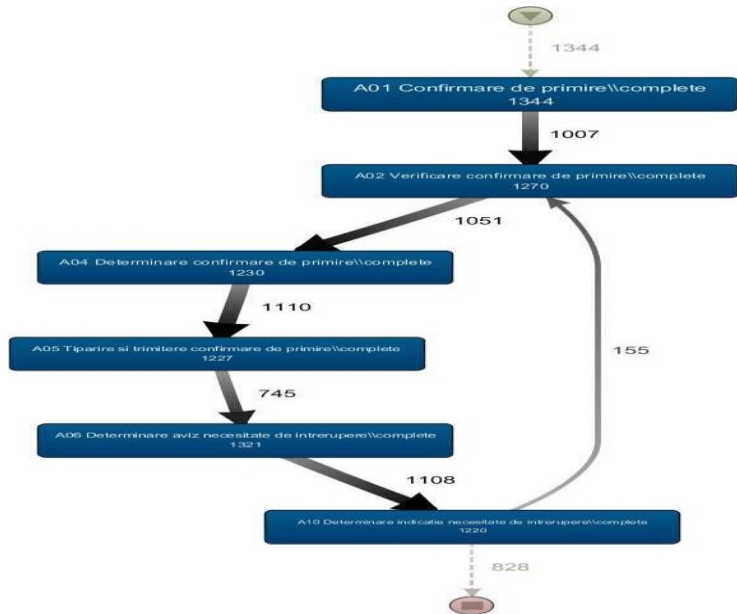


Fig. 6.20 Model Fuzzy descoperit din jurnalul inițial/filtrat cu cele mai frecvente activități și arce

Rezultatul algoritmului este un model fuzzy care nu se poate converti în rețea Petri [140]. Algoritmul Fuzzy se folosește atunci când se lucrează cu date complexe și nestructurate sau când se dorește o simplificare a modelului într-o manieră interactivă. În Disco se poate anima modelul cu evenimentele din jurnal pentru a observa dinamic comportamentul procesului.

6.2.3.4. Algoritmul Inductive Miner

Algoritmul Inductive Miner (3.4.2.5) este implementat în platforma ProM:

- prin plug-in-ul *Mine Petri Net with Inductive Miner (MPNIM)*. Rezultatul acestui algoritm este o rețea Petri.
- prin plug-in-ul *Mine Process Tree with Inductive Miner (MPTIM)*. Rezultatul acestui algoritm este un arbore de proces.
- plug-in-ul *Mine with Inductive Visual Miner (MIVM)*. Acest plug-in permite operații de filtrare folosind frecvența activităților și a arcelor, iar modelul ajustat poate fi exportat ca rețea Petri sau ca arbore de proces. Sunt permise și filtre asupra atributelor jurnalului. Peste modelul obținut se poate realiza o animație prin rularea cazurilor.

Modelul de proces dezvoltat din jurnalul inițial este descris în Fig. 6.21 și conversia lui în rețea Petri am denumit-o *Inițial-Inductiv*. Nivelul de prag pentru frecvența activităților este 1 (100%), iar pentru frecvența arcelor este 0,8 (80%).

În Fig. 6.22 s-a prezentat modelul de proces descoperit din jurnalul filtrat cu *FLuSH*, cu aceleași valori ale frecvenței ca pentru jurnalul inițial. S-a ales să se prezinte și deviațiile de la proces, reprezentate prin linie roșie punctată. În model, frecvențele sunt reprezentate pe activități, pe arce și pe arce deviate. Acest model a fost convertit în rețea Petri și în arbore de proces, care se pot vedea în aceeași figură. Conversia acestui model în rețea Petri am denumit-o *Filtrat-Inductiv*.

Algoritmul din spatele *MIVM* este foarte rapid și produce modele structurate ce pot fi îmbunătățite sau analizate cu alte plug-in-uri. Modelele rezultate pot fi îmbunătățite cu plug-in-ul *Replay a Log on Petri Net for Performance/Conformance Analysis (RLPNPC)* [12], ce adaugă date despre timp sau cu plug-in-ul *Perform Predictions of Business Process Features (PPBPF)* [105], ce poate adăuga alte atribute în model.

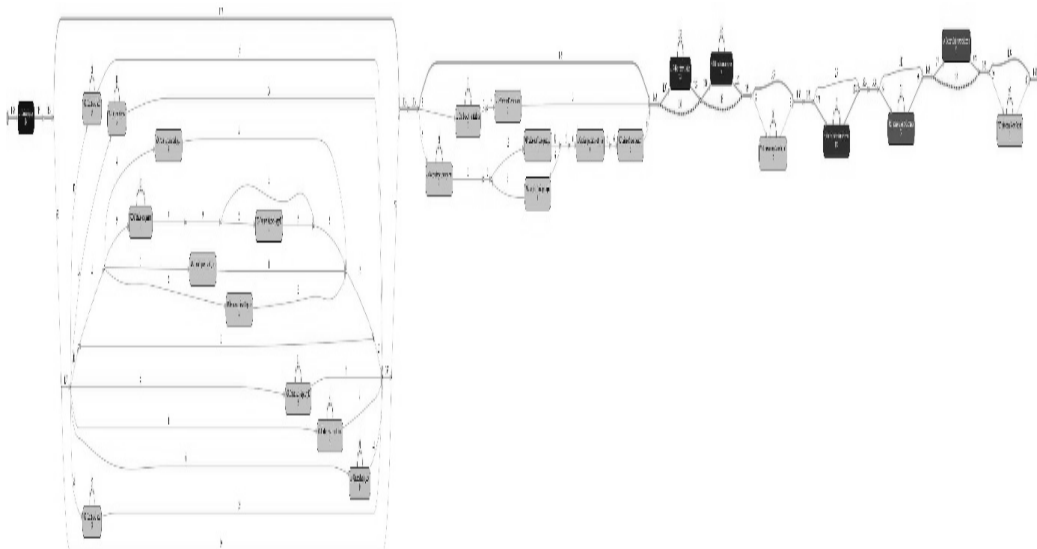


Fig. 6.21 Model creat din jurnalul inițial cu algoritmul Inductive Miner

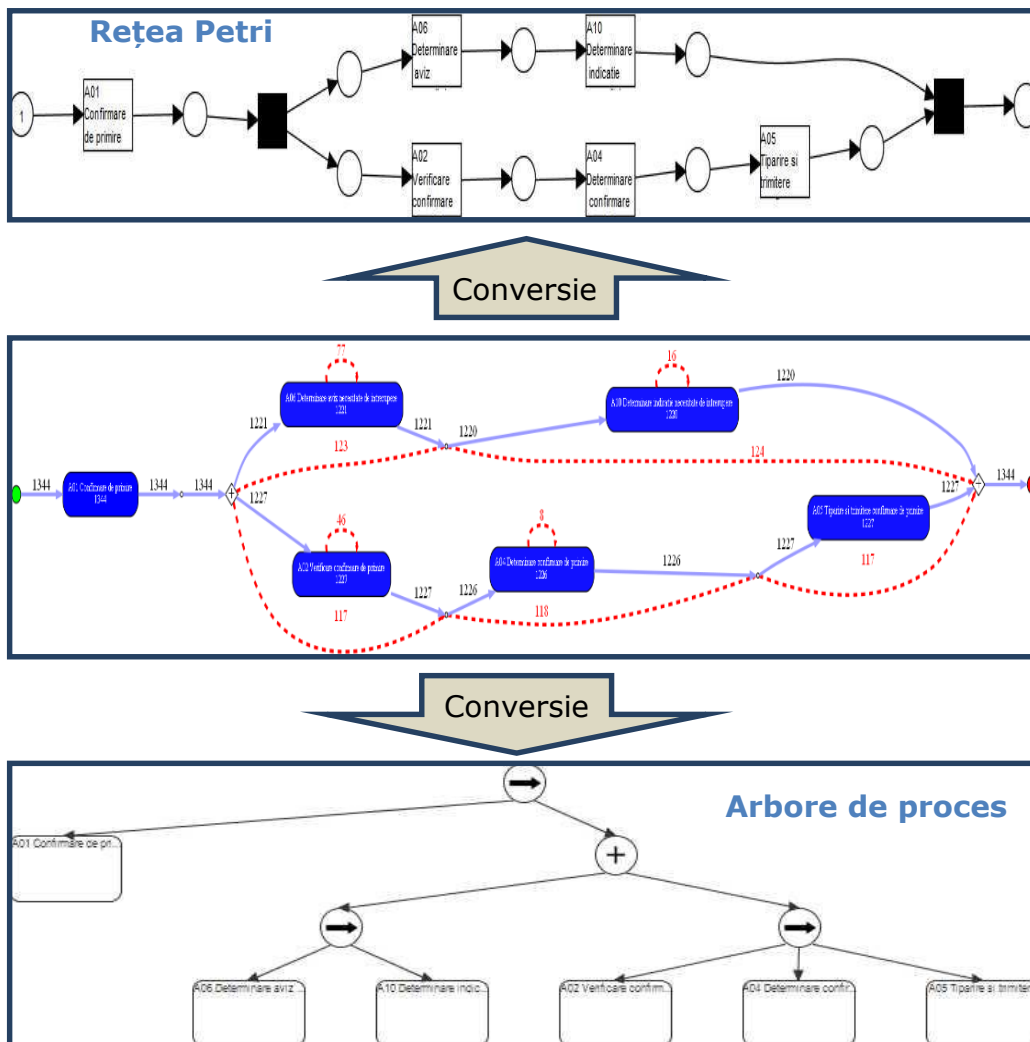


Fig. 6.22 Model creat din jurnalul filtrat cu algoritmul Inductive Miner și conversiile lui

6.2.4. Compararea și alegerea modelului de proces

Pentru a compara modelele obținute, se recurge la verificarea conformității prin calcularea potrivirii folosind tehnica reluării jurnalului peste o rețea Petri (3.5.2). S-a folosit plug-in-ul *Replay a Log on Petri Net for Conformance Analysis (RLPNCA)* [12]. Acest plug-in are ca intrare un jurnal de evenimente și modelul obținut în rețea Petri. În Tabelul 6.3 sunt prezentate valorile potrivirii, inclusiv ale modelului de proces documentat în manualul de proceduri. Pentru ca acest model de proces să fie folosit, prima dată s-a modelat în limbaj BPMN, iar apoi a fost convertit în Rețea Petri în ProM cu ajutorul plug-in-ului *Convert BPMN diagram to Petri net – Control Flow (CBDPNFC)*.

Tabelul 6.3. Compararea modelelor obținute cu ajutorul coeficientului de potrivire

Jurnal	Model	Potrivirea
Jurnal inițial	<i>Inițial-Alfa</i>	0,9191
	<i>Inițial-Euristic</i>	0,9191
	<i>Inițial-Fuzzy</i>	N/A
	<i>Inițial-Inductiv</i>	0,8494
	<i>Model normativ (din manualul de proceduri)</i>	0,8425
	Filtrat-Alfa	0,5986
	<i>Filtrat-Euristic</i>	0,8302
	<i>Filtrat-Fuzzy</i>	N/A
	Filtrat-Inductiv	0,911
Jurnal filtrat cu FLuSH	<i>Filtrat-Alfa</i>	0,8561
	<i>Filtrat-Euristic</i>	0,8503
	<i>Filtrat-Fuzzy</i>	N/A
	Filtrat-Inductiv	0,9295
	<i>Model normativ (din manualul de proceduri)</i>	0,8562

După cum se poate observa din Tabelul 6.3, cel mai bun model este cel obținut prin algoritmul inductiv (*Filtrat-Inductiv*), după ce jurnalul de evenimente a fost filtrat. Modelul este prezentat în Fig. 6.22. Potrivirea acestui model are valori mari, atât în comparație cu jurnalul inițial, cât și comparativ cu jurnalul filtrat cu *FLuSH*. Acest model este o rețea flux de lucru pentru că respectă proprietățile: siguranță, finalizare corespunzătoare, opțiunea de a finaliza și lipsa părților moarte. Fiindcă poate fi reprezentat ca rețea flux de lucru reprezintă alt avantaj pentru modelul obținut cu algoritmul Inductiv.

Filtrarea a avut un rol important asupra obținerii modelului, pentru că am reușit să creștem potrivirea de la 0,8494 la 0,9295.

Cea mai proastă variantă este algoritmul σ , care are valori foarte mici (0,5986) atunci când este folosit cu un jurnal filtrat. Explicația pentru valoarea mică a potrivirii este faptul că algoritmul σ nu ține cont de frecvența activităților. De asemenea, algoritmul σ nu ține cont de tranzițiile discrete (reprezentate de tranzițiile cu fundal negru). Dar prin faptul că s-a obținut acest rezultat, se adeverește că folosirea acestor tranziții discrete ajută foarte mult.

În Fig. 6.23 sunt prezentate devierile de la modelul creat cu *algoritmul Inductive Miner*. De exemplu, activitatea A02 s-a executat în 1227 cazuri din cele 1344. În urma reluării jurnalului peste modelul *Filtrat-Inductiv*, modelul a trebuit aliniat de 117 ori, făcându-se deplasări doar în model.

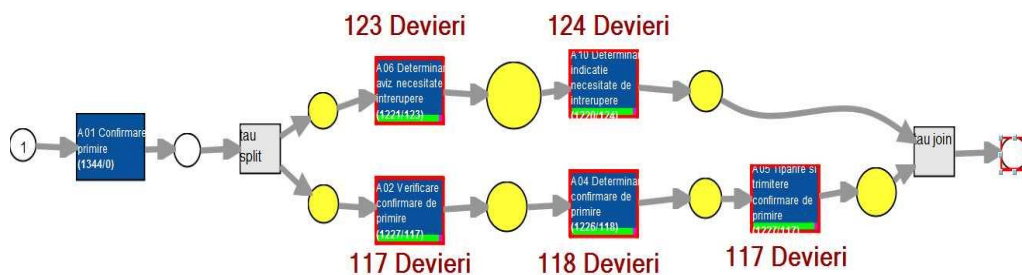


Fig. 6.23 Devieri de la modelul creat cu algoritmul Inductive Miner

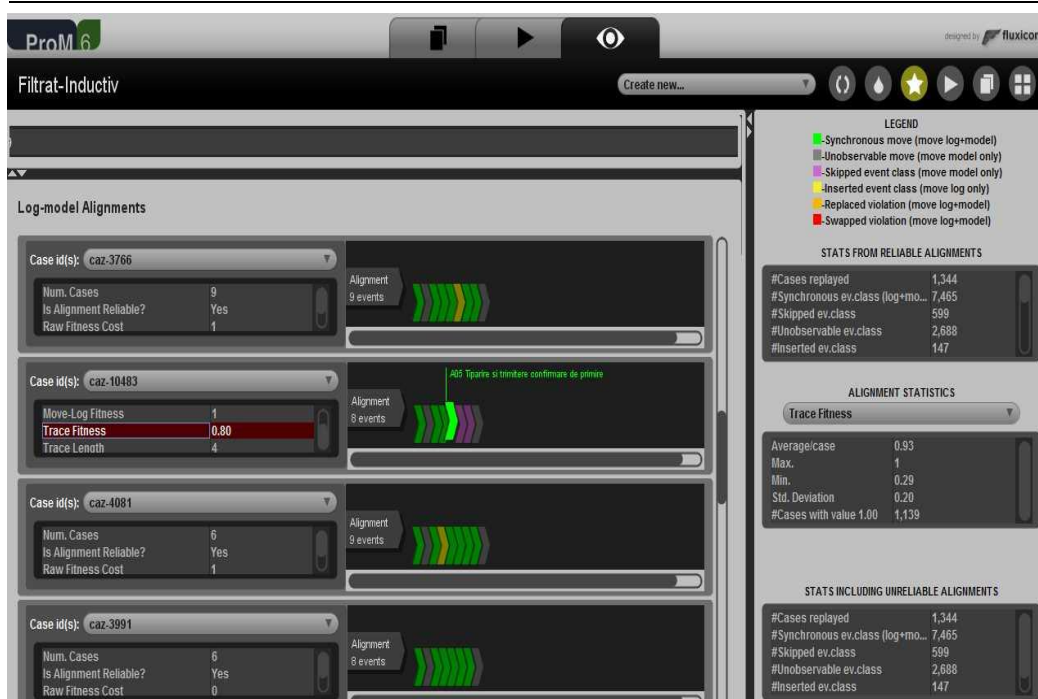


Fig. 6.24 Analiza variantelor de proces în urma verificării conformității

Pe acest model se poate observa și potrivirea la nivel de caz folosind opțiunea *Project alignment to log* din RLPNCA (Fig. 6.24). De exemplu, pentru cazul 10483 avem o potrivire de 0,8 și au fost rulate 7 cazuri similare. RLPNCA permite observarea alinierilor și afișează diferite metrice de conformitate.

6.2.5. Analiza performanței

Pentru analiza performanței este recomandată folosirea programului Disco. Disco se poate configura ca să prezinte modelul în funcție de metrice de performanță pentru activități și pentru arce, precum: durata totală, durata mediană, durata medie, durata minimă sau durata maximă. Durata mediană se referă la elementul din mijlocul unui șir ordonat, dacă șirul are număr impar de elemente sau la media aritmetică a celor două elemente din mijlocul unui șir ordonat, dacă șirul are număr par de elemente. Modelul fuzzy în Disco se poate vizualiza atât în termeni de frecvență, cât și în termeni de performanță.

Durata activităților nu o cunoaștem, deoarece în jurnalul inițial un eveniment conține doar data și timpul de completare a activității, iar nu și data de începere. Pentru aceasta, este posibilă doar analiza duratelor dintre activități, duratele fiind prezente pe arce. Ne propunem să găsim acele activități care au durata medie cea mai mare dintre activități.

Disco a fost configurat astfel: *Activities* = 100% (frecvența activităților), *Paths* = 100% (frecvența arcelor), *Performance* = *Mean duration* (afișare în funcție de performanță, durata medie), *Secondary* = *Absolute frequency* (afișare număr total cazuri). S-au exportat metricile procesului în format CSV. Fișierul exportat *Relations*

- *Mean Duration.csv* conține o matrice cu duratele medii dintre activități, iar *Relations - Absolute Frequency.csv* conține o matrice cu frecvențele dintre activități. Cele 2 fișiere au fost importate în Excel, iar apoi extrase acele arce dintre activități ce durează în medie mai mult de 3 zile și au o frecvență de peste 30 de cazuri. Arcele au fost evidențiate în Tabelul 6.4:

Același rezultat l-am obținut în Disco prin configurația: *Activities = 0%* , *Paths = 0%*), *Performance = Mean duration* , *Secondary = Case frequency* . S-a obținut procesul din Fig. 6.25. Amintim că jurnalul are 1434 de cazuri, iar durata medie de execuție este 5,4 zile pentru un caz. Se observă două gâturi (*bottlenecks*) ale procesului și anume:

- Între activitățile *A05* și *A06* durata medie este de 3,1 zile (57% din durata unui caz se petrece între aceste două activități!) pentru 791 cazuri (55% din totalul cazurilor).
- Între activitățile *A10* și *A02* durata medie este de 3,1 zile (57% din durata unui caz se petrece între aceste două activități!) pentru 155 cazuri (11% din totalul cazurilor).

Restul de activități au durate normale de execuție și nu prezintă interes. Ca obiectiv de cercetare s-a propus identificarea cauzelor acestor gâturi.

Tabelul 6.4. Durata medie și frecvența dintre activități pentru identificarea gâturilor

Nr.	Între activitățile	Durata medie	Frecvența
1	A02...→A03...	5 zile	43 cazuri
2	A03...→A06...	56 zile	2 cazuri
3	A04...→A03...	12,8 zile	6 cazuri
4	A05...→A13...	8,6 zile	2 cazuri
5	A05...→A07-2...	13 zile	3 cazuri
6	A05...→A06...	3,1 zile	791 cazuri
7	A05...→A11...	3,3 zile	10 cazuri
8	A05...→A07-5...	8,4 zile	3 cazuri
9	A06...→A03...	3,6 zile	2 cazuri
10	A07-1...→A07-2...	10,5 zile	2 cazuri
11	A08...→A09-3...	4,5 zile	8 cazuri
12	A09-1...→A07-3...	3,7 zile	2 cazuri
13	A10...→A11...	4,1 zile	34 cazuri
14	A14...→A15...	3 zile	38 cazuri
15	A20...→A02...	3,4 zile	3 cazuri
16	A10...→A02...	3,1 zile	155 cazuri

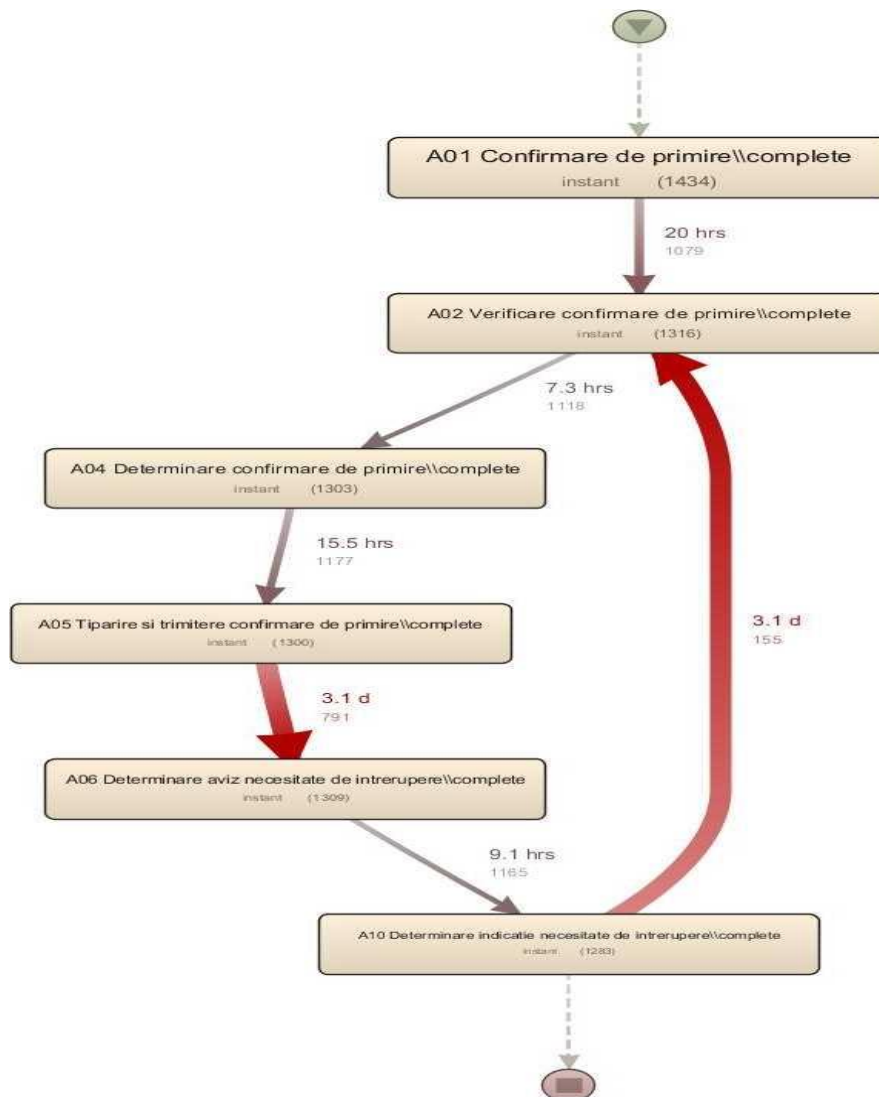


Fig. 6.25 Model fuzzy reprezentat cu frecvențe și durate medii între activități

Tot în Disco putem analiza distribuția evenimentelor în timp și distribuția cazurilor active în timp Fig. 6.26. Se poate vedea vizual când se execută cele mai multe/puține evenimente, care sunt perioadele de vârf, care sunt perioadele moarte. Din jurnalul nostru se poate observa că în 1 mai 2012 există un maxim de 146,39 evenimente pe zi. Având în vedere că 1 mai este zi liberă, apare întrebarea de ce s-a lucrat în acea zi și de ce a fost acest maxim. Se poate observa că există alternări ale perioadelor de ocupare cu perioadele în care nu se execută nici o activitate. Acest tipar arată că în weekend-uri nu se lucrează. Perioadele cu cele mai multe cazuri active (cea mai ocupată) sunt Decembrie 2011 – Ianuarie 2012 (marcată cu 1 în Fig. 6.26) și Aprilie - Mai 2012 (marcată cu 2 în Fig. 6.26). Perioada cu cele mai puține cazuri active este August - Octombrie 2012 (marcată cu 3 în Fig. 6.26).

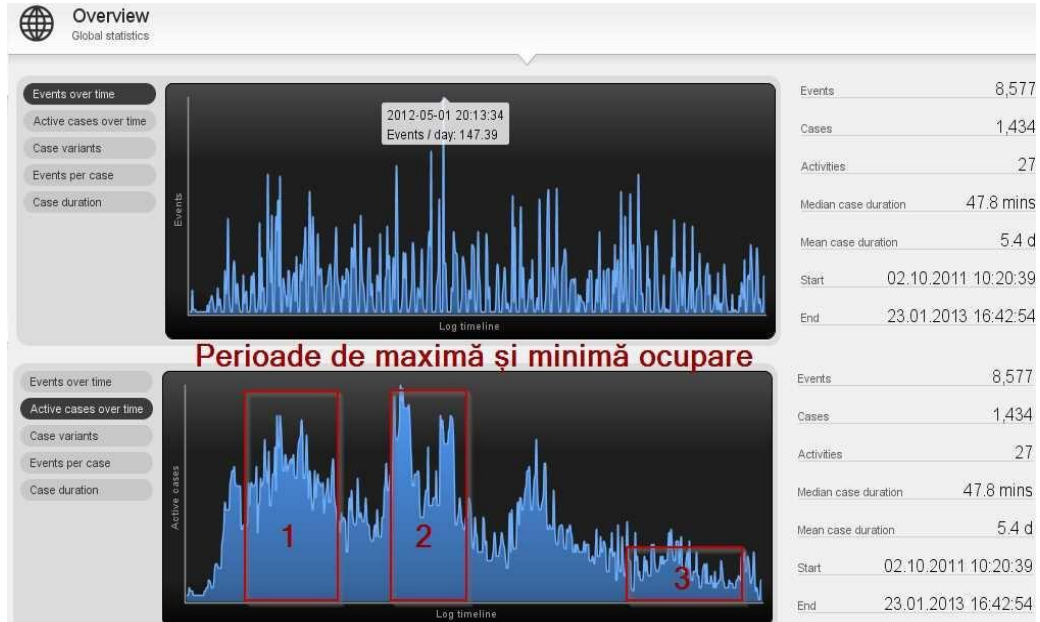


Fig. 6.26 Frecvența în timp a evenimentelor și a cazurilor active din jurnalul inițial

Se pot analiza și variante ale cazurilor (Fig. 6.27). Există 116 variante de cazuri. Primele 5 variante cuprind cumulativ 80% din cazuri. Prima variantă cuprinde 713 cazuri (50% din cazuri), are o durată medie de 4 zile și 8 ore și conține următoarele 6 evenimente în ordinea din Tabelul 6.5.

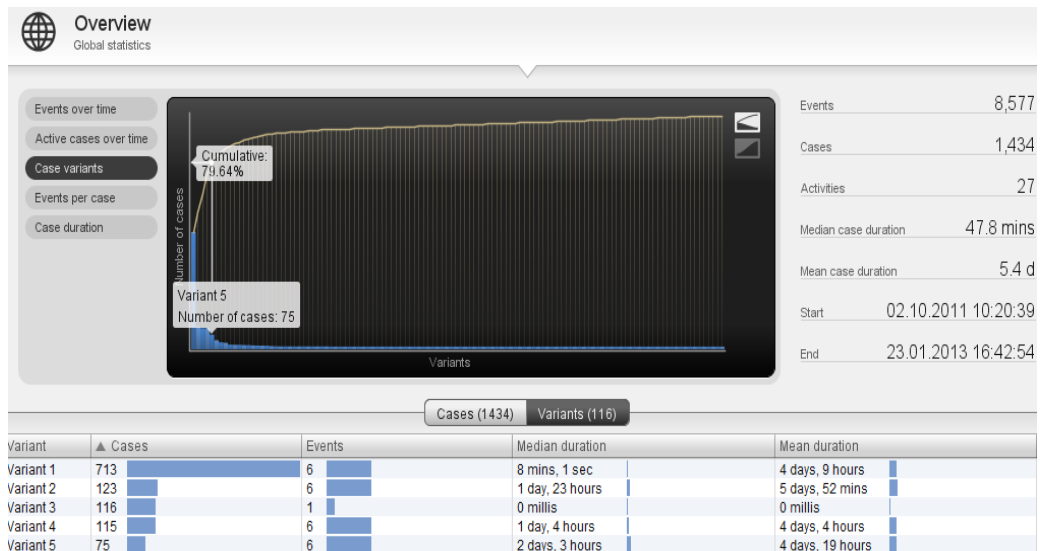


Fig. 6.27 Cele mai frecvente variante de cazuri

Tabelul 6.5. Varianta de caz cea mai frecventă

Nr.	Activități
1	A01 Confirmare de primire
2	A02 Verificare confirmare de primire
3	A04 Determinare confirmare de primire
4	A05 Tiparire si trimitere confirmare de primire
5	A06 Determinare aviz necesitate de intrerupere
6	A10 Determinare indicatie necesitate de intrerupere

Putem concluda că pentru orice caz început există o probabilitate de 0,5 ca el să se desfășoare conform variantei de mai sus și să dureze aproximativ 4 zile și 8 ore. Varianta 102 are 1 caz cu 8 evenimente și are durata cea mai mare și anume 97 zile și 18 ore. Pentru această variantă, cele mai mari întârzieri se datorează activităților A07-1 și A07-2.

6.2.6. Analiza organizațională

În cadrul ProM există mai multe plug-in-uri pentru Mining-ul rețelelor sociale și pentru Mining-ul organizațional. Toate aceste plug-in-uri creează rețele sociale în care elementele rețelei sunt resursele care execută activitățile. Prin analiza rețelelor sociale obținute se pot descoperi diferite tipare de lucru dintre resurse. Jurnalul a fost supus anonimizării resurselor, iar membrii organizației poartă nume simbolice, gen *Resursa02*. Plug-in-urile folosite în analiză [174] sunt:

- Plug-in-ul *Mine for a Handover-of-Work Social Network (MHWSN)* creează o rețea socială privind transferul lucrului între două resurse succesive. În Fig. 6.28 se poate vedea care sunt resursele centrale (care au cele mai multe conexiuni) și care sunt resursele izolate. În figură, resursele sunt reprezentate de poligoane: cu cât o resursă are mai multe conexiuni, cu atât ea are mai multe laturi (de exemplu, un triunghi reprezintă o resursă izolată, iar un poligon asemănător unui cerc reprezintă o resursă centrală). Sensul unui arc indică sensul transferului de lucru. Resursele centrale identificate sunt: *Resursa10*, *admin1*, *admin2*, *Resursa24* și *Resursa02*. Resursele izolate sunt în număr mare, printre care remarcăm *Resursa42* sau *Resursa31*.
- Plug-in-ul *Mine for a Subcontracting Social Network (MSSN)* creează o rețea socială privind subcontractarea lucrului. Subcontractarea este definită ca frecvența cu care o activitate executată de o resursă A se află între 2 activități executate de o resursă B. Rezultatul acestui plug-in este reprezentat în Fig. 6.29. Resursele centrale sunt *Resursa10*, *Admin1* și *Resursa24*. Se observă că sensurile arcelor sunt către aceste noduri, aceasta însemnând că lucrul este subcontractat acestor resurse. De aici, putem trage concluzia că fluxul de lucru este dirijat de la resursa inițiatoare către resursa centrală, iar apoi înapoi către resursa inițiatoare. Acest tipar se poate traduce ca un tipar de aprobare, adică resursa inițiatoare cere aprobare resursei centrale.
- Plug-in-ul *Mine for a Working-Together Social Network (MWTSN)* creează o rețea socială privind frecvența cu care două resurse lucrează împreună la activități din același caz. Dacă resursele lucrează împreună la cazuri, ele vor avea o relație mai puternică decât resursele care lucrează rar împreună. Rezultatul acestui plug-in este reprezentat în Fig. 6.30. Resursele centrale sunt: *Admin1*, *Resursa10* și *Admin2*. Aceste resurse au sensurile arcelor către ele, de unde rezultă că există

multe resurse care lucrează împreună cu acestea. Importanța resurselor este dată de mărimea nodurilor și de numărul de laturi ale poligoanelor.

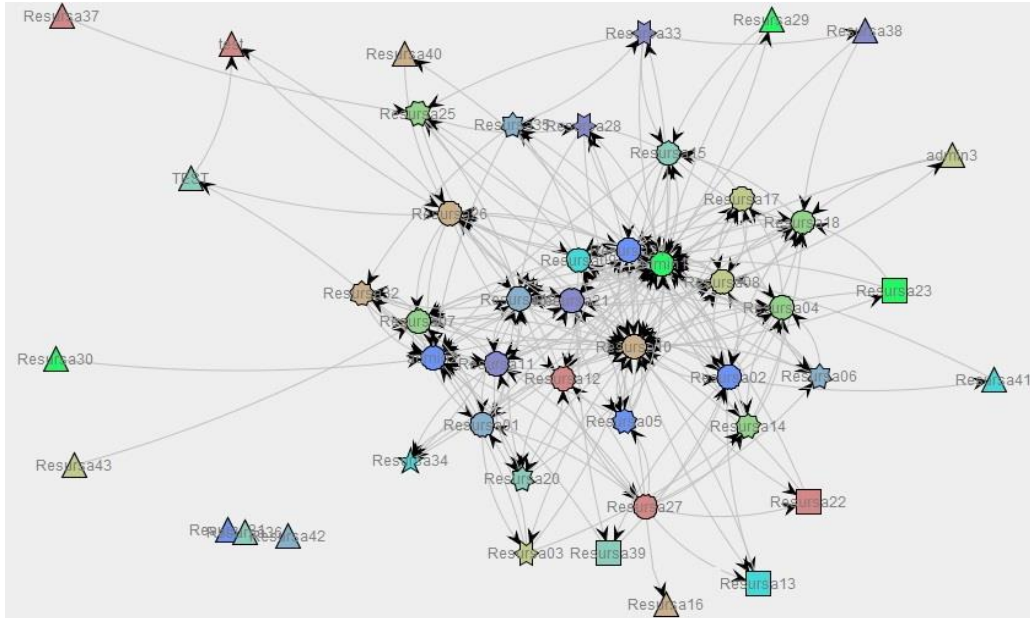


Fig. 6.28 Rețea socială reprezentând transferul de lucru între resurse

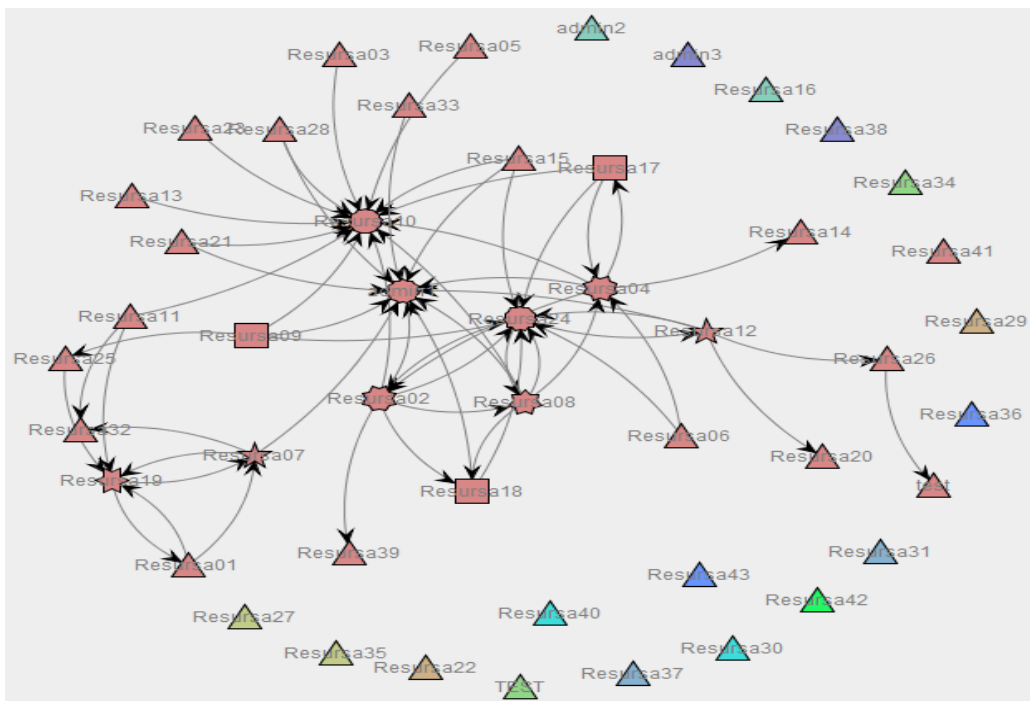


Fig. 6.29 Rețea socială reprezentând subcontractarea lucrului între resurse

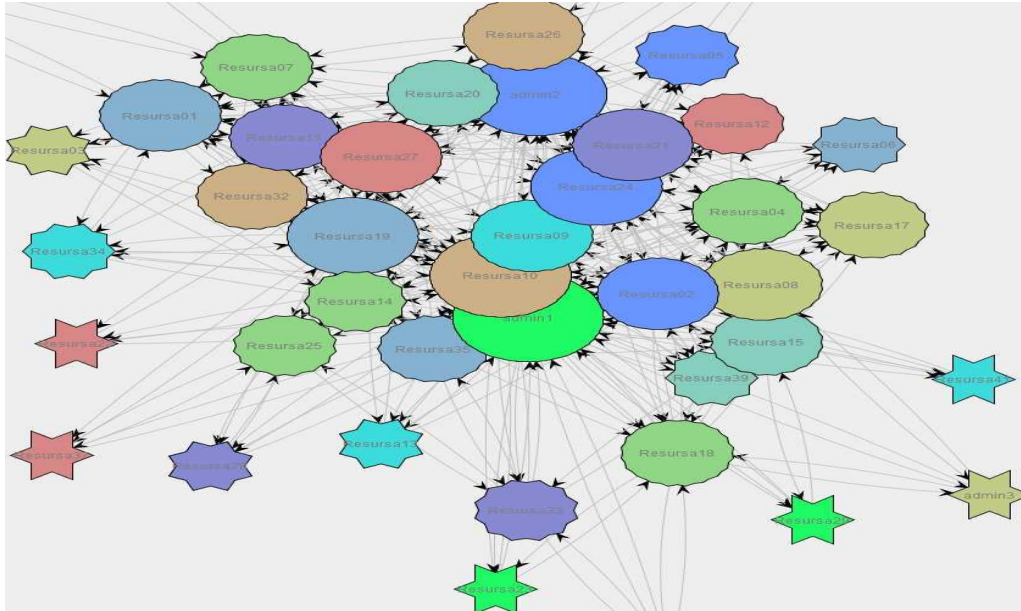


Fig. 6.30 Rețea socială reprezentând resursele ce lucrează împreună la aceleași cazuri

- Plug-in-ul *Mine for a Similar-Task Social Network (MSTSN)* creează o rețea socială privind frecvența cu care resursele execută activități similare. Ipoteza este că resursele care fac lucruri similare au relații mai puternice decât resursele care fac lucruri complet diferite. Fiecare resursă are un „profil”, bazat pe modul în care aceasta desfășoară frecvent activități specifice. Există multe modalități de a măsura „distanța” dintre două profile, permițând astfel mai multe metrice, precum: distanța euclidiană, distanța Hamming, coeficientul de corelare Pearson, coeficientul de similaritate.
- Plug-in-ul *Mine for a Reassignment Social Network (MRSN)* creează o rețea socială privind frecvența cu care resursele sunt realocate pentru a executa o anumită activitate. Cu acest plug-in se pot descoperi relații ierarhice: dacă lucrul este transferat frecvent de la o resursă A la o resursă B, iar nu invers, probabil că A este într-o relație ierarhică cu B.

Plug-in-urile *MSTSN* și *MRSN* nu prezintă rezultate interpretabile din jurnalul inițial, de aceea nu au fost prezentate capturi ecran din ProM.

6.2.7. Raportul final al analizei

În cadrul acestei etape se realizează concluziile analizei prin raportarea rezultatelor la obiectivele de cercetare propuse (stabilite). În această etapă, toate datele care au fost strânse anterior se centralizează, rezultând un raport de diagnosticare a procesului și eventual unele recomandări de îmbunătățire. Acest raport va fi prezentat reprezentanților organizației ale cărei procese au fost studiate.

Am precizat că *principala problemă* a acestui proces o constituie tranziția din A05 în A06. Cu ajutorul programului Disco și apoi cu ajutorul programului Excel, s-a filtrat jurnalul inițial, astfel încât să conțină toate cazurile ce au durată de execuție

dintre activitățile A05 în A06 mai mult de 3 zile. S-au obținut astfel 126 de cazuri (adică aproximativ 9% din totalul de cazuri). Din Tabelul 6.6 se observă că activitatea A05 este executată preponderent de resursa *admin1*, ceea ce înseamnă că este o activitate ce necesită privilegii. Acest lucru este întărit și de faptul că resursa *admin1* nu apare în activitatea următoare, A06. În jurnalul inițial, activitatea A05 este executată de 1300 de ori, iar resursa *admin1* o execută de 332 ori (adică de peste 25%). Resursa *admin1* execută în total 352 de activități din jurnalul inițial, deci activitatea A05 este executată în proporție de 94,32% din totalul de activități ale resursei *admin1*. Putem concluziona că resursa *admin1* este o resursă specializată. De asemenea, resursa *admin1* execută activitatea A15 de 12 ori adică 3,41%. Atât A05 cât și A15 sunt activități de tipărire, deci resursa *admin1* este specializată pe tipărire. Importanța resursei *admin1* este întărită și de analiza rezultată cu plug-in-ul *MWTSN* (Fig. 6.30), în care se observă că *admin1* este o resursă centrală ce lucrează împreună cu alte resurse pentru îndeplinirea unui caz.

Ca soluție, ar trebui să existe mai multe resurse care să poată executa activitatea A05 și în absența resursei *admin1* sau ar trebui achiziționate mai multe echipamente pentru a grăbi tipărirea.

Alt tipar ce trebuie analizat din cazurile cu durată mare între activitățile A05 și A06 (Tabelul 6.6) este apariția frecventă a utilizatorului *Resursa01* în execuția acestor activități. *Resursa01* este cea mai ocupată dintre toate cele 48 de resurse. *Resursa01* execută 1228 activități, adică 14,32% din totalul de activități, deci reprezintă resursa cea mai ocupată. Există probabilitatea ca din cauza faptului că *Resursa01* este cea mai ocupată, aceasta să influențeze direct performanța cazurilor. Ca o recomandare, activitățile acestei resurse ar trebui delegate și către alte resurse, pentru a evita blocajele și pentru a obține un timp de răspuns îmbunătățit.

Altă problemă o reprezintă tranziția din A10 în A02. Pentru aceasta s-a filtrat jurnalul inițial în Disco astfel: activitatea A10 este direct urmată de A02, durata cazului este mai mare de două zile, iar durata dintre activitățile A10 și A02 să fie mai mare de o zi. S-a obținut o altă variantă de proces (Fig. 6.31), diferită de cea din (Fig. 6.25), în care activitatea A06 urmează după A01. Procesul filtrat are 48 de cazuri.

În Tabelul 6.7 este reprezentat cine execută activitățile din A10, iar apoi cine execută activitățile din A02. Se poate observa că A10 este executată de anumite resurse (majoritar de *Resursa11* și *Resursa09*), iar A02 este executată de alte resurse (*Resursa24* și *Resursa19*). Interesant este faptul că *Resursa24* și *Resursa19* nu execută niciodată activitatea A10. Putem concluda că transferul de lucru între activitățile A10 și A02 întâmpină probleme, atunci când implică resurse diferite. *Resursa19* și *Resursa24* fac parte din grupul *NEDEFINIT*, deci putem concluziona că atunci când se transferă lucrul către acest grup avem de a face cu blocaje de proces.

Tabelul 6.6. Resursele ce execută secvența de activități A05→A06

Activitatea A05		Activitatea A06	
Resursă	Număr cazuri (total 126)	Resursă	Număr cazuri (total 126)
<i>admin1</i>	50	<i>Resursa01</i>	32
<i>Resursa01</i>	44	<i>Resursa08</i>	12
<i>Alte resurse</i>	32	<i>Alte resurse</i>	82

Tabelul 6.7. Resursele ce execută secvența de activități A10→A02

Activitatea A10		Activitatea A02	
Resursă	Număr cazuri (total 48)	Resursă	Număr cazuri (total 48)
Resursa11	12	Resursa24	13
Resursa09	7	Resursa19	11
Alte resurse	29	Alte resurse	24

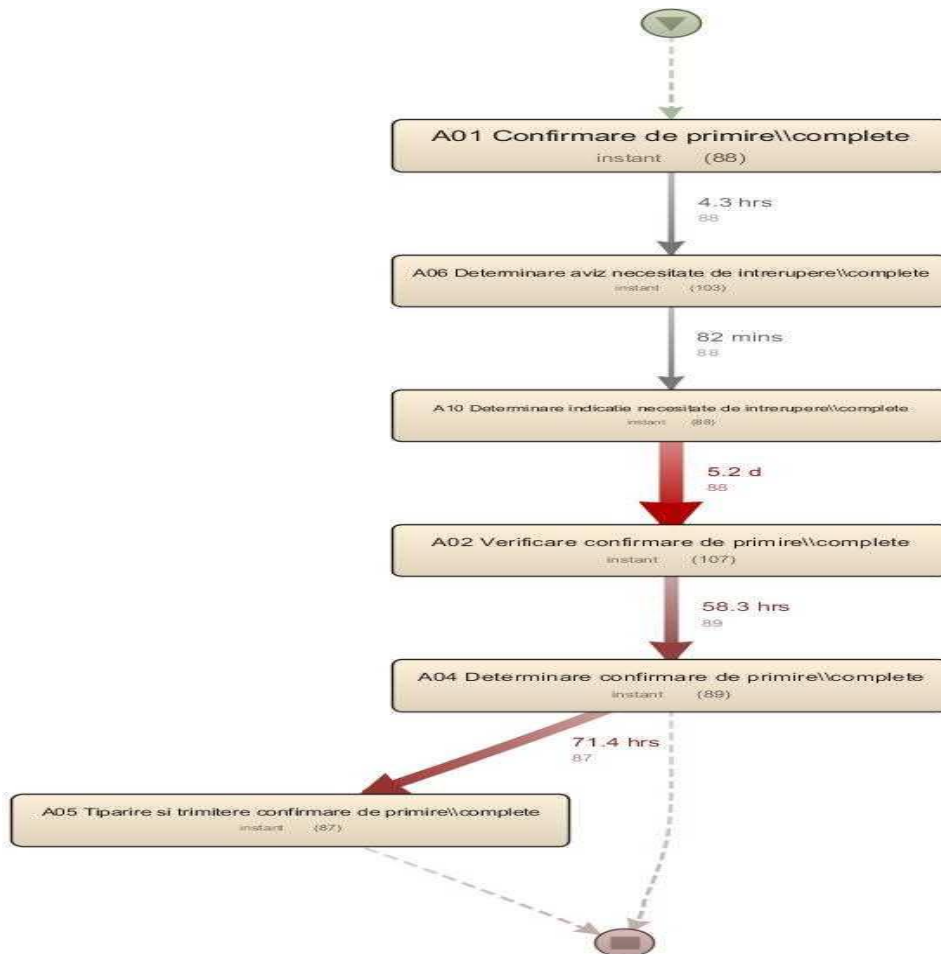


Fig. 6.31 Modelul fuzzy filtrat ce conține secvența A10→A02

Ca recomandare, ar trebui ca aceeași resursă ce execută A10 să execute A02. Studiindu-se jurnalul s-a observat că există resurse generaliste care pot executa ambele activități (de exemplu *Resursa01* sau *Resursa02*). Altă recomandare este evitarea transferului de lucru către grupul *NEDEFINIT*, atunci când trebuie executată activitatea A02.

Celălalt obiectiv al cercetării a fost analiza comparativă a procesului din punct de vedere al modului de inițializare. S-a dorit comparația procesului atunci când aplicația este primită pe Internet sau este primită la ghișeu. În Disco, s-au filtrat acele

cazuri a căror aplicație este via Internet și s-au identificat 1250 cazuri (87%). Cazurile a căror aplicație este la ghișeu sunt în număr de 109 (8%). În Tabelul 6.8 sunt prezentate diferențele de bază dintre cele două cazuri. În Fig. 6.32 se observă diferențele dintre modelele de proces obținute pentru aplicațiile recepționate pe Internet, respectiv la ghișeu. Amândouă modelele de proces au fost configurate în Disco cu *Activities = 0%* și *Paths = 0%* pentru a vedea modelul de bază. Se observă că în modelul de bază arcul dintre *A10* și *A02* pentru modelul *Ghișeu* lipsește, și el este înlocuit de alt arc dintre activitățile *A10* și *A04*. Nu înseamnă că arcul *A10-A02* nu există, ci doar că acesta nu este cel mai frecvent, iar în cazul modelului *Ghișeu*, acesta nu este prioritar arcului *A10-A04*.

Tabelul 6.8. Compararea cazurilor a căror solicitare este via Internet cu cele de la Ghișeu

Element analizat	Internet	Ghișeu
Număr cazuri	1250	109
Durata medie	5,2 zile	5,9 zile
Durata dintre activitățile <i>A05-A06</i>	2,8 zile	4,1 zile
Durata dintre activitățile <i>A10-A02</i>	3,1 zile	2,3 zile (pentru 15 cazuri)

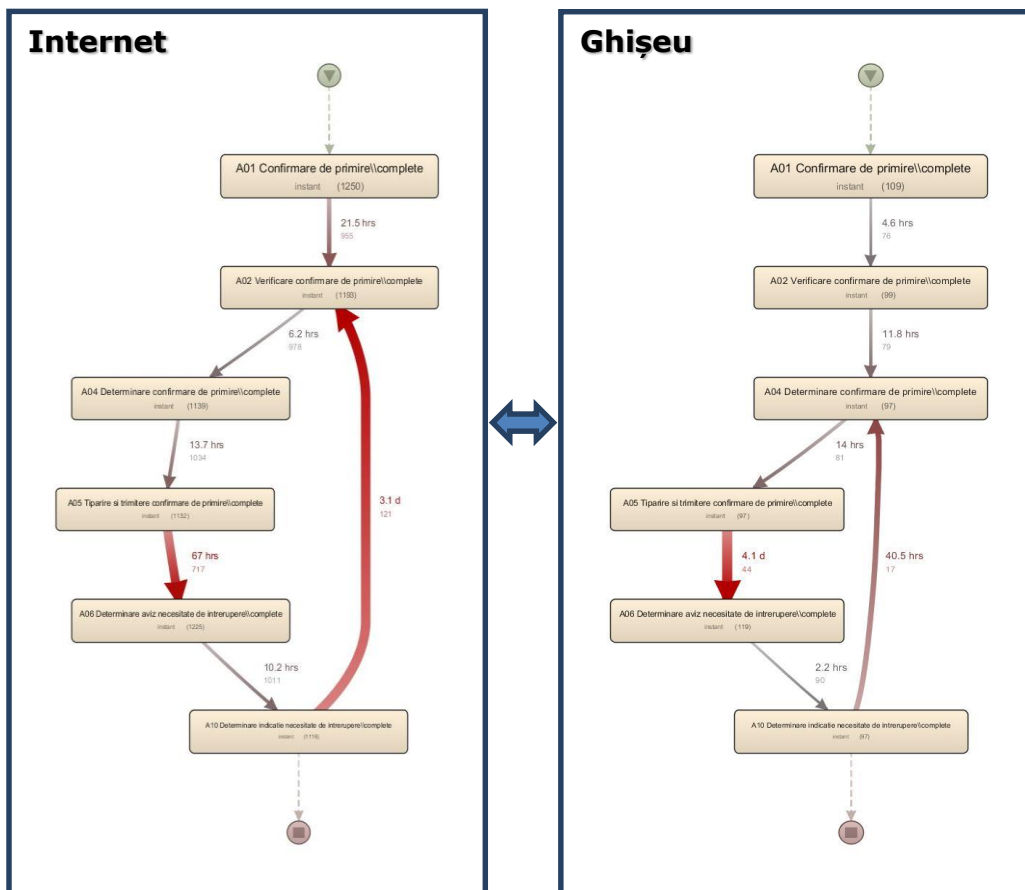


Fig. 6.32 Compararea proceselor a căror solicitare este via Internet cu cele de la Ghișeu

6.2.8. Transferul rezultatelor

Raportul final întocmit va fi prezentat experților de afaceri, iar rezultatele obținute vor fi dezbătute. Rezultatele obținute trebuie prezentate într-o manieră neutră, arătându-se în principal problemele procesului. Rezultatele pot crea disconfort unora dintre părțile interesate din cadrul organizației, de aceea se impune maximă atenție în prezentarea acestora.

Obiectivul principal al acestei etape este conștientizarea problemelor de către experții de afaceri. Aceștia sunt oamenii care cunosc cum trebuie să meargă afacerea și ei decid dacă problemele identificate sunt într-adevăr probleme relevante sau nu.

Datorită interacțiunii dintre experții de afaceri și analiștii de proces din cadrul acestei etape, pot apărea noi obiective de cercetare. De exemplu, investigarea cu un grad mai mare de detaliu a faptului că procesul este diferit în cazul cererilor primite pe Internet față de cererile primite la Ghișeu.

Dacă nu apar alte obiective de cercetare, iar experții de afaceri sunt mulțumiți cu ceea ce s-a găsit, atunci se poate trece mai departe. În caz contrar, se va reîncepe o iterație de investigare prin metodologia *Process Mining* pornind de la noile obiective de cercetare.

Metodologia *Process Mining* este cea mai eficientă atunci când analiștii de proces lucrează în strânsă colaborare cu experții de afaceri, într-un mod iterativ și interactiv. Acest lucru se poate observa atunci când se compară rezultatele dintr-un proiect clasic de BPM cu rezultatele unui proiect BPM bazat pe metodologia *Process Mining*.

Dacă nu se apelează la experții de afaceri, atunci transferul cunoștințelor este redus. Lipsa cunoașterii domeniului de afaceri în stagiul de inspecție poate duce la filtrări greșite și irelevante, ce pot duce mai departe la reprezentări incorecte de procese. Conform [68], proiectul BPM trebuie să includă obligatoriu analiști de proces și experți de afaceri, iar colaborarea dintre aceștia să fie strânsă. Această colaborare duce la iterații de analiză rapide și la idei concrete de îmbunătățire a procesului. Este recomandat ca metodologia *Process Mining* să fie cunoscută membrilor echipei care sunt experți în afacere. Este recomandat ca domeniul de afacere să fie cunoscut analiștilor de proces.

Se recomandă un workshop inițial (minim o zi) experților în afaceri pentru ca aceștia să fie familiari cu tehnicile *Process Mining* [68].

Chiar dacă la început întrebările de cercetare sunt abstracte, pe măsură ce se execută mai multe iterații, ele devin concrete. Explorările prin *Process Mining* pot genera rezultate neașteptate care pot conduce la definirea de întrebări concrete de cercetare. Compararea modelelor de proces este dificilă. În momentul de față, în aplicația ProM durează foarte mult să se filtreze date, să se compare date sau să se descopere modele. Acest lucru se poate îmbunătăți dacă se folosește plug-in-ul RapidProM pentru Rapid Miner [113] sau dacă se folosesc cuburi de proces [10].

6.3. Concluzii

În acest capitol s-a validat printr-un caz practic metodologia *CertiBPM-PM* care a fost propusă în capitolul 5, fiind parcurse toate etapele specifice *Process Mining* definite în noua metodologie. Pentru a se putea pune în aplicare metodologia s-a folosit un *jurnal de evenimente obținut din sistemele informatice* ale organizației, iar apoi a avut loc o *inspecție a jurnalului*, urmată de *descoperirea procesului* prin câțiva algoritmi și *alegerea celui mai bun model*. Alegerea s-a făcut prin compararea

coeficientului de potrivire obținut prin *verificarea conformității*. Pe urmă s-a analizat procesul din punctul de vedere al *performanței* și au fost identificate gâtuirile, ulterior găsiindu-se și cauzele acestor gâtuiuri. *Analiza organizațională* a relevat acele tipare de lucru ale personalului implicat în proces, care influențează performanța procesului. Următorul pas a fost *întocmirea raportului final al analizei și prezentarea acestuia către managementul organizației*. S-a putut observa că jurnalul de evenimente are multe activități, majoritatea fiind folosite rar, astfel încât s-a decurs la *filtrarea jurnalului* pentru ca acesta să conțină doar cele mai utilizate activități. După filtrare, modelul de proces generat a fost mult superior față de cel inițial.

În contextul metodologiei propuse, tehnicile *Process Mining* folosite au reușit în timp scurt să investigheze procesul, obținându-se rezultate clare ce pot fi folosite în luarea deciziei manageriale.

7. CONCLUZII GENERALE, CONTRIBUȚII PERSONALE ȘI PERSPECTIVELE CERCETĂRII

7.1. Concluzii generale asupra cercetării realizate

Prin teza de doctorat „**MANAGEMENTUL PROCESELOR DE AFACERI - PROVOCĂRI ȘI PERFORMANȚE ACTUALE**” s-a urmărit prezentarea disciplinei BPM, care se află la granița dintre management și tehnologia informației, cu elementele ei de noutate identificate și propuse. Acest demers a fost posibil datorită profilului de competențe ale autorului acestei lucrări, managementul și tehnologia informației fiind arii de interes ale acestuia, autorul tezei fiind absolvent al unor specializări, la universități din România, aferente ambelor domenii de știință. Ca urmare, tema de cercetare abordată, dar mai ales modul de rezolvare prin aprofundare a cercetărilor, precum și soluțiile la care s-a ajuns au fost posibile doar datorită mix-ului de competențe ale autorului, programul doctoral fiind o continuare firească a evoluției sale profesionale, valorificând la un înalt nivel cunoștințele din cele două domenii de specializare.

În **capitolul 1** s-a prezentat cadrul contextual și conceptual al cercetării, iar cele mai relevante concluzii sunt la nivelul disciplinei BPM și relative la stadiul actual al dezvoltării acesteia.

Disciplina BPM se află într-o continuă evoluție, iar această evoluție este declanșată de schimbarea continuă a mediului economic, organizațiile fiind mereu nevoite să se reorganizeze pentru a supraviețui. BPM formulează o propunere de abordare bazată pe proces care să facă față dinamicii schimbării. S-au identificat mai multe definiții ale termenului „proces”, dar aceste definiții sunt într-o permanentă transformare, întrucât organizațiile percep procesele în contexte diferite. Fenomenul actual cu care se confruntă companiile este explozia datelor, fenomen denumit simbolic „Big Data” și are o directă influență asupra disciplinei BPM, punând un accent ridicat pe prelucrarea inteligentă a datelor folosite de proces. S-au identificat două arii de interes ce reprezintă actualele tendințe: *Data Mining* și *Process Mining*. *Process Mining* este derivat din *Data Mining* și implică tehnici de creare a cunoștințelor prin analiza proceselor. Aceste tendințe care aparțin științei calculatoarelor au fost studiate în detaliu în contextul metodologiei de management specific BPM, rezultând obiectivul acestei teze: elaborarea unei metodologii de implementare a unui proiect BPM în strânsă legătură cu cele două tendințe, în special cu *Process Mining*. Implicarea autorului în proiectul *CertiBPM* a contribuit la stabilirea unei baze de cunoștințe specifice implementării de proiecte BPM în organizații.

În **capitolul 2** al tezei de doctorat au fost derulate cercetări asupra referențialului bibliografic referitoare la problematica celor două tendințe, *Data Mining* și *Process Mining*. Pentru domeniul *Process Mining* au fost identificate reprezentările de bază ale modelelor de procese, atât modele matematice, cât și modele vizuale sub formă de diagramă utilizate în companii. Reprezentările modelelor de proces se pot compara cu tipurile de date care sunt utile unui programator în vederea scrierii de cod. În concluzie se poate afirma că, aprofundarea și detalierea cunoașterii în domeniul acestor modele este necesară pentru a înțelege algoritmi specifici *Process Mining* care vor fi folosiți ulterior. Reprezentările bune ale modelelor de procese

trebuie să suporte diferite tipare: de concurență, de decizie, de iterație și altele, iar un analist de proces trebuie să știe să răspundă cum/când/unde trebuie folosite aceste modele. Pentru domeniul *Data Mining*, s-au identificat acele tehnici care aduc valoare în contextul *Process Mining*. Aceste cunoștințe fundamentale stau la baza înțelegerii tehnicilor de *Process Mining*.

În **capitolul 3** al tezei de doctorat au fost derulate cercetări asupra referențialului bibliografic privind problematica tehnicilor specifice *Process Mining*. Structura capitolului urmează evoluția unui proiect de *Process Mining*. Concluziile succint formulate se referă, în principal, la tehnicile *Process Mining* care utilizează date din evenimente obținute din sistemele informatice ale unei companii pentru a *descoperi* modele de proces, pentru a *verifica conformitatea* modelelor descoperite și pentru a *extinde* aceste modele cu informații despre gâturi, decizii sau folosirea resurselor. Aceste tehnici operează cu evenimente observate, iar nu cu modele manuale.

Pe baza cunoștințelor acumulate, structurate și a concluziilor formulate în capitolul 2, a fost posibilă explicarea algoritmilor de *descoperire* de procese și de *verificare a conformității*. Modelele de proces sunt îmbunătățite prin adăugarea de noi perspective (puncte de decizie, rețele sociale privind lucrul, performanța), făcând investigarea de proces mai valoroasă. Un aspect principal pentru pornirea demersului de *Process Mining* îl reprezintă extragerea datelor pentru întocmirea jurnalului de evenimente.

Prin cercetările asupra referențialului de specialitate, s-au concluzionat următoarele:

- *Process Mining* se bucură de o mare popularitate în cadrul academic, acest domeniu având o evoluție exponențială din punct de vedere al cunoștințelor. Există foarte mulți algoritmi de descoperire a procesului care s-au adăugat în ultimii ani.
- *Process Mining* reprezintă o punte importantă de legătură între *Data Mining* și modelarea proceselor de afaceri și de analiză [13].
- Datorită faptului că unui model de proces i se pot adăuga mai multe perspective, *Process Mining* nu se rezumă doar la descoperirea de modele.
- *Process Mining* nu este limitat doar la analiza offline. Există tehnici de detecție a devierilor și de predicție, care se bazează pe analiza online a datelor care sunt comparate cu date istorice.
- Calitatea datelor trebuie asigurată, pentru că rezultatele obținute sunt direct influențate de aceasta.
- Modelele obținute din jurnalele de eveniment reprezintă un punct de vedere al realității. Fiecare perspectivă reprezintă un alt punct de vedere al realității.
- În cadrul unei organizații, procesele trebuie văzute ca niște hărți. Managementul se va folosi de aceste hărți pentru a conduce compania către obiectivele stabilite.
- Procesele sunt într-o continuă schimbare, deci trebuie create cadre ce investighează în mod continuu procesele prin tehnicile *Process Mining*.

În **capitolul 4** al tezei de doctorat au fost derulate cercetări asupra mijloacelor software care se pot utiliza în contextul *Process Mining*. S-au ales mijloace software reprezentative, care au fost folosite în cadrul câtorva proiecte de analiză de procese. Aceste mijloace software sunt: ProM și Disco pentru *Process Mining* și RapidMiner pentru *Data Mining*. RapidMiner se poate integra cu platforma ProM prin plug-in-ul RapidProM. Concluziile acestui capitol sunt:

- Au fost alese mijloace software care sunt prezente atât în zona academică, cât și în zona comercială.

- Folosind platforma ProM se poate avea la acces la ultimele noutăți software ale domeniului *Process Mining*.
- Platforma ProM are foarte multe plug-in-uri și este greoi de utilizat. Cele mai utilizate funcționalități din ProM sunt prezente în Disco.
- Mijloacele software pentru *Data Mining* pot aduce beneficii prin inspecția jurnalului de evenimente și identificarea de tipare din aceste date.
- Mijloacele software pentru *Process Mining* și *Data Mining* sunt complementare, fiind indicată folosirea acestora împreună.

Capitolul 5 al tezei de doctorat tratează metodologiile de implementare BPM în cadrul unei organizații. A fost explicată metodologia *CertiBPM* care este o metodologie clasică de implementare a BPM. Din domeniul *Process Mining*, au fost cercetate acele metodologii existente pentru implementarea unui proiect specific *Process Mining*. Avantajele metodologiei *CertiBPM* specifice managementului, împreună cu avantajele tehnicilor *Process Mining* au condus la definirea unei noi metodologii, pe care am numit-o *CertiBPM-PM* (*CertiBPM for Process Mining*). Concluziile formulate în urma realizării cercetărilor prezentate în acest capitol sunt:

- Există puține metodologii pentru implementarea proiectelor de *Process Mining*.
- Domeniul *Process Mining* este foarte mult cercetat academic. În planul utilizării în companii, tehnicile *Process Mining* sunt încă în fază incipientă.
- Metodologia propusă combină robustețea unui proiect clasic BPM cu informațiile pentru afaceri oferite prin tehnicile *Process Mining*.

Capitolul 6 al tezei de doctorat se ocupă de validarea metodologiei *CertiBPM-PM*. Etapele descrise în capitolul 5 specifice *Process Mining* sunt executate asupra unui jurnal de date din cadrul unei instituții publice. Procesul este analizat cu această metodologie. Astfel, sunt descoperite diferite modele de proces cu diverși algoritmi, iar cel mai bun model este ales. Sunt identificate gâturi și posibile cauze ale acestora. Sunt descoperite tipare de lucru referitoare la oamenii care execută procesul, ajungând la concluzia că unele moduri de lucru influențează negativ performanța procesului. De asemenea, sunt descoperite modele de proces care sunt specifice diverselor canale de aplicare. Concluziile enunțate în urma finalizării cercetărilor aplicative (incluse în capitolul 6) sunt:

- Trebuie stabilite obiective clare atât pentru obținerea jurnalului, cât și pentru inspecția acestuia.
- Sunt necesare filtrări asupra jurnalului de evenimente pentru a răspunde corect întrebărilor de cercetare.
- Gâtuirile trebuie identificate atât în termeni de performanță (durate lungi între activități), cât și în termeni de frecvență (număr de cazuri).
- Rezultate mai bune sunt obținute prin folosirea în tandem a mijloacelor software.
- Pentru obținerea unor anumite rezultate au fost necesare comparări suplimentare de date prin folosirea calculului tabelar din Excel.
- Pentru compararea modelelor s-a folosit doar coeficientul potrivirii și nu s-a ținut cont de toate cele patru dimensiuni ale calității (potrivire, precizie, simplitate, generalizare).
- Mijloacele software permit identificarea vizuală foarte facilă de tipare: variante de cazuri, perioade de maximă/minimă ocupare, prelucrări de activități în loturi, resurse generaliste sau specializate, rata de sosire a cazurilor ș.a.

7.2. Contribuții personale

7.2.1. Contribuții în planul cercetărilor asupra referențialului bibliografic

Preliminar enunțării concluziilor în planul cercetărilor efectuate asupra referențialului bibliografic, actual și relevant, trebuie menționat faptul că pe perioada realizării programului doctoral, autorul a urmărit realizarea de cercetări pentru dobândirea de cunoștințe și expertiză în domeniul BPM și *Process Mining* (formarea prin cercetare, accesând diferite programe de formare). Astfel se menționează, în ordine cronologică, cele două certificări obținute

- ✓ Cunoștințele BPM de bază au fost acumulate și certificate prin absolvirea cursurilor **ECQA Certified Business Process Manager Foundation și ECQA Certified Business Process Manager Advanced**, în cadrul proiectului **Certified Business Process Manager – CertiBPM (contr. nr. LLP-LdV/TOI/10/RO/010)**, în anul 2012. În urma acestei activități au putut fi realizate primele sinteze și analize asupra referențialului bibliografic (capitolul 1), precum și formularea unei prime idei asupra metodologiei de implementare a proiectelor de BPM în organizații (parțial capitolul 5);
- ✓ Cunoștințele de bază pentru *Process Mining* au fost obținute prin absolvirea cursului **Process Mining: Data Science in Action** oferit de Universitatea Tehnică din Eindhoven prin platforma Coursera¹⁷, în anul 2015. Acest demers a contribuit la actualizarea și extinderea perspectivelor de cercetare a domeniului BPM, la elaborarea rafinată a sintezelor prezentate în capitolele 2, 3 și 4, aferente tezei de doctorat.

Cercetările asupra referențialului bibliografic au urmărit construirea unei **sinteze bibliografice pe tema implementării proiectelor BPM cu ajutorul Process Mining** care să constituie un fundament, nu doar pentru actuala teză, ci și pentru cercetări ulterioare în domeniu. Astfel, au putut fi puse în evidență tendințele actuale referitoare la abordarea BPM prin prisma *Process Mining*. Sinteza asupra referențialului bibliografic este utilă și din punct de vedere practic, aceasta putând fi valorificată de specialiști (IT și de management) din companii în dorința lor de a introduce conceptul, principiile, abordările și mijloacele software specifice *Process Mining* în propriile organizații.

Principalele contribuții personale aduse în urma cercetărilor asupra referențialului bibliografic sunt:

- Analiza și sinteza (inclusiv dobândirea de cunoștințe specifice) domeniului BPM și care au condus la descrierea modului de evoluție al științei și a tendințelor actuale: *Process Mining* și *Data Mining* (capitolul 1).
- Realizarea stadiului actual al cunoașterii pe următoarele teme (capitole ale lucrării):
 - ✓ modelarea proceselor de afaceri și tehnici pentru *Data Mining* utile în *Process Mining* (capitolul 2);
 - ✓ pregătirea și extragerea datelor, algoritmi de descoperire a proceselor, verificarea conformității proceselor și perspectivele *Process Mining* (capitolul 3);
 - ✓ metodologii de implementare *Process Mining* (capitolul 5).

¹⁷ <https://www.coursera.org/course/procmin>

7.2.2. Contribuții în planul cercetărilor teoretice

Principalele contribuții personale în planul cercetărilor teoretice sunt:

- Elaborarea unei definiții proprii pentru *procesul de afaceri* dată în contextul schimbării continue a sistemului organizațional;
- Enunțarea obiectivelor urmărite în cercetare (prin discuții informale cu experții implicați în proiectul CertiBPM) și explicitarea modului de atingere a acestora prin logica adoptată a acțiunilor și activităților de investigare întreprinse și care au fost asimilate apoi ca structură a tezei de doctorat;
- Elaborarea unei metodologii de implementare a unui proiect BPM cu elemente de *Process Mining*. **Metodologia elaborată, și denumită CertiBPM-PM, se constituie ca o combinație a modului/metodologiei de implementare a unui proiect BPM în versiunea CertiBPM, cu modul de conducere a unui proiect de Process Mining așa cum este el descris în metodologia PDM. La această combinație s-a adăugat conceptul iterativ aferent metodologiei PM².**

7.2.3. Contribuții în planul cercetărilor aplicative și experimentale

Principalele contribuții personale în planul cercetărilor aplicative și experimentale sunt:

- Rezultatele (în termenii avantaje - dezavantaje, beneficii – limite) cercetării unor mijloace software (disponibile în prezent pe piață) care să faciliteze implementarea unui proiect BPM. Au fost stabilite criteriile pentru alegerea acestor mijloace software și anume: să aibă costuri reduse (open-source), să fie capabile de a oferi suport pentru etapele metodologiei *CertiBPM-PM*, să se poată interconecta între ele și să fie ușor de utilizat (capitolul 4).
- Rezultatele cercetărilor aplicative pentru testarea și validarea **metodologiei CertiBPM-PM** (capitolul 6).

Concluziile cercetărilor realizate și contribuțiile originale aduse în domeniul științei și practicii manageriale confirmă atingerea pe deplin a obiectivului propus inițial al tezei de doctorat. Din cercetările teoretico-aplicative a rezultat că metodologia *CertiBPM-PM* și-a dovedit categoric utilitatea practică.

Rezultatele obținute în urma aplicării acesteia contribuie la o mai bună înțelegere la nivelul managementului unei organizații, a ceea ce se întâmplă în realitate cu procesele sale. Investigarea corectă și pedantă a proceselor contribuie la o mai bună luare a deciziilor, fiind utilă atât managementului operațional, cât și managementului strategic.

7.3. Direcții viitoare de cercetare

Cercetările teoretice și aplicative efectuate în cadrul tezei de doctorat creează perspective de dezvoltare pe următoarele direcții:

- Se poate încerca adăugarea de elemente *Process Mining* și în alte stagii ale metodologiei *CertiBPM* de implementare a unui proiect BPM:
 - ✓ În activitatea *Modelarea și documentarea proceselor (TO-BE)*, modelul care este propus spre îmbunătățire de către analiștii de proces se poate converti din format BPMN în rețea Petri, iar apoi se pot rula pe acesta teste

- de conformitate cu jurnalul de evenimente. Modelul TO-BE este astfel foarte repede analizat;
- ✓ În activitatea *Simularea și analiza proceselor*, modelul TO-BE în format de rețea Petri poate fi simulat și cu alte mijloace software precum CPN-Tools, abordare prezentată de Mărușter în [116];
 - ✓ În activitatea de *Implementare a noilor procese în organizație* se poate implementa un cadru pentru supravegherea noilor procese. Acesta poate calcula automat potrivirea modelului de proces implementat cu cazurile care se vor rula și poate semnala automat eventualele devieri. Se poate configura și un prag, depășirea acestuia alertând în mod automat persoanele responsabile. Această propunere a plecat de la conceptul de *Deteție* detaliat în Cadrul de lucru pentru *Process Mining* rafinat, din secțiunea 3.8.
 - Este interesant de aplicat metodologia *CertiBPM-PM* în diferite companii sau organizații locale, iar apoi rezultatele obținute să fie comparate cu cele ale unor companii din același domeniu din străinătate (ce operează pe piața globală) sau în alte zone ale țării. De asemenea, este interesant de aplicat metodologia *CertiBPM-PM* în procese similare care se execută în diferite puncte de lucru (de exemplu, procesul de deschidere al unui cont la diferite agenții bancare).
 - Crearea unui manual al metodologiei *CertiBPM-PM* în cadrul căreia tehnicile *Process Mining* să fie ușor asimilabile de către persoane ce nu au pregătire sau competențe certificate în domeniul științei calculatoarelor sau informaticii.

BIBLIOGRAFIE

1. van der Aalst W (2013) " Mine your own business ": using process mining to turn big data into real value. Proceedings of the European Conference on Information Systems
2. van der Aalst WM (2011) Process mining: discovering and improving Spaghetti and Lasagna processes. 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM). doi: 10.1109/CIDM.2011.6129461
3. van der Aalst WM, Adriansyah A, Van Dongen B (2011) Causal nets: A modeling language tailored towards process discovery. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6901 LNCS:28–42. doi: 10.1007/978-3-642-23217-6_3
4. van der Aalst WM, ter Hofstede A, Kiepuszewski B, Barros a. P (2003) Workflow patterns. Distributed and Parallel Databases 14:5–51. doi: 10.1023/A:1022883727209
5. van der Aalst WMP (1998) the Application of Petri Nets To Workflow Management. Journal of Circuits, Systems and Computers 08:21–66. doi: 10.1142/S0218126698000043
6. van der Aalst WMP (2008) Challenges in Business Process Analysis. Proceedings of the 9th International Conference on Enterprise Information Systems (ICEIS '07), LNBIP 12 27–42. doi: 10.1007/978-3-540-88710-2_3
7. van der Aalst WMP (2009) TomTom for business process management (TomTom4BPM). Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5565 LNCS:2–5. doi: 10.1007/978-3-642-02144-2_2
8. van der Aalst WMP (2011) Process Mining: Discovery, Conformance and Enhancement of Business Processes. doi: 10.1007/978-3-642-19345-3
9. van der Aalst WMP (2013) Decomposing Petri nets for process mining: A generic approach. Distributed and Parallel Databases 31:471–507. doi: 10.1007/s10619-013-7127-5
10. van der Aalst WMP (2013) Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining. Asia Pacific Business Process Management: First Asia Pacific Conference, AP-BPM 2013, Beijing, China, August 29-30, 2013 Selected Papers 1–22.
11. van der Aalst WMP (2015) Extracting Event Data from Databases to Unleash Process Mining. In: vom Brocke J, Schmiedel T (eds) BPM - Driving Innovation in a Digital World SE - 8. Springer International Publishing, pp 105–128
12. van der Aalst WMP, Adriansyah A, van Dongen B (2012) Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2:182–192. doi: 10.1002/widm.1045

13. van der Aalst WMP, Andriansyah A, De Medeiros A, Arcieri F, Baier T, Blickle T, Bose J, van den Brand P, Brandtjen R, Buijs J (2012) Process mining manifesto. BPM 2011 Workshops Proceedings 169–194. doi: 10.1016/j.is.2011.10.006
14. van der Aalst WMP, Dustdar S (2012) Process mining put into context. IEEE Internet Computing 16:82–86.
15. van der Aalst WMP, van Hee KM, ter Hofstede a. HM, Sidorova N, Verbeek HMW, Voorhoeve M, Wynn MT (2011) Soundness of workflow nets: Classification, decidability, and analysis. Formal Aspects of Computing 23:333–363. doi: 10.1007/s00165-010-0161-4
16. van der Aalst WMP, Reijers H a., Song M (2005) Discovering social networks from event logs. Computer Supported Cooperative Work 14:549–593. doi: 10.1007/s10606-005-9005-9
17. van der Aalst WMP, Reijers H a., Weijters a. JMM, van Dongen BF, Alves de Medeiros a. K, Song M, Verbeek HMW (2007) Business process mining: An industrial application. Information Systems 32:713–732. doi: 10.1016/j.is.2006.05.003
18. van der Aalst WMP, Rubin V, Verbeek HMW, Van Dongen BF, Kindler E, Günther CW (2010) Process mining: A two-step approach to balance between underfitting and overfitting. Software and Systems Modeling 9:87–111. doi: 10.1007/s10270-008-0106-z
19. van der Aalst WMP, Schonenberg MH, Son M (2011) Time prediction based on process mining. Information Systems 36:450–475. doi: 10.1016/j.is.2010.09.001
20. van der Aalst WMP, Stahl C (2011) Modeling Business Processes: A Petri Net-Oriented Approach. MIT press, Cambridge
21. van der Aalst WMP, Weijters T, Mărușter L (2004) Workflow mining: Discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering 16:1128–1142. doi: 10.1109/TKDE.2004.47
22. van der Aalst WMP, Weske M (2012) Business Process Management Concepts, Languages, Architectures, Second Edi. doi: 10.1007/978-3-642-28616-2
23. Academia Română (1998) Dicționarul explicativ al limbii române, ediția a II-a. Editura Univers Enciclopedic
24. Academia Română (2009) Dicționarul explicativ al limbii române (ediția a III-a, 2009, revăzută și adăugită). Univers Enciclopedic Gold
25. Adriansyah a., Van Dongen BF, Van Der Aalst WMP (2011) Conformance checking using cost-based fitness analysis. Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC 55–64. doi: 10.1109/EDOC.2011.12
26. Aldea CC, **Olariu C** (2014) Selecting the Optimal Software Solution under Conditions of Uncertainty. Procedia - Social and Behavioral Sciences 109:333–337. doi: <http://dx.doi.org/10.1016/j.sbspro.2013.12.467>
27. Alwahaishi S, Jaffar A, Vondrák I, Snášel V (2012) Business Process Models Representation by Deducing Interpretative Evidences on Intuitively Common Symbols. International Journal of Productivity Management and Assessment Technologies 1:29–39. doi: 10.4018/ijpmat.2012010103
28. Anderson JA (1995) An Introduction to Neural Networks. MIT Press

29. Azevedo A (2015) Data Mining and Business Intelligence. In: Azevedo A, Santos MF (eds) Integration of Data Mining in Business Intelligence Systems. IGI Global, pp 1–11
30. Bolt A, Sepúlveda M (2014) Process Remaining Time Prediction Using Query Catalogs. In: Lohmann N, Song M, Wohed P (eds) Business Process Management Workshops SE - 5. Springer International Publishing, pp 54–65
31. Börger E (2012) Approaches to modeling business processes: A critical analysis of BPMN, workflow patterns and YAWL. *Software and Systems Modeling* 11:305–318. doi: 10.1007/s10270-011-0214-z
32. Borja V, Mucientes M, Lama M (2015) ProDiGen : mining complete , precise and minimal structure process models with a genetic algorithm. *Information Sciences, Innovative Applications of Artificial Neural Networks in Engineering* 294:315–333. doi: 10.1016/j.ins.2014.09.057
33. Bose RPJC, Mans RS, van der Aalst WMP (2013) Wanna Improve Process Mining Results? It ' s High Time We Consider Data Quality Issues Seriously. *Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013 (2013)* 127–134.
34. Bowman FO (2004) The Curious History and Distressing Implications of the Criminal Provisions of the Sarbanes-Oxley Act and the Sentencing Guidelines Amendments That Followed. *Ohio State Journal of Criminal Law* 23:373–442.
35. Bozkaya M, Gabriels J, Werf JM van der (2009) Process Diagnostics: A Method Based on Process Mining. In: *International Conference on Information, Process, and Knowledge Management. IEEE*, pp 22–27
36. Brabänder E, Davis R (2007) ARIS design platform–getting started with BPM. Springer-Verlag London Limited, London
37. Bramer M (2013) Principles of Data Mining. doi: 10.1007/978-1-4471-4884-5
38. Bratosin C, Sidorova N, van der Aalst WM (2010) Distributed genetic process mining. *Evolutionary Computation (CEC), 2010 IEEE Congress on* 1–8. doi: 10.1109/CEC.2010.5586250
39. Buhl HU, Röglinger M, Moser F, Heidemann J (2013) Big data: A fashionable topic with(out) sustainable relevance for research and practice? *Business and Information Systems Engineering* 5:65–69. doi: 10.1007/s12599-013-0249-5
40. Buijs J (2014) Receipt phase of an environmental permit application process ("WABO"), CoSeLoG project. Eindhoven University of Technology. Dataset. doi: <http://dx.doi.org/10.4121/uuid:a07386a5-7be3-4367-9535-70bc9e77dbe6>
41. Buijs J, van Dongen B, van der Aalst WMP (2014) Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity. *International Journal of Cooperative Information Systems*. doi: 10.1142/S0218843014400012
42. Buijs JC a. M (2010) Mapping Data Sources to XES in a Generic Way. Master's Thesis 123.

43. Buijs JCAM, van Dongen BF, van der Aalst WMP (2012) On the role of fitness, precision, generalization and simplicity in process discovery. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7565 LNCS:305–322. doi: 10.1007/978-3-642-33606-5_19
44. Burattin A, Sperduti A, Aalst WMP van der (2014) Control-flow Discovery from Event Streams. Proceedings of the IEEE Congress on Evolutionary Computation. doi: 10.1109/CEC.2014.6900341
45. Cînda AI (2013) Întreprinderea agilă - Strategii manageriale și modele organizaționale. Universitatea "Politehnica" Timișoara
46. Chrissis MB, Konrad M, Shrum S (2011) CMMI for Development: Guidelines for Process Integration and Product Improvement. Pearson Education
47. Clempner J (2014) Verifying soundness of business processes: A decision process Petri nets approach. Expert Systems with Applications 41:5030–5040. doi: 10.1016/j.eswa.2014.03.005
48. Conger S (2015) Six Sigma and Business Process Management. In: vom Brocke J, Rosemann M (eds) Handbook on Business Process Management 1 SE - 6. Springer Berlin Heidelberg, pp 127–146
49. Davenport TH (1993) Process innovation : reengineering work through information technology. Harvard Business School Press, Boston, Mass
50. Davenport TH, Short JE (1990) The New Industrial Engineering : Information Technology And Business Process Redesign. Routledge
51. Delen D, Demirkan H (2013) Data, information and analytics as services. Decision Support Systems 55:359–363. doi: 10.1016/j.dss.2012.05.044
52. Diaz M (2009) Petri Nets Fundamental Models, Verification and Applications, 1st ed. Wiley-ISTE
53. Domingos P (1999) The role of Occam's Razor in knowledge discovery. Data Mining and Knowledge Discovery 3:409–425. doi: 10.1023/a:1009868929893
54. van Dongen B, Busi N, Pinna GM, van der Aalst WMP (2007) An Iterative Algorithm for Applying the Theory of Regions in Process Mining. Workshop on Formal Aspects of Business Processes and Web Services
55. van Dongen BF, de Medeiros AKA, Verbeek HMW, Weijters A, van der Aalst WMP (2005) The ProM framework: A new era in process mining tool support. In: APPLICATIONS AND THEORY OF PETRI NETS 2005, PROCEEDINGS. pp 444–454
56. van Dongen BF, Shabani S (2015) Relational XES : Data Management for Process Mining. BPM Center Report BPM-15-02
57. Drăgan M (2010) Rețelele Petri colorate și utilitatea lor în modelarea, validarea și simularea sistemelor. Meridian Ingineresc 2:14–21.
58. Draghici A, Albulescu C, **Olariu C**, Draghici G (2012) Improving Business Process Management in Manufacturing – A Certification And Qualification Model. New face of TMCR, Modern Technologies, Quality and Innovation
59. Draghici A, Draghici G, **Olariu C**, Canda A (2012) A way to support SPI strategy through CertiBPM training and certification program in Romania. Communications in Computer and Information Science 301 CCIS:276–287. doi: 10.1007/978-3-642-31199-4_24

60. Draghici A, Draghici G, **Olariu C**, Canda A (2012) Proposal For The Romanian Master Programs International Recognition Through CertiBPM Initiative. In: Avasilcai S, Tjoa AM, Hutu CA, Weippl E (eds) "E" Era & Higher Education, Proceedings of the 7th International Conference Quality Management in Higher Education (QMHE 2012). Austrian Computer Society, pp 383–390
61. Draghici A, Draghici G, **Olariu C**, Canda A (2012) Romanian Market Acceptance for Business Process Management Skills Development. *Procedia Technology* 5:537–546. doi: 10.1016/j.protcy.2012.09.059
62. Draghici A, Draghici G, **Olariu C**, Canda A (2012) ICT Infrastructure Created to Support CertiBPM Project and Training Program. *Procedia Technology* 5:527–536. doi: 10.1016/j.protcy.2012.09.058
63. Draghici A, **Olariu C**, Rozman T (2012) Business Process Management Training and Certification Program: A Slovenian-Romanian Collaboration Experience. In: Knowledge and Learning: Global Empowerment; Proceedings of the Management, Knowledge and Learning International Conference 2012. International School for Social and Business Studies, Celje, Slovenia, pp 651–663
64. Draghici A, **Olariu C**, Rozman T (2014) Certified Business Process Manager: a transfer of innovation experience. *International Journal of Innovation and Learning* 16:12. doi: 10.1504/IJIL.2014.063371
65. Dumas M, La Rosa M, Mendling J, Reijers H a. (2013) Fundamentals of Business Process Management. doi: 10.1007/978-3-642-33143-5
66. Dunham MH (2006) Data mining: Introductory and advanced topics. Pearson Education India
67. van Eck ML, Buijs JCAM, van Dongen BF (2015) Genetic Process Mining: Alignment-Based Process Model Mutation. *Business Process Management Workshops, Lecture Notes in Business Information Processing* 202:291–303. doi: 10.1007/978-3-319-15895-2_25
68. van Eck ML, Lu X, Leemans SJJJ, van der Aalst WMP (2015) PM²: A Process Mining Project Methodology. In: Zdravkovic J, Kirikova M, Johannesson P (eds) *Advanced Information Systems Engineering SE - 19*. Springer International Publishing, pp 297–313
69. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI magazine* 17:37–53. doi: 10.1609/aimag.v17i3.1230
70. Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (1996) *Advances in Knowledge Discovery and Data Mining*. The MIT Press
71. Foley É, Guillemette MG (2012) What is business intelligence? *Organizational Applications of Business Intelligence Management: Emerging Trends* 52.
72. Grundspenkis J, Mislevics A (2010) Intelligent Agents for Business Process Management Systems. In: Pankowska M (ed) *Infonomics for Distributed Business and Decision-Making Environments: Creating Information System Ecology*. IGI Global, Hershey, PA, USA, pp 97–131
73. Günther CW, Aalst WMP Van Der (2007) Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. *Business Process Management - Lecture Notes in Computer Science* 4714:328–343. doi: 10.1007/978-3-540-75183-0

74. Günther CWC, Verbeek E (2014) XES standard 2.0. In: Fluxicon Process Laboratories. http://www.xes-standard.org/_media/xes/xesstandarddefinition-2.0.pdf. Accessed 21 May 2015
75. Gupta JND, Sharma SK (2004) Intelligent enterprises of the 21st century. IGI Global, Hershey, PA
76. Hammer M (1990) Reengineering work: don't automate, obliterate. *Harvard business review* 68:104–112.
77. Hammer M, Champy J (1993) Reengineering the Corporation: A Manifesto for Business Revolution. HarperBusiness
78. Han J, Kamber M, Pei J (2012) Data Mining: Concepts and Techniques, 3rd Ed. Morgan Kaufmann Publishers
79. Hâncean MG (2014) Rețelele sociale. Editura Polirom
80. Hand D, Mannila H, Smyth P (2002) Principles of Data Mining. The MIT Press
81. Harmon P (2014) Business Process Change: A Business Process Management Guide for Managers and Process Professionals, 3rd ed. Morgan Kaufmann, Waltham, MA
82. Harmon P (2015) The Scope and Evolution of Business Process Management. In: vom Brocke J, Rosemann M (eds) Handbook on Business Process Management 1 SE - 3. Springer Berlin Heidelberg, pp 37–80
83. Harmon P, Wolf C (2012) The State of Business Process Management 2012. In: BPTrends. http://www.bptrends.com/bpt/wp-content/surveys/2012-_BPT_SURVEY-3-12-12-CW-PH.pdf. Accessed 25 May 2015
84. Harmon P, Wolf C (2014) The State of Business Process Management 2014. In: BPTrends. <http://www.bptrends.com/bpt/wp-content/uploads/BPTrends-State-of-BPM-Survey-Report.pdf>. Accessed 25 May 2015
85. Haro JM de, Carrión JA (2012) Leadership Talent Development in the New Economy. In: Juana-Espinosa S de, Fernandez-Sanchez JA, Manresa-Marhuenda E, Valdes-Conca J (eds) Human Resource Management in the Digital Economy: Creating Synergy between Competency Models and Information. IGI Global, Hershey, PA, USA, pp 258–289
86. Hatzakis T (2008) Closing the {Business-IT} {'Gap':} A {Trust-Building} Approach. {AMCIS} 2008 Proceedings
87. Herschel G, Linden A, Kart A (2015) Magic Quadrant for Advanced Analytics Platforms. Stamford
88. ter Hofstede A, van der Aalst WM (2005) YAWL: Yet Another Workflow Language. *Information Systems* 30:245–275.
89. ter Hofstede AHM, van der Aalst WMP, Adams M, Russell N (2010) Modern Business Process Automation - YAWL and its Support Environment. doi: 10.1007/978-3-642-03121-2
90. International Organization for Standardization (ISO) ISO 9001 Revision. http://www.iso.org/iso/home/standards/management-standards/iso_9000/iso9001_revision.htm. Accessed 1 Jan 2015
91. Jablonski S (2009) Process Modeling for Holistic Process Management. In: Cardoso J, Aalst W van der (eds) Handbook of Research on Business Process Modeling. IGI Global, Hershey, PA, USA, pp 49–68
92. Jensen K, Kristensen LM (2009) Coloured Petri Nets Modelling and Validation of Concurrent Systems. doi: 10.1007/b95112
93. Jeston J, Nelis J (2014) Business Process Management. Taylor & Francis

94. Johansson HJ (1993) Business process reengineering: breakpoint strategies for market dominance. Wiley
95. Kalenkova AA, Leoni M De, van der Aalst WMP (2014) Discovering , Analyzing and Enhancing BPMN Models Using ProM. CEUR Workshop Proceedings 1295:36–41.
96. Kaplan RS, Norton DP (1996) The balanced scorecard: translating strategy into action. Harvard Business Press
97. Kilinc I, Oncu MA, Tasgit YE (2014) The Business World as a Battlefield. In: Hacıoğlu Ü, Dincer H (eds) Globalization and Governance in the International Political Economy. IGI Global, Hershey, PA, USA, pp 306–318
98. Kindler E (2004) On the Semantics of EPCs: A Framework for Resolving the Vicious Circle. In: Desel J, Pernici B, Weske M (eds) Business Process Management SE - 6. Springer Berlin Heidelberg, pp 82–97
99. Kocbek M, Jost G, Hericko M, Polancic G (2015) Business process model and notation: The current state of affairs. Computer Science and Information Systems 12:509–539. doi: 10.2298/CSIS140610006K
100. Lam VSW (2015) The Basis and Core Knowledge of Business Process Management. In: Khosrow-Pour M (ed) Encyclopedia of Information Science and Technology, Third Edition. IGI Global, Hershey, PA, USA, pp 5303–5314
101. Lanz A, Weber B, Reichert M (2014) Time patterns for process-aware information systems. Requirements Engineering 19:113–141. doi: 10.1007/s00766-012-0162-3
102. Leemans SJJ, Fahland D, van der Aalst WMP (2013) Discovering Block-Structured Process Models from Event Logs - A Constructive Approach. Application and Theory of Petri Nets and Concurrency, Lecture Notes in Computer Science 7927:311–329. doi: 10.1007/978-3-642-38697-8_17
103. Leemans SJJ, Fahland D, van der Aalst WMP (2014) Discovering block-structured process models from event logs containing infrequent behaviour. Lecture Notes in Business Information Processing 171:66–78. doi: 10.1007/978-3-319-06257-0
104. de Leoni M, van der Aalst WMP (2013) Data-aware process mining: discovering decisions in processes using alignments. Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13 1454. doi: 10.1145/2480362.2480633
105. de Leoni M, van der Aalst WMP, Dees M (2014) A General Framework for Correlating Business Process Characteristics. In: 12th conference on business process management Haifa Israel. Springer Berlin Heidelberg, pp 250–266
106. Leyking K, Ziemann J (2008) Service-Oriented Architecture for Business Management. In: Putnik GD, Cruz-Cunha MM (eds) Encyclopedia of Networked and Virtual Organizations. IGI Global, Hershey, PA, USA, pp 1437–1445
107. Liu G (2014) Some Complexity Results for the Soundness Problem of Workflow Nets. Services Computing, IEEE Transactions on 7:322–328. doi: 10.1109/TSC.2013.36
108. Liu X, Chen J, Ji Y, Yu Y (2015) Q-learning Algorithm for Task Allocation Based on Social Relation. In: Cao J, Wen L, Liu X (eds) Process-Aware Systems SE - 5. Springer Berlin Heidelberg, pp 49–58

109. Lohmann N, Verbeek E (2009) Petri net transformations for business processes—a survey. *Transactions on Petri Nets and Other* 37–65.
110. Mămăligă T, Fletcher GHL (2013) Realizing a Process Cube Allowing for the Comparison of Event Data. Master’s Thesis
111. Mans R, van der Aalst WMP, Vanwersch RB (2015) Data Quality Issues. In: *Process Mining in Healthcare SE - 6*. Springer International Publishing, pp 79–88
112. Mans RS, van der Aalst WMP, Vanwersch RJB (2015) Process Mining in Healthcare. doi: 10.1007/978-3-319-16071-9
113. Mans RS, Aalst WMP Van Der, Verbeek HMW (2014) Supporting Process Mining Workflows with RapidProM. *CEUR Workshop Proceedings* 1295:56–61.
114. Marcu F, Maneca C (1986) *Dicționar de neologisme*, Ed. a 3-a. Editura Academiei Republicii Socialiste România
115. Mărușter L (2003) A machine learning approach to understand business processes. Technische Universiteit Eindhoven
116. Mărușter L, Beest NRTP (2009) Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems* 21:267–297. doi: 10.1007/s10115-009-0224-0
117. de Medeiros AKA, van der Aalst WMP, Weijters A (2003) Workflow mining: Current status and future directions. *On The Move to Meaningful Internet Systems 2003 CoopIS DOA and ODBASE* 2888:389–406. doi: 10.1007/978-3-540-39964-3_25
118. De Medeiros AKA, Van Dongen BF, Van Der Aalst WMP, Weijters AJMM (2004) Process mining: Extending the α -algorithm to mine short loops. Eindhoven University of Technology Eindhoven 1–25.
119. De Medeiros AKA, Weijters AJMM, Van Der Aalst WMP (2007) Genetic process mining: An experimental evaluation. *Data Mining and Knowledge Discovery* 14:245–304. doi: 10.1007/s10618-006-0061-7
120. Mierswa I, Wurst M, Klinkenberg R, Scholz M, Euler T (2006) YALE: Rapid prototyping for complex data mining tasks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2006:935–940. doi: 10.1145/1150402.1150531
121. Mikut R, Reischl M (2011) Data mining tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1:431–443. doi: 10.1002/widm.24
122. Mosteller F, Tukey JW (1977) *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley Publishing Company
123. Motahari-Nezhad HR, Saint-Paul R, Casati F, Benatallah B (2011) Event correlation for process discovery from web service interaction logs. *VLDB Journal* 20:417–444. doi: 10.1007/s00778-010-0203-9
124. Murata T (1989) Petri Nets: Properties , Analysis and Applications. *Proceedings of the IEEE* 77:541–580. doi: 10.1109/5.24143
125. Nong Y (2003) *The Handbook of Data Mining (Human Factors and Ergonomics)*. Lawrence Earlbaum Associates, Publishers, Mahwah, New Jersey London
126. Oanea O (2007) Verification of Soundness and Other Properties of Business Processes. Technische Universiteit Eindhoven

127. **Olariu C**, Aldea CC (2014) Managing Processes for Virtual Teams – A BPM Approach. *Procedia - Social and Behavioral Sciences* 109:380–384. doi: 10.1016/j.sbspro.2013.12.476
128. **Olariu C**, Canda A, Aldea CC, Draghici A (2012) A Proposed Approach for the Project Meeting Process Improvement in the Case of Virtual Teams. In: *RMEE 2012 proceedings, „A new dilemma: between East and Vest”*. Editura Todesco Publishing House, pp 335–343
129. **Olariu C**, Canda A, Draghici A, Rozman T (2013) Planning Business Process Management Implementation by a Human Resources Development Support Initiative. In: *Advances in Intelligent Systems and Computing*. Springer-Verlag Berlin, pp 669–680
130. **Olariu C**, Gogan M, Rennung F (2014) Switching the Center of Software Development from IT to Business Experts Using Intelligent Business Process Management Suites. *SOFA 2014*
131. Ould MA (2005) *Business Process Management: A Rigorous Approach*. British Computer Society
132. Pater AM (2006) *Contribuții la dezvoltarea algoritmilor de extragere a datelor din bazele de date multi-nivel, teză de doctorat*. Universitatea Politehnica Timișoara
133. Paul Puah KY, Nelson Tang KH (2000) Business process management, a consolidation of BPR and TQM. *Management of Innovation and Technology, 2000 ICMIT 2000 Proceedings of the 2000 IEEE International Conference on* 1:110–115 vol.1. doi: 10.1109/ICMIT.2000.917294
134. Paulk MC, Averill EL (1994) *The Capability Maturity Model : guidelines for improving the software process*. Addison-Wesley, Reading, MA [etc.]
135. Pawlewski P (2012) *Petri Nets - Manufacturing and Computer Science*. doi: 10.5772/2578
136. Polato M, Sperduti A, Burattin A, de Leoni M (2014) Data-Aware Remaining Time Prediction of Business Process Instances. *Proceedings of IEEE International Joint Conference on Neural Networks*. doi: 10.1109/IJCNN.2014.6889360
137. Porter ME (2008) *Competitive advantage: Creating and sustaining superior performance*. Simon and Schuster
138. Rebuge Á, Ferreira DR (2012) Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems* 37:99–116. doi: 10.1016/j.is.2011.01.003
139. Rogge-Solti A, Weske M (2015) Prediction of business process durations using non-Markovian stochastic Petri nets. *Information Systems*. doi: 10.1016/j.is.2015.04.004
140. Rozinat A *ProM Tips — Which Mining Algorithm Should You Use?* <https://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/>. Accessed 14 Aug 2015
141. Rozinat A (2005) *Conformance Testing: Measuring the Alignment Between Event Logs and Process Models*. Master’s Thesis. Eindhoven University of Technology
142. Rozinat A (2010) *Process mining: conformance and extension*. PhD Thesis. Eindhoven University of Technology

143. Rozinat A, van der Aalst WMP (2006) Conformance Testing: Measuring the Fit and Appropriateness of Event Logs and Process Models. In: Business Process Management Workshops , Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp 163–176
144. Rozinat A, van der Aalst WMP (2006) Decision Mining in ProM. Proceedings of the 4th International Conference on Business Process Management (BPM 2006) 4102:420–425. doi: 10.1007/11841760_33
145. Rozinat A, van der Aalst WMP (2006) Decision mining in business processes. BPM Center Report BPM-06-10
146. Rozinat A, van der Aalst WMP (2008) Conformance checking of processes based on monitoring real behavior. *Information Systems* 33:64–95. doi: 10.1016/j.is.2007.07.001
147. Rummler GA, Brache AP (1995) Improving Performance: How to Manage the White Space on the Organization Chart.
148. Russell N (2005) Workflow resource patterns: Identification, representation and tool support. *Advanced Information* 216–232. doi: 10.1007/11431855_16
149. Seleem YZ, Mohamed MH, Hussain KF (2013) Improving genetic process mining using Honey Bee algorithm. In: 2013 Second International Conference on Informatics & Applications (ICIA). IEEE, pp 59–65
150. Sidorova N, Stahl C, Trčka N (2011) Soundness verification for conceptual workflow nets with data: Early detection of errors with the most precision possible. *Information Systems* 36:1026–1043. doi: 10.1016/j.is.2011.04.004
151. Silver B (2011) BPMN Method and Style, 2nd Edition, with BPMN Implementer’s Guide: A Structured Approach for Business Process Modeling and Implementation Using BPMN 2. Cody-Cassidy Press, Aptos, Calif
152. Smith H, Fingar P (2003) Business Process Management: The Third Wave. Meghan-Kiffer Press
153. Song J, Kim M, Kim H, Kim K (2010) A Framework: Workflow-Based Social Network Discovery and Analysis. *Computational Science and Engineering (CSE)*, 2010 IEEE 13th International Conference on 421–426. doi: 10.1109/CSE.2010.74
154. Song M, van der Aalst WMP (2007) Supporting process mining by showing events at a glance. In Proceedings of 17th Annual Workshop on Information Technologies and Systems 139–147.
155. Song M, van der Aalst WMP (2008) Towards comprehensive support for organizational mining. *Decision Support Systems* 46:300–317. doi: 10.1016/j.dss.2008.07.002
156. Syropoulos A (2001) Mathematics of Multisets. In: Calude C, Păun G, Rozenberg G, Salomaa A (eds) Multiset Processing SE - 17. Springer Berlin Heidelberg, pp 347–358
157. Tka M, Ghannouchi SA (2014) Comparison of Business Process Models as Part of BPR Projects. *Information Resources Management Journal (IRMJ)* 27:53–66. doi: 10.4018/irmj.2014010104
158. Urh B, Kern T, Roblek M (2010) Business Process Modification Management. In: Management Association IR (ed) Business Information Systems. IGI Global, pp 1748–1759

159. Verbeek HMW, Buijs JC a M, Van Dongen BF, Van Der Aalst WMP (2011) XES, XESame, and ProM 6. Lecture Notes in Business Information Processing 72 LNBIP:60–75. doi: 10.1007/978-3-642-17722-4_5
160. Verbeek HMW, Buijs JCAM, Dongen BF Van, van der Aalst WMP (2010) ProM 6 : The Process Mining Toolkit. Proc of BPM Demonstration Track 615:34–39.
161. Vercellis C (2009) Business intelligence: data mining and optimization for decision making. Editorial John Wiley and Sons
162. Warboys BC (1999) Business Information Systems: A Process Approach. McGraw-Hill
163. De Weerd J, De Backer M, Vanthienen J, Baesens B (2012) A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. Information Systems 37:654–676. doi: 10.1016/j.is.2012.02.004
164. Weijters AJMM, Van Der Aalst WMP, Medeiros AKA De (2006) Process Mining with the HeuristicsMiner Algorithm. Cirp Annals-manufacturing Technology 166:1–34. doi: 10.1.1.118.8288
165. Weijters AJMM, Ribeiro JTS (2011) Flexible Heuristics Miner (FHM). Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on 310–317. doi: 10.1109/CIDM.2011.5949453
166. Weiss M, Amyot D (2005) Business Process Modeling with URN. International Journal of E-Business Research (IJEBR) 1:63–90. doi: 10.4018/jebr.2005070104
167. Wen L, Van Der Aalst WMP, Wang J, Sun J (2007) Mining process models with non-free-choice constructs. Data Mining and Knowledge Discovery 15:145–180. doi: 10.1007/s10618-007-0065-y
168. van der Werf JME, van Dongen BF, Hurkens CA, Hee KM Van, Serebrenik A (2008) Process Discovery using Integer Linear Programming. Applications and Theory of Petri Nets 5062:368–387.
169. Weske M (2007) Business process management: concepts, languages, architectures. Leipzig, Alemania: Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-540-73522-9
170. WfMC (1999) Workflow Management Coalition Terminology & Glossary. Management 39:1–65. doi: 10.1016/S0019-0578(00)00014-8
171. Wynn MT, Ouyang C, Adams M (2013) YAWL4Industry : Reflections on using YAWL for Industry Projects. Proceedings of the 1st YAWL Symposium 982:26–32.
172. Yakovlev A (2013) Business Intelligence in Corporate Governance and Business Processes Management. In: Naidenova X, Ignatov DI (eds) Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems. IGI Global, Hershey, PA, USA, pp 249–269
173. process - definition of process in English from the Oxford dictionary. <http://www.oxforddictionaries.com/definition/english/process>. Accessed 7 Feb 2015
174. Social Network Miner. <http://www.processmining.org/online/snminer>. Accessed 1 Aug 2015

ANEXA 1 – Poster conținând notații pentru modelarea proceselor de afaceri (BPMN)

Poster - Notația pentru Modelarea Proceselor de Afaceri (BPMN)

Obiecte Grafice ale unui Diagramă de Proces de Afaceri

Obiecte Grafice de Conectare ale unui Diagramă de Proces de Afaceri

Business Process Diagram Notation – Symbols and Connectors

Modelul de afaceri: modelul în BPMN / BPMN Model

Eventimente

1) Simbolul este utilizat pentru a reprezenta un eveniment care este un punct de pornire sau de sfârșit al unei activități sau al unui flux.

Tip	Simbol	Descriere
Start		Eveniment de început
End		Eveniment de sfârșit
Intermediate		Evenimente intermediare

Activități

2) Simbolul este utilizat pentru a reprezenta o activitate care este o sarcină sau o sub-proces.

3) Simbolul este utilizat pentru a reprezenta un sub-proces inclus.

4) Simbolul este utilizat pentru a reprezenta un eveniment de activitate.

Portațiuni

5) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate care este un eveniment de activitate.

6) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Culoare

7) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

8) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Obiecte grafice de conectare

9) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

10) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Mecanismul Fluxurilor de Securitate

11) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

12) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Utilizarea corectă a fluxurilor

13) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

14) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Utilizarea corectă a evenimentelor temporale

15) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

16) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Utilizarea corectă a fluxurilor în cadrul căușelor

17) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

18) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Utilizarea corectă a mecanismului fluxului de securitate

19) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

20) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Utilizarea corectă a simbolurilor

21) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

22) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Soluții pentru fluxuri de lucru

23) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

24) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Descrierea Posterului BPMN

25) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

26) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

Obiecte Grafice de Conectare ale unui Diagramă de Proces de Afaceri

27) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

28) Simbolul este utilizat pentru a reprezenta un eveniment de activitate care este un eveniment de activitate.

ANEXA 2 – Jurnal de evenimente în format XES

```

<log>
  <extension name="Lifecycle" prefix="lifecycle" uri="http://www.xes-
standard.org/lifecycle.xesext"/>
  <extension name="Time" prefix="time" uri="http://www.xes-
standard.org/time.xesext"/>
  <extension name="Concept" prefix="concept" uri="http://www.xes-
standard.org/concept.xesext"/>
  <extension name="Semantic" prefix="semantic" uri="http://www.xes-
standard.org/semantic.xesext"/>
  <extension name="Organizational" prefix="org" uri="http://www.xes-
standard.org/org.xesext"/>
  <extension name="Order" prefix="order"
uri="http://my.company.com/xes/order.xesext"/>
  <global scope="trace">
    <string key="concept:name" value="unknown"/>
  </global>
  <global scope="event">
    <string key="concept:name" value="unknown"/>
    <string key="lifecycle:transition" value="unknown"/>
    <string key="org:resource" value="unknown"/>
  </global>
  <classifier name="Activity classifier" keys="concept:name
lifecycle:transition"/>
  <string key="concept:name" value="Example log" />
  <trace>
    <string key="concept:name" value="Order 1" />
    <float key="order:totalValue" value="2142.38" />
    <event>
      <string key="concept:name" value="Create" />
      <string key="lifecycle:transition" value="complete" />
      <string key="org:resource" value="Wil" />
      <date key="time:timestamp" value="2009-01-
03T15:30:00.000+01:00" />
      <float key="order:currentValue" value="2142.38" />
      <string key="details" value="Order creation details">
        <string key="requestedBy" value="Eric" />
        <string key="supplier" value="Fluxi Inc." />
        <date key="expectedDelivery" value="2009-01-
12T12:00:00.000+01:00" />
      </string>
    </event>
  </trace>
</log>

```

ANEXA 3 – Activitățile jurnalului de evenimente considerate în cadrul cercetărilor aplicative

Nr.	Activitate
1	A01 Confirmare de primire
2	A02 Verificare confirmare de primire
3	A03 Ajustare confirmare de primire
4	A04 Determinare confirmare de primire
5	A05 Tiparire si trimitere confirmare de primire
6	A06 Determinare aviz necesitate de intrerupere
7	A07-1 Draft intern aviz aspect 1
8	A07-2 Draft intern aviz aspect 2
9	A07-3 Draft intern aviz oprire pentru aspect 3
10	A07-4 Draft intern aviz oprire pentru aspect 4
11	A07-5 Draft intern aviz aspect 5
12	A08 Draft si trimitere cerere pentru aviz
13	A09-1 Procesare sau primire aviz de la grupul 1
14	A09-2 Procesare sau primire aviz de la grupul 2
15	A09-3 Procesare sau primire aviz de la grupul 3
16	A09-4 Procesare sau primire aviz de la grupul 4
17	A10 Determinare indicatie necesitate de intrerupere
18	A11 Creare document X - cerere fara licenta
19	A12 Verificare document X - cerere fara licenta
20	A13 Ajustare document X - cerere fara licenta
21	A14 Determinare document X - cerere fara licenta
22	A15 Tiparire document X - cerere fara licenta
23	A16 Raportare motive pentru intrerupere cerere
24	A17 Vizualizare raport Y indicatie pentru oprire
25	A18 Ajustare raport Y indicatie pentru oprire
26	A19 Determinare raport Y indicatie pentru oprire
27	A20 Tiparire raport Y indicatie pentru oprire