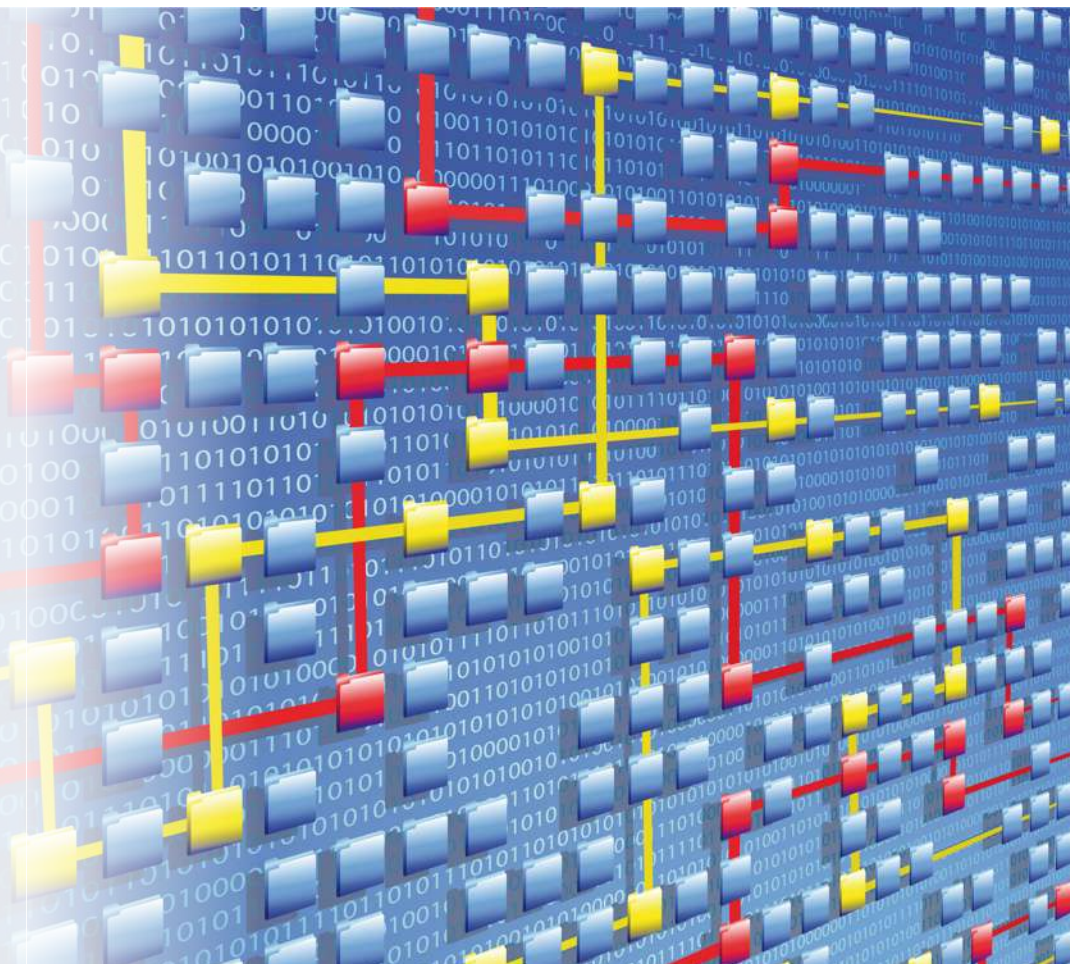


# Διακήρυξη για την Εξόρυξη Διαδικασιών



Διακήρυξη καλείται η «εξαγγελία ή γνωστοποίηση που γίνεται γραπτά ή προφορικά με επίσημο ή πανηγυρικό τρόπο» από μία ομάδα ανθρώπων. Η παρούσα διακήρυξη συγγράφηκε από μέλη και υποστηρικτές της Ομάδας Κρούσης Εξόρυξης Διαδικασιών της ΙΕΕΕ. Ο σκοπός της ομάδας είναι η προώθηση της έρευνας, της ανάπτυξης, της εκπαίδευσης, της εκτέλεσης, της εξέλιξης και της κατανόησης της εξόρυξης διαδικασιών.

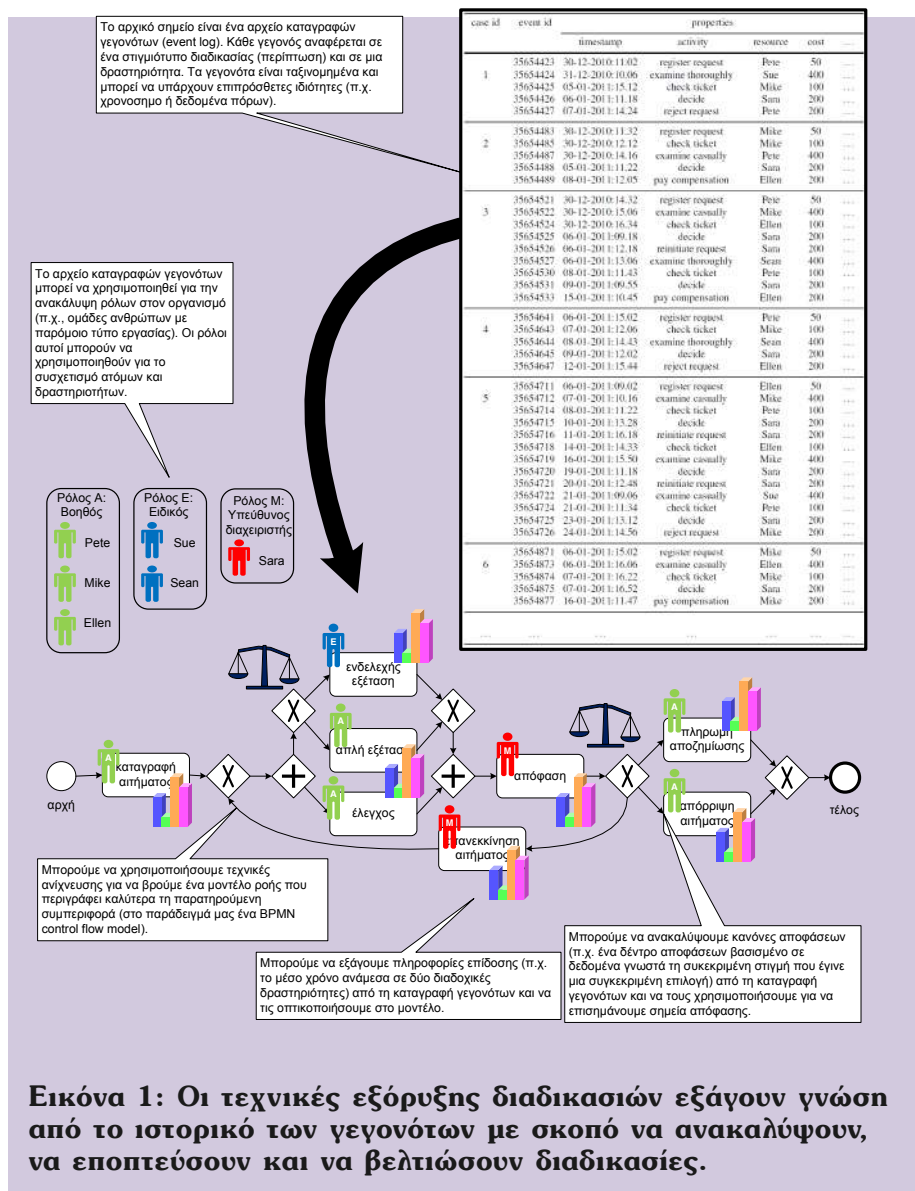
Η εξόρυξη διαδικασιών είναι ένας σχετικά νέος ερευνητικός τομέας ο οποίος ορίζεται ανάμεσα στην υπολογιστική νοημοσύνη και την εξόρυξη δεδομένων και στη μοντελοποίηση και ανάλυση των διαδικασιών. Η έννοια της εξόρυξης διαδικασιών συνίσταται στην ανακάλυψη, εποπτεία και στον εμπλουτισμό των διαδικασιών επί της ουσίας (δηλ. όχι των διαδικασιών επί των τύπων) μέσω της εξαγωγής γνώσης από το ιστορικό των γεγονότων, το οποίο υπάρχει συνήθως διαθέσιμο στα σημερινά (πληροφοριακά) συστήματα. Η εξόρυξη διαδικασιών περιλαμβάνει την (αυτόματη) ανακάλυψη διαδικασιών (δηλ. την εξαγωγή μοντέλων μιας διαδικασίας από καταγραφές γεγονότων), τον έλεγχο συμμόρφωσης (δηλ. τον εντοπισμό και την

παρακολούθηση των αποκλίσεων μεταξύ πρότυπου μοντέλου διαδικασίας και ιστορικού γεγονότων), την εξόρυξη κοινωνικών δικτύων - δικτύων οργανισμού, την αυτόματη δημιουργία και προσομοίωση μοντέλων, την επέκταση των μοντέλων, την επιδιόρθωση των μοντέλων, την πρόγνωση περιστατικών και τις

## Περιεχόμενα:

|                                   |    |
|-----------------------------------|----|
| Εξόρυξη Διαδικασιών - Τεχνολογική |    |
| Στάθμιση                          | 3  |
| Κατευθυντήριες Οδηγίες            | 6  |
| Προκλήσεις                        | 11 |
| Επίλογος                          | 16 |
| Λεξικό Όρων                       | 17 |

Οι τεχνικές εξόρυξης διαδικασιών (Process Mining) έχουν τη δυνατότητα να εξάγουν γνώση από αρχεία καταγραφών γεγονότων (event logs), τα οποία είναι συνήθως διαθέσιμα στα σημερινά πληροφοριακά συστήματα. Οι τεχνικές αυτές αποτελούν τα μέσα με τα οποία είναι δυνατόν να ανακαλυφθούν, να εποπτευτούν, ακόμη και να βελτιωθούν οι διαδικασίες σε διάφορα πεδία. Δύο είναι οι κύριοι παράγοντες που συντελούν στο ολοένα αυξανόμενο ενδιαφέρον για την εξόρυξη διαδικασιών. Αφενός, ολοένα και περισσότερα γεγονότα καταγράφονται, παρέχοντας με αυτόν τον τρόπο ένα λεπτομερές ιστορικό των διαδικασιών. Αφετέρου, σε ένα ανταγωνιστικό και ραγδαία μεταβαλλόμενο περιβάλλον, προκύπτει η ανάγκη για βελτίωση και υποστήριξη των επιχειρηματικών διαδικασιών. Η παρούσα διακήρυξη είναι δημιουργία της Ομάδας Κρούσης Εξόρυξης Διαδικασιών της ΙΕΕΕ και έχει ως σκοπό την προώθηση του ζητήματος της εξόρυξης διαδικασιών. Πλέον τούτου, δια του καθορισμού ενός συνόλου κατευθυντήριων οδηγιών και μιας λίστας ερευνητικών προκλήσεων, η διακήρυξη ευελπιστεί να διατελέσει ως ένα χρήσιμο εγχειρίδιο για προγραμματιστές, ερευνητές, συμβούλους, μάντζερς αλλά και απλούς χρήστες. Ο στόχος είναι να συμβάλλει στην ωρίμανση της εξόρυξης διαδικασιών ως ένα εργαλείο συνεισφοράς στον (ανα)σχεδιασμό, στον έλεγχο και στην υποστήριξη των επιχειρηματικών διαδικασιών των οργανισμών.



επιστήμης αποτελούν η συνεχής βελτίωση διαδικασιών (Continuous Process Improvement), η βελτίωση επιχειρηματικών διαδικασιών (Business Process Improvement), η διοίκηση ολικής ποιότητας (Total Quality Management) και η 6σ (Six Sigma). Το κοινό ανάμεσα σε όλες τις προαναφερθείσες προσεγγίσεις είναι πως επικεντρώνουν στις διαδικασίες και στις δυνατότητες βελτίωσης τους. Η εξόρυξη διαδικασιών καθαυτή είναι μία τεχνολογία που διευρύνει προσεγγίσεις όπως οι παραπάνω.

Ενώ τα εργαλεία επιχειρηματικής ευφυΐας και οι πρακτικές διοίκησης όπως η 6σ και η Δ.Ο.Π. στοχεύουν στη βελτίωση των επιχειρησιακών επιδόσεων, π.χ. μείωση των χρόνων και των σφαλμάτων παραγωγής, οι οργανισμοί δίνουν πλέον μεγάλη έμφαση και στην εταιρική διακυβέρνηση, στη διαχείριση των κινδύνων και στη συμμόρφωση με τα πρότυπα. Κανονισμοί όπως αυτοί της Πράξης Sarbanes-Oxley και η συμφωνία της Βασιλείας II σκιαγραφούν την προσοχή στα ζητήματα συμμόρφωσης. Οι τεχνικές εξόρυξης διαδικασιών, εκτός των άλλων, αποτελούν ένα μέσο για έναν πλέον ενδελεχή έλεγχο συμμόρφωσης και για την εξακρίβωση της εγκυρότητας και της αξιοπιστίας των πληροφοριών σχετικά με τις κεντρικές διαδικασίες ενός οργανισμού.

Κατά την περασμένη δεκαετία, τα δεδομένα που αφορούν στις καταχωρήσεις των γεγονότων γίνανε περισσότερο διαθέσιμα και προσβάσιμα, ενώ ωρίμασαν και οι τεχνικές εξόρυξης διαδικασιών. Παράλληλα, όπως προαναφέρθηκε, οι τάσεις της διοίκησης προς τη βελτίωση των διαδικασιών (π.χ. 6σ, Δ.Ο.Π., κτλ) και τη συμμόρφωση (Sarbanes-Oxley, Βασιλεία) μπορούν να επωφεληθούν από την εξόρυξη διαδικασιών. Αποτελεί ευτυχή συγκυρία το γεγονός της εφαρμογής αλγορίθμων της εξόρυξης διαδικασιών τόσο σε ακαδημαϊκά όσο και σε εμπορικά συστήματα. Σήμερα, λειτουργεί ένα σύνολο δραστηρίων ερευνητών που ασχολούνται με την εξόρυξη διαδικασιών και το ζήτημα έχει αναδυθεί ως ένα από τα «καυτά» ζητήματα στο χώρο της διοίκησης επιχειρηματικών διαδικασιών. Επιπρόσθετα, διακρίνεται ένα έντονο ενδιαφέρον της βιομηχανίας για το ζήτημα της εξόρυξης διαδικασιών. Ολοένα και περισσότεροι κατασκευαστές και έμποροι λογισμικού προσθέτουν λειτουργίες της εξόρυξης διαδικασιών στα προϊόντα τους. Παραδείγματα λογισμικών προϊόντων με δυνατότητες για εξόρυξη διαδικασιών αποτελούν: ARIS Process Performance Manager (Software AG), Comprehend (Open Connect), Discovery

συστάσεις βάση ιστορικών στοιχείων.

Η εξόρυξη διαδικασιών επιτελεί τη σημαντική διαδικασία της γεφύρωσης μεταξύ της εξόρυξης δεδομένων και

### Συγκεκριμένοι στόχοι της ομάδας κρούσης:

- 1) Να ενημερώσει και να κρατά ενήμερους τους χρήστες, προγραμματιστές, συμβούλους, μάντζερς και τους ερευνητές για την τεχνολογική στάθμη στο πεδίο της εξόρυξης διαδικασιών
- 2) Να προωθήσει τη χρήση των τεχνικών και των εργαλείων εξόρυξης διαδικασιών και να παρακινήσει νέες εφαρμογές τους
- 3) Να διαδραματίσει έναν ρόλο στις προσπάθειες τυποποίησης των δεδομένων καταγραφής γεγονότων
- 4) Να διοργανώνει σεμινάρια, ημερίδες, συνέδρια και
- 5) Να δημοσιεύει άρθρα, βιβλία, ειδικές εκδόσεις περιοδικών και πολυμεσικά αρχεία

της μοντελοποίησης και ανάλυσης επιχειρηματικών διαδικασιών. Κάτω από τη σκεπή της επιχειρηματικής ευφυΐας, διατυπώθηκε μία πληθώρα τεχνικών όρων για να περιγράψουν τα παρόλα αυτά απλοϊκά εργαλεία πανοπτικών ταμπλό (Dashboard) και εκτύπωσης αναφορών. Η εποπτεία επιχειρηματικής δραστηριότητας (Business Activity Monitoring) αναφέρεται σε τεχνολογίες που επιτρέπουν την παρακολούθηση των επιχειρηματικών διαδικασιών σε πραγματικό χρόνο. Η σύνθετη επεξεργασία γεγονότων (Complex Event Processing) αναφέρεται στις τεχνολογίες που επεξεργάζονται πλήθος γεγονότων και τα χρησιμοποιούν προκειμένου να επιβλέψουν, να καθοδηγήσουν ή ακόμη και να βελτιστοποιήσουν επιχειρηματικές πτυχές σε πραγματικό χρόνο. Η εταιρική διοίκηση επιδόσεων (Corporate Performance Management) είναι ένας ακόμα τεχνικός όρος που χρησιμοποιείται για τη μέτρηση της επίδοσης μιας διαδικασίας ή ενός οργανισμού. Πρόσθετοι όροι σχετικών προσεγγίσεων της διοικητικής

Η ομάδα κρούσης ιδρύθηκε το 2009 στο πλαίσιο της τεχνικής επιτροπής εξόρυξης δεδομένων (DMTC) της κοινότητας Computational Intelligence Society της IEEE. Στην παρούσα σύνθεση της ομάδας κρούσης συμμετέχουν εταιρείες λογισμικού (π.χ. Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Business nescscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), εταιρείες συμβούλων (π.χ. ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chili, BWI Systeme GmbH, Excellentia BPM, Rabobank), και ερευνητικές δομές (π.χ. TU/e, University of Padua, Universitat Politècnica de Catalunya, New Mexico State University, IST - Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universität zu Berlin, Queensland University of Technology, Vienna University of Economics and Business, Stevens

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

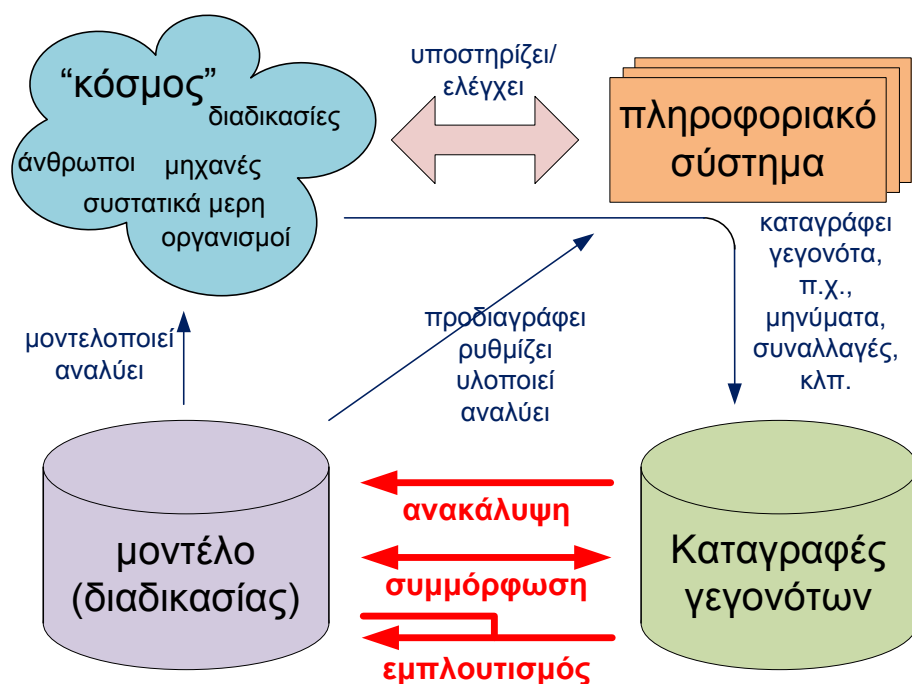


Από την ίδρυση της, το 2009, η ομάδα έχει πραγματοποιήσει διάφορες δράσεις για την επίτευξη των παραπάνω στόχων. Για παράδειγμα, (συν-) διοργανώθηκαν αρκετές συναντήσεις - εργαστήρια στα πλαίσια συνεδρίων όπως π.χ. τα εργαστήρια στα συνέδρια Ευφυΐας Επιχειρηματικών Διαδικασιών (BPI'09, BPI'10, και BPI'11) και ειδικές σύνοδοι σε κύρια συνέδρια της IEEE (π.χ. CIDM'11). Πραγματοποιήθηκαν επίσης αρκετές δραστηριότητες διάχυσης της γνώσης όπως σεμινάρια (π.χ. WCCI'10 και PMPM'09), θερινά σχολεία (ESSCaSS'09, ACPN'10, CICH'10, κτλ), δημοσίευση πολυμεσικών αρχείων

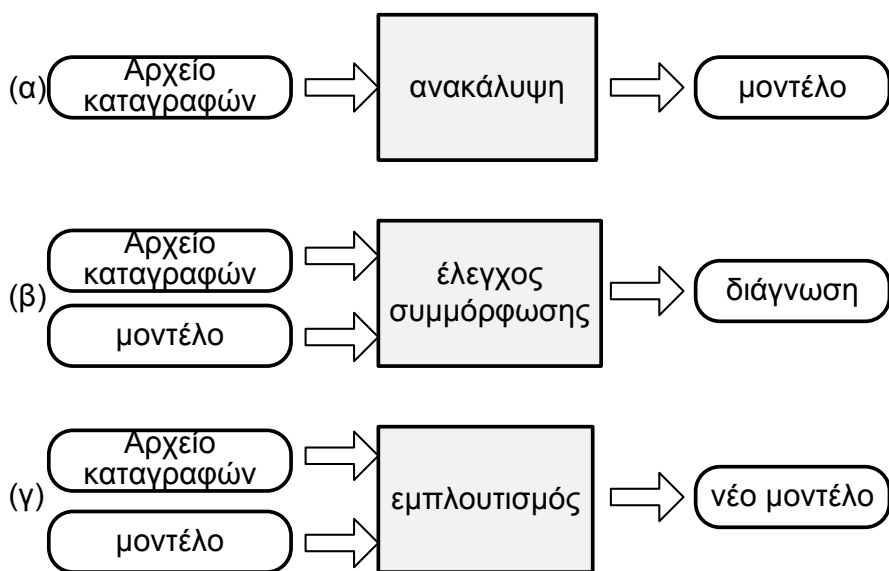
[www.processmining.org](http://www.processmining.org)), καθώς και αρκετές δημοσιεύσεις, ανάμεσα στις οποίες το πρώτο βιβλίο για την εξόρυξη διαδικασιών, πρόσφατα δημοσιευμένο από τον εκδοτικό οίκο Springer. Η ομάδα κρούσης (συν-) διοργάνωσε επίσης και τον πρώτο διαγωνισμό Ευφυΐας Επιχειρηματικών Διαδικασιών (BPIC'11): Πρόκειται για ένα διαγωνισμό στον οποίο οι συμμετέχοντες καλούνται να εξάγουν αξιοποιήσιμη γνώση από ένα ευμεγέθους και πολύπλοκο ιστορικό γεγονότων. Το 2010 επίσης, η ομάδα κρούσης δημιούργησε το πρότυπο XES ([www.xes-standard.org](http://www.xes-standard.org)), ένα πρότυπο σχήμα καταγραφής, το οποίο είναι επεκτάσιμο και υποστηρίζεται από τη βιβλιοθήκη OpenXES ([www.openxes.org](http://www.openxes.org)) και εργαλεία όπως τα ProM, XESame, Nitro, κτλ. Ο ενδιαφερόμενος αναγνώστης προτρέπεται να επισκεφτεί τον ιστοχώρο [www.win.tue.nl/ieeetfpm/](http://www.win.tue.nl/ieeetfpm/) για περισσότερες πληροφορίες σχετικά με τις δράσεις της ομάδας κρούσης.

Οι επεκτεινόμενες δυνατότητες των πληροφοριακών συστημάτων καθώς και άλλων συστημάτων που στηρίζονται σε υπολογιστικούς πόρους, διατρέχονται από το νόμο του Moore. Ο Gordon Moore, συνιδρυτής της Intel, προέβλεψε το 1965 πως ο αριθμός των τρανζίστορ στα ολοκληρωμένα κυκλώματα θα διπλασιάζεται κάθε χρόνο. Πράγματι, τα τελευταία πενήντα χρόνια ο ρυθμός ανάπτυξης τους υπήρξε όντως εκθετικός (αν και σε μικρότερο βαθμό). Αυτές οι εξελίξεις συντέλεσαν στη θεαματική αύξηση του αποκαλούμενου «ψηφιακού σύμπαντος» (δηλ. όλων των δεδομένων που αποθηκεύονται ή συναλλάσσονται ψηφιακά). Επιπλέον, το ψηφιακό και το πραγματικό σύμπαν τείνουν να ευθυγραμμίζονται ολοένα και περισσότερο.

Ακριβώς αυτή η ανάπτυξη ενός



**Εικόνα 2: Οριοθέτηση των τριών κύριων λειτουργιών της εξόρυξης διαδικασιών: α) Ανακάλυψη, β) Έλεγχος συμμόρφωσης, και γ) Εμπλουτισμός**



**Εικόνα 3: Οι τρεις λειτουργίες της εξόρυξης διαδικασιών σε όρους δεδομένων εισόδου / εξόδου: α) ανακάλυψη, β) Έλεγχος συμμόρφωσης, και γ) Εμπλουτισμός**

ψηφιακού σύμπαντος, καλά ευθυγραμμισμένου με τις διαδικασίες των οργανισμών, καθιστά δυνατή την καταγραφή και την ανάλυση γεγονότων. Τα γεγονότα μπορούν να ποικίλουν από την ανάληψη μετρητών από μία μηχανή αυτόματων συναλλαγών, τη ρύθμιση ενός ακτινοσκοπικού μηχανήματος από ένα γιατρό, την εκτύπωση μιας άδειας οδήγησης για έναν πολίτη, την υποβολή μιας φορολογικής δήλωσης, μέχρι και την εκτύπωση ενός ηλεκτρονικού εισιτηρίου ενός ταξιδιώτη. Η πρόκληση συνίσταται στην ερμηνεία τέτοιων δεδομένων γεγονότων με τέτοιο τρόπο ώστε να παράγονται χρήσιμα αποτελέσματα, όπως για παράδειγμα η αναγνώριση στενωπών, η πρόληψη προβλημάτων, η καταγραφή παραβιάσεων κανόνων, η σύσταση θεραπείας σφαλμάτων, ο εξορθολογισμός διαδικασιών. Τα παραπάνω σκιαγραφούν απόλυτα το στόχο της εξόρυξης διαδικασιών.

Σημείο εκκίνησης της εξόρυξης δεδομένων είναι η εγγραφή - καταχώρηση ενός γεγονότος (event log). Όλες οι τεχνικές εξόρυξης δεδομένων κάνουν τη βασική παραδοχή πως είναι δυνατόν να καταγραφεί μία αλληλουχία γεγονότων, κάθε ένα από τα οποία αναφέρεται σε μία εργασία (δηλ. ένα σαφώς προσδιορισμένο βήμα μιας διαδικασίας) και συνδέεται με μία συγκεκριμένη περίπτωση εφαρμογής (δηλ. με ένα στιγμιότυπο διαδικασίας). Οι καταχωρήσεις γεγονότων είναι δυνατόν να περιέχουν πρόσθετες πληροφορίες για τα γεγονότα. Για την ακρίβεια, όπου αυτό είναι δυνατόν, οι τεχνικές εξόρυξης διαδικασιών χρησιμοποιούν πρόσθετη πληροφορία όπως ο πόρος (ανθρώπινος ή υλικός -

π.χ. συσκευή) ο οποίος εκτελεί ή εκκινεί την εργασία, το χρονόσημο (timestamp) του γεγονότος, ή ακόμη και πληροφοριακά στοιχεία της εγγραφής (π.χ. το μέγεθος μιας παραγγελίας).

Όπως φαίνεται στην Εικόνα 2, οι καταχωρήσεις των γεγονότων μπορούν να χρησιμοποιηθούν για να υποστηρίξουν τις τρεις λειτουργίες της εξόρυξης διαδικασιών. Η πρώτη λειτουργία είναι αυτή της ανακάλυψης (discovery). Σε μία τεχνική ανακάλυψης παράγεται ένα μοντέλο διαδικασίας από ένα ιστορικό γεγονότων (ένα σύνολο καταχωρήσεων), χωρίς καμία εκ των προτέρων πληροφορία. Η ανακάλυψη διαδικασιών είναι η προεξέχουσα λειτουργία της εξόρυξης διαδικασιών. Πολλοί οργανισμοί εκπλήσονται όταν διαπιστώνουν πως υπάρχουν τεχνικές που είναι ικανές να ανακαλύψουν επιχειρηματικές τους διαδικασίες απλά και μόνο μέσα από ένα δείγμα των καταχωρήσεων των γεγονότων τους. Η δεύτερη λειτουργία είναι η συμμόρφωση (conformance). Αυτή η λειτουργία προβλέπει τη σύγκριση του υπάρχοντος μοντέλου της διαδικασίας με τις καταγραφές γεγονότων από την πραγματική εκτέλεση της. Ο έλεγχος συμμόρφωσης μπορεί να χρησιμοποιηθεί για να ελέγξει αν η πραγματικότητα (όπως αυτή αποδίδεται μέσα από τις καταγραφές των γεγονότων) συμβαδίζει με το μοντέλο και για το αντίστροφο. Διαφορετικά είδη μοντέλων μπορούν να μελετηθούν, καθώς ο έλεγχος συμμόρφωσης μπορεί να εφαρμοστεί σε μοντέλα ροής, σε μοντέλα δομής οργανισμών, σε συστήματα επιχειρηματικών κανόνων, σε κείμενους νόμους και διατάξεις, κτλ.

Η τρίτη λειτουργία της εξόρυξης διαδικασιών είναι ο εμπλουτισμός

## Τρία ΔΕΝ σε λαθεμένες αντιλήψεις για την Εξόρυξη Διαδικασιών:

1. Η εξόρυξη διαδικασιών **δεν** περιορίζεται στην ανακάλυψη της ροής των εργασιών.

Η ανακάλυψη μοντέλων διαδικασιών από καταγραφές γεγονότων ενθουσιάζει και παρακινεί τόσο τους επαγγελματίες όσο και τους ακαδημαϊκούς, και αντιμετωπίζεται συχνά ως το πιο συναρπαστικό κομμάτι της εξόρυξης διαδικασιών. Εν τούτοις, η εξόρυξη διαδικασιών δεν περιορίζεται μόνο στην ανακάλυψη της ροής των εργασιών. Αφενός, η λειτουργία της ανακάλυψης είναι μία από τις τρεις βασικές λειτουργίες της εξόρυξης διαδικασιών (ανακάλυψη, συμμόρφωση και εμπλουτισμός). Αφετέρου, δεν ανακαλύπτεται μόνο η ροή των εργασιών αλλά και άλλες διαστάσεις της διαδικασίας όπως η οργανωσιακή, η χρονική, η διάσταση της ανά περίπτωση προσέγγισης, και οι οποίες έχουν ιδιαίτερη σημασία.

2. Η εξόρυξη διαδικασιών **δεν** είναι απλά μία έκφανση της εξόρυξης δεδομένων

Η εξόρυξη διαδικασιών μπορεί να ιδωθεί ως ο «συνδετικός κρίκος» μεταξύ της εξόρυξης δεδομένων και της παραδοσιακής, βασισμένης σε μοντέλα, διαχείρισης διαδικασιών. Οι περισσότερες τεχνικές εξόρυξης δεδομένων δεν έχουν τον παραμικρό προσανατολισμό σε διαδικασίες. Τα μοντέλα διαδικασιών, τα οποία μπορούν να περιέχουν παραλληλίες και ταυτοχρονισμούς εργασιών, δεν είναι συγκρίσιμα με τις απλές δομές που προκύπτουν από της εξόρυξη δεδομένων, όπως τα δένδρα αποφάσεων και οι κανόνες συσχέτισης. Επομένως, απαιτούνται απόλυτα καινούργιες τεχνικές αναπαράστασης και νέοι αλγόριθμοι.

3. Η εξόρυξη διαδικασιών **δεν** περιορίζεται σε ανάλυση «νεκρού» χρόνου.

Οι τεχνικές εξόρυξης διαδικασιών εξαγουν γνώση από παρελθούσες καταχωρήσεις γεγονότων. Παρόλο που χρησιμοποιούνται δεδομένα για γεγονότα που έχουν ήδη συμβεί, τα αποτελέσματα μπορούν να εφαρμοστούν σε περιπτώσεις διαδικασιών που βρίσκονται σε εξέλιξη. Για παράδειγμα, ο χρόνος περάτωσης μία μερικώς εξυπηρετηθείσας παραγγελίας, μπορεί να προβλεφθεί χρησιμοποιώντας ένα μοντέλο διαδικασίας που ανακαλύφθηκε.

(enhancement). Η ιδέα αυτής της λειτουργίας είναι η επέκταση ή η βελτίωση ενός υπάρχοντος μοντέλου με χρήση πληροφοριών που προέρχονται από καταχωρήσεις γεγονότων. Ενώ ο έλεγχος συμμόρφωσης σταθμίζει την ταύτιση μεταξύ του μοντέλου της διαδικασίας και της πραγματικότητας, η τρίτη λειτουργία της εξόρυξης διαδικασιών στοχεύει στην αλλαγή και στην επέκταση του προκατασκευασμένου μοντέλου. Για παράδειγμα, η χρήση χρονοσήμων μπορεί να συμβάλλει στον εμπλουτισμό των μοντέλων με δυνατότητες όπως οι υποδείξεις στενωπών, ο υπολογισμός διεκπαιρωτικής ικανότητας, ή ακόμη η εκτίμηση επιπέδων εξυπηρέτησης..

Η Εικόνα 3 αναπαριστά τις τρεις λειτουργίες της εξόρυξης διαδικασιών σε όρους εισόδου / εξόδου. Οι τεχνικές της ανακάλυψης δέχονται καταχωρήσεις γεγονότων και παράγουν ένα μοντέλο. Το παραγόμενο μοντέλο είναι συνήθως ένα μοντέλο ροής της διαδικασίας (π.χ. ένα διάγραμμα Petri net, BPMN, EPC, ή UML), ωστόσο το μοντέλο μπορεί να περιέχει και άλλες διαστάσεις - οπτικές γωνίες (π.χ. κοινωνική δικτύωση). Οι τεχνικές του ελέγχου συμμόρφωσης χρειάζονται πλέον των καταγραφών γεγονότων και ένα πρότυπο μοντέλο ως δεδομένα εισόδου. Στην έξοδο, παραδίδουν διαγνωστικές αναφορές που περιγράφουν τις διαφορές και τις ομοιότητες μεταξύ του προτύπου μοντέλου και των καταγραφών. Οι τεχνικές του εμπλουτισμού χρειάζονται τα ίδια δεδομένα εισόδου όπως αυτές του ελέγχου συμμόρφωσης (μοντέλο και καταγραφές). Στην έξοδο τους παράγουν ένα νέο βελτιωμένο μοντέλο ή ένα μοντέλο στο οποίο έχουν γίνει σημαντικές επεκτάσεις.

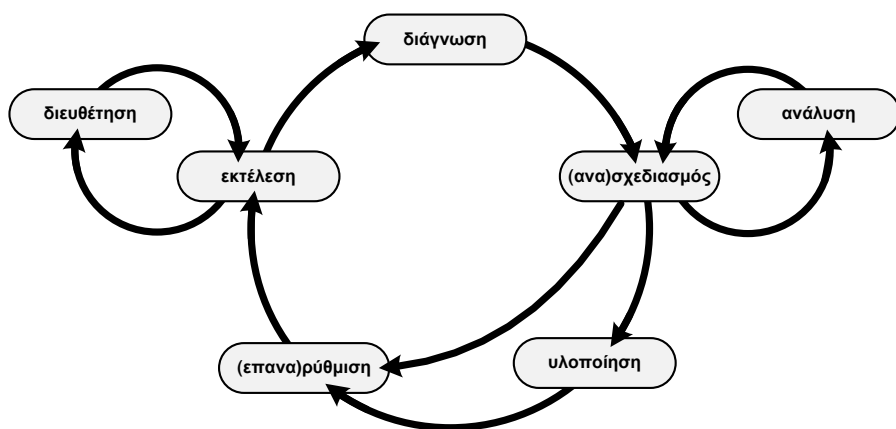
Η εξόρυξη διαδικασιών μπορεί να

καλύψει διαφορετικές διαστάσεις - οπτικές γωνίες μιας διαδικασίας. Η διάσταση του ελέγχου ροής εστιάζει στον έλεγχο της ροής, δηλαδή στη διάταξη των επιμέρους εργασιών. Ο στόχος της εξόρυξης όσο αφορά σε αυτήν τη διάσταση είναι να βρεθούν επαρκείς απεικονίσεις όλων των πιθανών διατάξεων. Τα αποτελέσματα αποδίδονται συνήθως με τη μορφή κάποιου διαγράμματος Petri Net, ή με κάποια άλλη γλώσσα μοντελοποίησης διαδικασιών (π.χ. EPC, BPMN, ή διαγράμματα δραστηριοτήτων UML). Η οργανωσιακή διάσταση εστιάζει στις πληροφορίες που υπάρχουν στις καταχωρήσεις και αφορούν στους πόρους του οργανισμού, δηλαδή ποιοί/ ποιά (άνθρωποι, συστήματα, θέσεις εργασίας ή τμήματα) εμπλέκονται στη διαδικασία και με ποιό τρόπο αυτοί σχετίζονται. Ο στόχος σε αυτήν την περίπτωση είναι να παρουσιαστεί μία δομή του οργανισμού είτε ταξινομώντας τους ανθρώπους και αντιστοιχίζοντας τους σε εργασιακούς ρόλους, είτε να αναδειχτεί το κοινωνικό δίκτυο του οργανισμού. Η διάσταση της ανά περίπτωση προσέγγισης (case perspective) εστιάζει στις ιδιότητες των περιπτώσεων. Προφανώς, κάθε περίπτωση διαδικασίας μπορεί να χαρακτηριστεί από τη διάταξη των επιμέρους εργασιών σε αυτήν ή από τα δρώντα σε αυτήν στοιχεία. Ωστόσο, μία περίπτωση διαδικασίας μπορεί να χαρακτηριστεί επίσης και από τις τιμές που παίρνουν σε αυτήν τα αντίστοιχα στοιχεία δεδομένων. Για παράδειγμα, αν μία περίπτωση αναφέρεται σε μία παραγγελία ανεφοδιασμού, θα ήταν χρήσιμο να αναδεικνύονται οι τιμές στα πεδία που αφορούν το όνομα του προμηθευτή ή το πλήθος των προϊόντων που παραγγέλθηκαν. Η διάσταση του χρόνου ασχολείται τόσο

με το χρονισμό όσο και με τη συχνότητα των γεγονότων. Όταν τα γεγονότα φέρουν και το χρονικό στίγμα τους, είναι δυνατόν να εντοπιστούν στενωποί, να καταμετρηθούν τα επίπεδα εξυπηρέτησης, να ελεγχθούν τα επίπεδα αξιοποίησης των πόρων, και να προβλεφθούν οι αναμενόμενοι χρόνοι περάτωσης διαδικασιών που βρίσκονται σε εξέλιξη.

Υπάρχουν ορισμένες εσφαλμένες αντιλήψεις που σχετίζονται με την εξόρυξη διαδικασιών. Κάποιες εταιρείες λογισμικού, αναλυτές και ερευνητές, περιορίζουν το εύρος της εξόρυξης διαδικασιών θεωρώντας την ως μία ειδική έκφανση της εξόρυξης δεδομένων η οποία έχει στόχο την ανακάλυψη διαδικασιών και η οποία μάλιστα μπορεί να εφαρμοστεί σε νεκρό χρόνο (offline). Ωστόσο, η πραγματικότητα είναι διαφορετική, ενώ οι συνήθεις παρανοήσεις εντοπίζονται στα τρία σημεία που αναφέρονται στο πλαίσιο της προηγούμενης σελίδας.

Προκειμένου να οριοθετηθεί η εξόρυξη διαδικασιών, χρησιμοποιούμε τον κύκλο ζωής της Διοίκησης Επιχειρηματικών Διαδικασιών (ΔΕΔ) που απεικονίζεται στην Εικόνα 4. Ο κύκλος ζωής της ΔΕΔ απεικονίζει επτά φάσεις μιας επιχειρηματικής διαδικασίας και τα πληροφοριακά συστήματα που αντιστοιχούν σε αυτές. Στη φάση του (ανά)σχεδιασμού δημιουργείται ένα νέο μοντέλο διαδικασίας ή υιοθετείται ένα υπάρχον μοντέλο διαδικασίας. Στη φάση της ανάλυσης εξετάζεται το υποψήφιο μοντέλο και οι εναλλακτικές του. Μετά τη φάση του (ανά)σχεδιασμού, είτε υλοποιείται το μοντέλο (φάση υλοποίησης), ή το υπάρχον σύστημα (επανα)ρυθμίζεται (φάση επαναρύθμισης). Στη φάση της εκτέλεσης εφαρμόζεται το μοντέλο. Κατά τη διάρκεια αυτής της φάσης πραγματοποιείται και η εποπτεία της διαδικασίας. Επιπρόσθετα, είναι δυνατόν να γίνουν μικρές τροποποιήσεις στο μοντέλο χωρίς να πραγματοποιηθεί ανασχεδιασμός (φάση διευθέτησης). Στη φάση της διάγνωσης, αναλύεται η διαδικασία που εφαρμόστηκε, ενώ τα αποτελέσματα της διάγνωσης μπορούν να πυροδοτήσουν μία νέα φάση ανασχεδιασμού της διαδικασίας. Η εξόρυξη διαδικασιών είναι ένα πολύτιμο εργαλείο για τις περισσότερες από τις φάσεις της Εικόνας 4. Προφανώς, η φάση της διάγνωσης μπορεί να επωφεληθεί από την εξόρυξη διαδικασιών. Ωστόσο, η συμβολή της εξόρυξης διαδικασιών δεν περιορίζεται σε αυτή τη φάση (διάγνωση). Για παράδειγμα, οι τεχνικές εξόρυξης διαδικασιών



**Εικόνα 4: Ο κύκλος ζωής της διοίκησης επιχειρηματικών διαδικασιών με τις διάφορες φάσεις μιας επιχειρηματικής διαδικασίας. Η εξόρυξη διαδικασιών εμπλέκεται (εν δυνάμει) σε όλες τις φάσεις (εκτός από τη φάση της υλοποίησης).**

## Κατευθυντήριες Οδηγίες:

**KO1:** Τα δεδομένα γεγονότων πρέπει να αντιμετωπίζονται ως διακεκριμένοι πολίτες.

**KO2:** Η εξαγωγή αρχείων καταγραφών πρέπει να καθοδηγείται από ερωτήματα

**KO3:** Η παραλληλία, η επιλογή και άλλες βασικές δομές ελέγχου ροής θα πρέπει να υποστηρίζονται

**KO4:** Τα γεγονότα πρέπει να συνδέονται με στοιχεία του μοντέλου

**KO5:** Τα μοντέλα πρέπει να αντιμετωπίζονται ως στοχοθετημένες αφαιρέσεις της πραγματικότητας

**KO6:** Η εξόρυξη διαδικασιών πρέπει να είναι μία διαρκής διαδικασία

Το παρόν κείμενο αποτελεί μετάφραση της πρωτότυπης διακήρυξης, η οποία δημοσιεύτηκε στα Αγγλικά στο "Business Process Management Workshops 2011, Lecture Notes in Business Information Processing, Vol. 99, Springer-Verlag, 2011. Η διακήρυξη έχει επίσης μεταφραστεί στα Ισπανικά, στα Ιταλικά, στα Γαλλικά, στα Κινέζικα, στα Πορτογαλικά, στα Κορεάτικα, στα Γερμανικά, στα Ιαπωνικά και στα Τούρκικα. Περισσότερες πληροφορίες μπορούν να βρεθούν στην ιστοσελίδα της Ομάδας Κρούσης: <http://www.win.tue.nl/ieetfpm/>

μπορούν να χρησιμοποιηθούν στη φάση της εκτέλεσης για λειτουργική υποστήριξη (operational support). Οι προβλέψεις και οι συστάσεις που βασίζονται σε μοντέλα που έχουν προκύψει από ιστορικά στοιχεία μπορούν να χρησιμοποιηθούν προς όφελος διαδικασιών που βρίσκονται σε εξέλιξη. Παρόμοιες δομές υποστήριξης αποφάσεων μπορούν να χρησιμοποιηθούν για να εισάγουν τροποποιήσεις στις διαδικασίες και για να καθοδηγήσουν τη διαδικασία επαναρύθμισης τους.

Ενώ η Εικόνα 4 αναπαριστά το συνολικό κύκλο ζωής της ΔΕΔ, η Εικόνα 5 εστιάζει σε συγκεκριμένες δραστηριότητες και συστατικά. Η Εικόνα 5 περιγράφει τα πιθανά στάδια σε ένα έργο εξόρυξης διαδικασιών. Κάθε έργο εξόρυξης διαδικασιών ξεκινά με έναν προγραμματισμό - σχεδιασμό και την αντίστοιχη αιτιολόγηση (Στάδιο 0). Μετά την εκκίνηση του έργου, τα δεδομένα για τα γεγονότα, τα μοντέλα, οι στόχοι και τα ερωτήματα που πρέπει να απαντηθούν πρέπει να εκμαieuτούν από τα συστήματα, τους εμπειρογνώμονες και τη διοίκηση της επιχείρησης (Στάδιο 1). Αυτό απαιτεί κατανόηση τόσο των διαθέσιμων πληροφοριών («Τί μπορεί να είναι χρήσιμο για την ανάλυση;»), όσο και κατανόηση του πεδίου εφαρμογής («Ποιές ερωτήσεις είναι σημαντικές;») και των δυνατών μορφών των αποτελεσμάτων που απεικονίζονται στην Εικόνα 5 (δηλ. ιστορικά δεδομένα, μοντέλα σχεδιασμένα με το χέρι, στόχοι και ερωτήματα). Στο Στάδιο 2, κατασκευάζεται το μοντέλο ροής και συνδέεται με τις καταγραφές των γεγονότων. Σε αυτό το σημείο μπορούν να χρησιμοποιηθούν οι τεχνικές ανακάλυψης διαδικασιών. Τα μοντέλα διαδικασιών που θα ανακαλυφθούν μπορεί να είναι σε θέση να απαντήσουν ορισμένα από τα ερωτήματα του προηγούμενου σταδίου, και να πυροδοτήσουν τις δράσεις ανασχεδιασμού ή τροποποιήσεων. Επιπλέον, οι καταγραφές των γεγονότων μπορούν να φιλτραριστούν ή να προσαρμοστούν κάνοντας χρήση του μοντέλου (π.χ. να εκτοπιστούν σπάνιες δραστηριότητες ή απομονωμένα (έκτοπα) σημεία, να εισαχθούν γεγονότα που απουσιάζουν). Κάποιες φορές, χρειάζεται σημαντική προσπάθεια για να μπορέσουν να συσχετιστούν τα γεγονότα που ανήκουν στο ίδιο στιγμιότυπο μιας διαδικασίας. Τα εναπομείναντα γεγονότα σχετίζονται με οντότητες του μοντέλου της διαδικασίας. Όταν η διαδικασία είναι σχετικά δομημένη, το μοντέλο ελέγχου της ροής μπορεί να εμπλουτιστεί και με άλλες διαστάσεις (π.χ. δεδομένα, χρονική διάταξη, πόροι) στο Στάδιο 3. Η αντιστοίχιση μεταξύ των καταγραφών των γεγονότων, και του μοντέλου που προκρίθηκε στο Στάδιο 2, χρησιμοποιείται για την επέκταση του μοντέλου (π.χ. τα χρονόσημα των εμπλεκόμενων γεγονότων χρησιμοποιούνται για να εκτιμηθούν οι χρόνοι αναμονής των επιμέρους δραστηριοτήτων). Μετά και από αυτό το στάδιο ίσως απαντιούνται κάποια πρόσθετα ερωτήματα ή μπορούν να πυροδοτηθούν πρόσθετες δράσεις ανασχεδιασμού και τροποποιήσεων. Τελικά, τα μοντέλα που κατασκευάστηκαν στο Στάδιο 3, μπορούν να χρησιμοποιηθούν και για τη λειτουργική υποστήριξη (Στάδιο 4). Η γνώση που εξάγεται από τα παρελθόντα δεδομένα των γεγονότων συνδυάζεται με πληροφορίες για τις διαδικασίες που βρίσκονται σε εξέλιξη. Κάτι τέτοιο μπορεί να είναι χρήσιμο για να παρέμβει κανείς σε αυτές, για να προβλέψει ή για να κάνει συστάσεις. Τα στάδια 3 και 4 μπορούν να πραγματοποιηθούν εφόσον η διαδικασία είναι επαρκώς σταθερή και δομημένη.

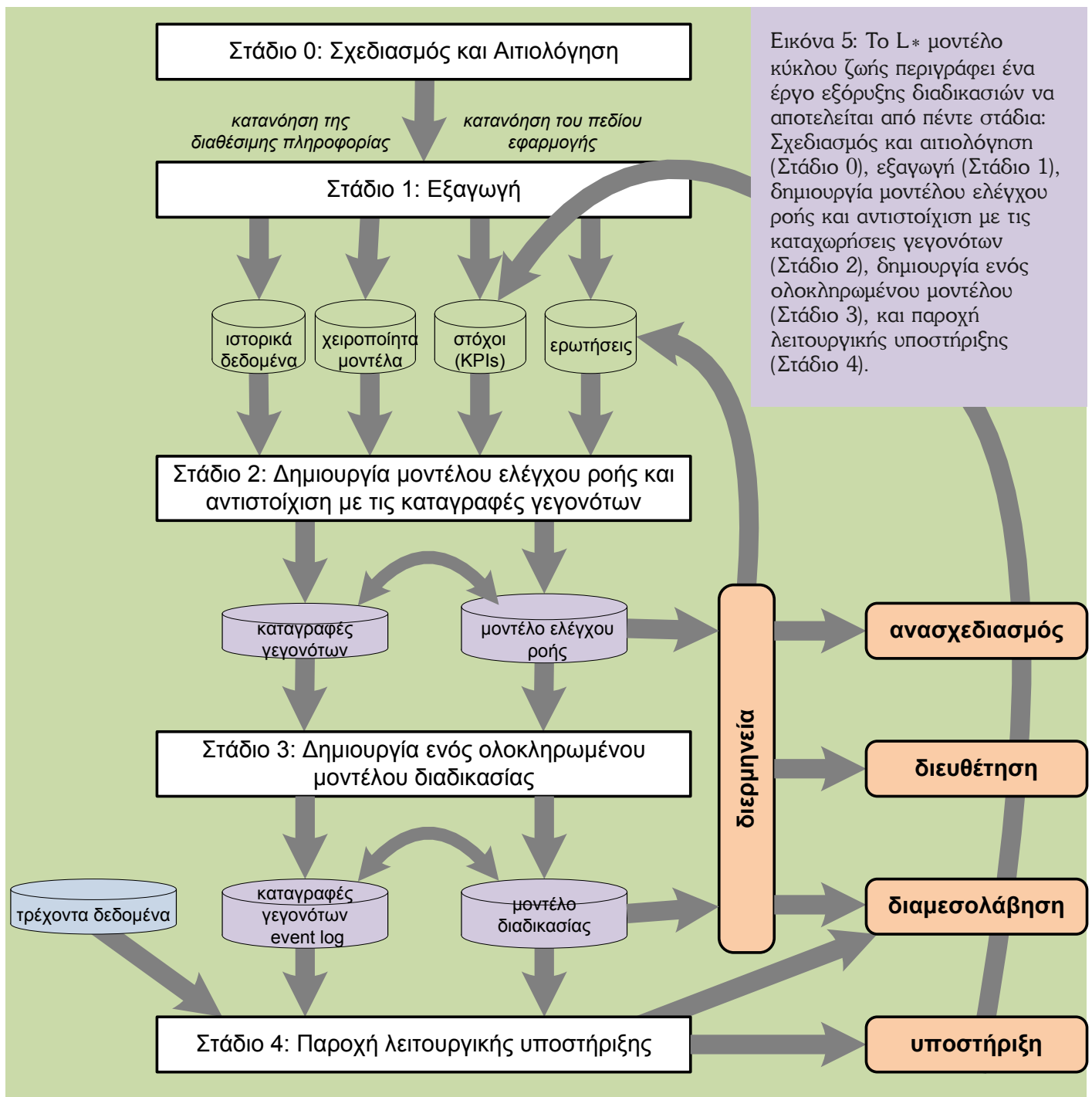
Αυτή τη στιγμή, υπάρχουν τεχνικές και εργαλεία που μπορούν να υποστηρίξουν όλα τα στάδια που εμφανίζονται στην Εικόνα 5. Ωστόσο, η εξόρυξη διαδικασιών είναι ένα σχετικά νέο πλαίσιο ιδεών και τα περισσότερα από τα διαθέσιμα εργαλεία είναι επί του παρόντος μάλλον ανώριμα. Επιπλέον, συχνά οι ενδοχόμενοι χρήστες δεν είναι ενήμεροι για τις δυνατότητες και τους περιορισμούς της εξόρυξης δεδομένων. Επομένως, η παρούσα διακήρυξη παρουσιάζει μερικές κατευθυντήριες οδηγίες (κανόνες) και προκλήσεις τόσο για τους χρήστες της εξόρυξης διαδικασιών, όσο και για τους ερευνητές και προγραμματιστές που ενδιαφέρονται να προωθήσουν την παρούσα κατάσταση.

### 3. Κατευθυντήριες Οδηγίες

Όπως με κάθε νέα τεχνολογία, υπάρχουν κάποια προφανή λάθη που μπορούν να γίνουν κατά την εφαρμογή της εξόρυξης διαδικασιών σε πραγματικές συνθήκες. Για το λόγο αυτό, απαριθμούμε έξι κατευθυντήριες οδηγίες (ΚΟ), έξι κανόνες που αποτρέπουν τους χρήστες / αναλυτές από το να διαπράττουν αυτά τα λάθη.

**KO1:** Τα δεδομένα γεγονότων πρέπει να αντιμετωπίζονται ως διακεκριμένοι πολίτες.

Σημείο εκκίνησης για κάθε δραστηριότητα της εξόρυξης διαδικασιών είναι τα γεγονότα που καταγράφονται. Αναφερόμαστε στις συλλογές των γεγονότων ως καταγραφές γεγονότων (event logs), ωστόσο αυτό δεν συνεπάγεται πως τα γεγονότα είναι αποθηκευμένα σε ειδικά αρχεία καταγραφών. Τα γεγονότα μπορεί να είναι αποθηκευμένα σε πίνακες βάσεων δεδομένων, σε καταγραφές μηνυμάτων ή συναλλαγών, σε αρχεία ηλεκτρονικής αλληλογραφίας και σε άλλες πηγές δεδομένων. Περισσότερο σημαντική από τη μορφή αποθήκευσης είναι η ποιότητα



Εικόνα 5: Το L\* μοντέλο κύκλου ζωής περιγράφει ένα έργο εξόρυξης διαδικασιών να αποτελείται από πέντε στάδια: Σχεδιασμός και αιτιολόγηση (Στάδιο 0), εξαγωγή (Στάδιο 1), δημιουργία μοντέλου ελέγχου ροής και αντιστοίχιση με τις καταχωρήσεις γεγονότων (Στάδιο 2), δημιουργία ενός ολοκληρωμένου μοντέλου (Στάδιο 3), και παροχή λειτουργικής υποστήριξης (Στάδιο 4).

τέτοιων καταχωρήσεων γεγονότων. Η ποιότητα ενός αποτελέσματος της εξόρυξης διαδικασιών εξαρτάται άρρηκτα από τα δεδομένα εισόδου. Επομένως, τα δεδομένα γεγονότων πρέπει να αντιμετωπίζονται ως διακεκριμένα στοιχεία των πληροφοριακών συστημάτων που υποστηρίζουν τις προς ανάλυση διαδικασίες. Δυστυχώς συχνά, οι καταχωρήσεις γεγονότων δεν είναι παρά ένα «παραπροϊόν» που χρησιμοποιείται είτε για αποσφαλμάτωση ή για αποτύπωση χαρακτηριστικών. Για παράδειγμα, οι ιατρικές συσκευές της Philips Healthcare καταγράφουν γεγονότα απλά επειδή οι κατασκευαστές του λογισμικού τους είχαν εισάγει εντολές

εκτύπωσης στον κώδικα. Παρόλο που υπάρχουν ορισμένες άτυπες οδηγίες για την προσθήκη τέτοιων εντολών στον κώδικα, απαιτείται μία πιο συστηματική προσέγγιση για να βελτιωθεί η ποιότητα των καταχωρήσεων των γεγονότων. Τα δεδομένα των γεγονότων θα πρέπει να θεωρούνται ως διακεκριμένοι πολίτες (αντί για πολίτες δεύτερης κατηγορίας).

Υπάρχουν αρκετά κριτήρια για να κριθεί η ποιότητα των δεδομένων των γεγονότων. Τα γεγονότα θα πρέπει να είναι αξιόπιστα, δηλ. θα πρέπει με ασφάλεια να υποθέτει κανείς πως τα γεγονότα που καταγράφηκαν έχουν όντως συμβεί και πως οι ιδιότητες αυτών των γεγονότων είναι οι σωστές.

Οι καταχωρήσεις των γεγονότων θα πρέπει να είναι πλήρεις, δηλ. δεδομένου ενός συγκεκριμένου πλαισίου, κανένα γεγονός δεν θα πρέπει να λείπει. Κάθε καταγεγραμμένο γεγονός πρέπει να έχει σαφώς ορισμένη σημασιολογία (semantics). Επιπλέον, τα δεδομένα των γεγονότων πρέπει να είναι ασφαλή, με την έννοια πως τα ζητήματα για το προσωπικό απόρρητο και την ασφάλεια έχουν αντιμετωπιστεί κατά την καταγραφή των γεγονότων. Για παράδειγμα, όσοι συμμετέχουν στη διαδικασία πρέπει να είναι ενήμεροι για το είδος των γεγονότων που καταγράφονται και τον τρόπο που αυτές οι καταγραφές θα χρησιμοποιηθούν.

| Επίπεδο | Περιγραφή                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Παραδείγματα                                                                                                                                                          |
|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ★★★★★   | Υψιστο επίπεδο: Το αρχείο καταγραφών είναι εξαιρετικής ποιότητας (δηλ. αξιόπιστο και πλήρες) και τα γεγονότα είναι σαφώς ορισμένα. Τα γεγονότα καταγράφονται με έναν αυτόματο, συστηματικό, αξιόπιστο και ασφαλές τρόπο. Τα ζητήματα του προσωπικού απόρρητου και η ασφάλεια αντιμετωπίζονται επαρκώς. Επιπλέον, τα γεγονότα που καταγράφονται (καθώς και όλα τα χαρακτηριστικά τους) έχουν ξεκάθαρη σημασιολογία. Αυτό υπονοεί την ύπαρξη μιας ή περισσότερων οντολογιών. Τα γεγονότα και τα χαρακτηριστικά τους αναφέρονται σε αυτές τις οντολογίες.                                                                                            | Σημασιολογικά αρχεία καταγραφών των συστημάτων διοίκησης επιχειρηματικών διαδικασιών.                                                                                 |
| ★★★★    | Τα γεγονότα καταγράφονται αυτόματα και με έναν συστηματικό και αξιόπιστο τρόπο, δηλ. οι καταγραφές είναι αξιόπιστες και πλήρεις. Σε αντίθεση με τα συστήματα του επιπέδου ★★★, έννοιες όπως στιγμιότυπο διαδικασίας και δραστηριότητα υποστηρίζονται ρητά και κατηγορηματικά.                                                                                                                                                                                                                                                                                                                                                                     | Αρχεία καταγραφών των παραδοσιακών συστημάτων διαχείρισης ροών εργασιών.                                                                                              |
| ★★★     | Τα γεγονότα καταγράφονται αυτόματα, αλλά δεν ακολουθείται καμία συστηματική προσέγγιση για την καταγραφή τους. Ωστόσο, αντίθετα με τις καταγραφές επιπέδου ★★, υπάρχει ένα εγγυημένο επίπεδο αντιστοίχισης των καταγραφών με τα πραγματικά γεγονότα (δηλ. τα αρχεία καταγραφών είναι αξιόπιστα αλλά όχι απαραίτητα και πλήρη). Για παράδειγμα, ας θεωρήσουμε τα γεγονότα που καταγράφονται από ένα σύστημα ERP. Παρόλο που τα γεγονότα μπορούν να εξαχθούν από ένα συνδυασμό πινάκων, η πληροφορία μπορεί να θεωρηθεί σωστή (π.χ. είναι ασφαλές να υποθέσουμε πως μία πληρωμή που έχει καταγραφεί έχει όντως πραγματοποιηθεί, και το αντίστροφο). | Πίνακες ERP συστημάτων, αρχεία καταγραφών CRM συστημάτων, αρχεία συναλλαγών συστημάτων μετάδοσης μηνυμάτων, αρχεία καταγραφών των συστημάτων υψηλής τεχνολογίας, κτλ. |
| ★★      | Τα αρχεία καταγραφών καταγράφονται αυτόματα, δηλ. ως παραπροϊόν ενός πληροφοριακού συστήματος. Η κάλυψη μπορεί να διαφέρει, δηλ. δεν ακολουθείται καμία συστηματική προσέγγιση που να καθορίζει ποια γεγονότα καταγράφονται. Επιπλέον, είναι πιθανόν κατά την καταγραφή να παρακάμπτεται το πληροφοριακό σύστημα. Επομένως, από τις καταγραφές μπορούν να λείπουν γεγονότα ή να μην έχουν καταγραφεί ορθώς.                                                                                                                                                                                                                                       | Αρχεία καταγραφών συστημάτων διαχείρισης εγγράφων, αρχεία καταγραφής σφαλμάτων ενσωματωμένων συστημάτων, λογιστικά φύλλα υπηρεσίας κτλ.                               |
| ★       | Κατώτατο επίπεδο: Τα αρχεία καταγραφών είναι κακής ποιότητας. Τα καταγεγραμμένα γεγονότα μπορούν να μην αντιστοιχούν στα πραγματικά, ενώ μπορεί να λείπουν και κάποια. Τέτοια χαρακτηριστικά απαντώνται συχνά στα αρχεία καταγραφών που έχουν δημιουργηθεί με το χέρι.                                                                                                                                                                                                                                                                                                                                                                            | Έγγραφα που έχουν δρομολογηθεί μέσα στην επιχείρηση, χειρόγραφα ιατρικών σημειώσεων, κτλ.                                                                             |

Πίνακας 1: Επίπεδα ωριμότητας αρχείων καταγραφών.

Ο Πίνακας 1 ορίζει πέντε επίπεδα ωριμότητας για τις καταγραφές των δεδομένων, από την άριστη ποιότητα (★★★★★) έως τη φτωχή ποιότητα (★). Για παράδειγμα, οι καταγραφές γεγονότων της Philips Healthcare βρίσκονται στο επίπεδο ★★★, δηλ. τα γεγονότα καταγράφονται μεν αυτόματα και η καταγεγραμμένη συμπεριφορά προσομοιάζει την πραγματικότητα, αλλά δεν υπάρχει καμία συστηματική προσέγγιση για την απόδοση σημασιολογίας στα γεγονότα και για την διασφάλιση ενός συγκεκριμένου ποσοστού κάλυψης. Οι τεχνικές εξόρυξης δεδομένων μπορούν

να εφαρμοστούν σε καταγραφές επιπέδων ποιότητας ★★★★★, ★★★★, και ★★★. Θεωρητικά, είναι δυνατόν να εφαρμοστούν τεχνικές εξόρυξης διαδικασιών σε καταγραφές γεγονότων επιπέδου ★★ ή ★. Ωστόσο, η ανάλυση τέτοιων καταγραφών είναι συνήθως προβληματική και τα αποτελέσματα δεν είναι αξιόπιστα. Στην πραγματικότητα, δεν έχει και πολύ νόημα να εφαρμοστεί η εξόρυξη διαδικασιών σε καταγραφές επιπέδου ★.

Προκειμένου να επωφεληθούν οι οργανισμοί από την εξόρυξη διαδικασιών, θα πρέπει να στοχεύουν σε καταγραφές γεγονότων όσο το

δυνατόν καλύτερης ποιότητας.

## ΚΟ2: Η εξαγωγή αρχείων καταγραφών πρέπει να καθοδηγείται από ερωτήματα

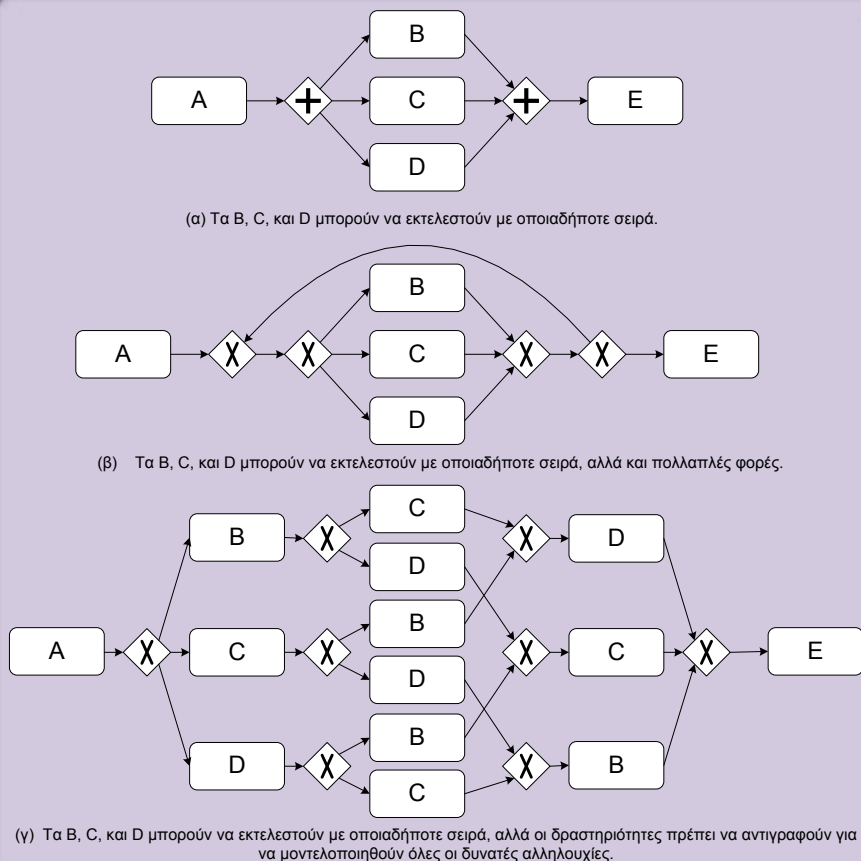
Όπως φαίνεται στην Εικόνα 5, είναι απαραίτητο η εξόρυξη διαδικασιών να καθοδηγείται από ερωτήματα. Χωρίς συγκεκριμένες ερωτήσεις, είναι εξαιρετικά δύσκολο να εξαχθούν δεδομένα γεγονότων μεστά νοήματος. Για παράδειγμα, ας θεωρήσουμε τις χιλιάδες των πινάκων που υπάρχουν στη βάση δεδομένων ενός συστήματος ERP όπως το SAP.

Χωρίς συγκεκριμένα ερωτήματα, είναι αδύνατον να επιλεγούν οι σχετικοί πίνακες για την εξαγωγή των δεδομένων.

Ένα μοντέλο μιας διαδικασίας, όπως αυτό της Εικόνας 1, περιγράφει τον κύκλο ζωής των ατομικών περιπτώσεων (δηλ. των στιγμιοτύπων της διαδικασίας) ενός συγκεκριμένου τύπου. Επομένως, πριν από την εφαρμογή μιας τεχνικής εξόρυξης διαδικασιών, θα πρέπει να επιλεγεί ο τύπος των περιπτώσεων που θα αναλυθούν. Η επιλογή αυτή πρέπει να καθοδηγείται από ερωτήματα που χρήζουν απάντησης, και αυτό μπορεί να μην είναι τετριμμένο. Για παράδειγμα, ας σκεφτούμε τη διαχείριση των παραγγελιών των πελατών. Κάθε παραγγελία ενός πελάτη μπορεί να αποτελείται από πολλαπλά στοιχεία, καθώς ο πελάτης είναι δυνατόν να παραγγείλει περισσότερα από ένα προϊόντα με μία παραγγελία. Μία παραγγελία μπορεί να έχει ως αποτέλεσμα πολλαπλές διανομές. Μία διανομή μπορεί να περιέχει ένα ή περισσότερα προϊόντα που παραγγέλθηκαν. Συνεπώς, υπάρχει μία συσχέτιση Πολλά-προς-Πολλά ανάμεσα στις παραγγελίες και τις διανομές και μία συσχέτιση Ένα-προς-Πολλά ανάμεσα στις παραγγελίες και τα στοιχεία τους. Διαθέτοντας μία βάση δεδομένων με τα γεγονότα που συνδέονται με τις παραγγελίες, τα στοιχεία παραγγελιών και τις διανομές τους, είναι δυνατόν να ανακαλυφθούν διαφορετικά μοντέλα διαδικασιών. Τα δεδομένα μπορούν να εξαχθούν με στόχο να περιγραφούν οι κύκλοι ζωής μεμονωμένων παραγγελιών. Ωστόσο, είναι επίσης πιθανό να εξαχθούν τα δεδομένα με στόχο να ανακαλυφθούν οι κύκλοι ζωής μεμονωμένων στοιχείων της παραγγελίας ή μεμονωμένων διανομών.

### ΚΟ3: Η παραλληλία, η επιλογή και άλλες βασικές δομές ελέγχου ροής θα πρέπει να υποστηρίζονται

Υπάρχει μία πληθώρα από γλώσσες μοντελοποίησης διαδικασιών (π.χ. BPMN, EPCs, Petri nets, BPEL, και διαγράμματα δραστηριοτήτων UML). Κάποιες από αυτές τις γλώσσες προσφέρουν πλήθος στοιχείων μοντελοποίησης (για παράδειγμα η σημειογραφία BPMN διαθέτει περισσότερα από 50 διακριτά γραφικά στοιχεία) ενώ κάποιες άλλες κινούνται στο εντελώς βασικό επίπεδο (για παράδειγμα τα δίκτυα Petri αποτελούνται μόνο από τρία διαφορετικά στοιχεία: θέσεις, μεταβάσεις και ακμές). Η περιγραφή του ελέγχου ροής είναι η ραχοκοκαλιά οποιουδήποτε μοντέλου διαδικασίας. Οι



**Εικόνα 6: Παράδειγμα που επιδεικνύει τα προβλήματα που προκύπτουν όταν η παράλληλη δρομολόγηση (δηλ. διασταυρώσεις – ενώσεις AND) δεν μπορεί να εκφραστεί άμεσα. Στο παράδειγμα, παράλληλες είναι μόνο τρεις δραστηριότητες (η B, η C, και η D). Μπορεί κανείς να φανταστεί τα ανάλογα αποτελέσματα όταν στα μοντέλα των διαδικασιών υπάρχουν 10 παράλληλες δραστηριότητες ( $2^{10} = 1024$  καταστάσεις και  $10! = 3,628,800$  πιθανές αλληλουχίες εκτέλεσης).**

βασικές δομές ροής εργασιών - γνωστές και ως μοτίβα (patterns) - που υποστηρίζονται από τις επικρατούσες γλώσσες είναι η ροή αλληλουχίας (sequence), η παράλληλη δρομολόγηση (διασταύρωση - ένωση AND), η αποκλειστική απόφαση (XOR-splits/joins), και οι επαναληπτικοί βρόχοι. Προφανώς αυτά τα μοτίβα θα πρέπει να υποστηρίζονται από τις τεχνικές της εξόρυξης διαδικασιών. Ωστόσο, κάποιες τεχνικές δεν είναι ικανές να αντιμετωπίσουν την παραλληλία και υποστηρίζουν μόνο τις αλυσίδες Markov/ συστήματα μετάβασης. Η Εικόνα 6 παρουσιάζει τα αποτελέσματα της χρήσης τεχνικών της εξόρυξης διαδικασιών που δεν μπορούν να ανακαλύψουν την παραλληλία. Ας υποθέσουμε μία καταγραφή γεγονότων  $L = \{ \langle A, B, C, D, E \rangle, \langle A, B, D, C, E \rangle, \langle A, C, B, D, E \rangle, \langle A, C, D, B, E \rangle, \langle A, D, B, C, E \rangle, \langle A, D, C, B, E \rangle \}$ . Η L περιέχει περιπτώσεις

που ξεκινούν με A και λήγουν με E. Οι δραστηριότητες B, C, και D μπορούν να λάβουν χώρα με οποιαδήποτε σειρά ανάμεσα στην A και την E. Το μοντέλο (σε σημειογραφία BPMN) της Εικόνας 6 (α) αναπαριστά την υποβόσκουσα διαδικασία χρησιμοποιώντας δύο πύλες KAI (AND). Υποθέτουμε τώρα πως η τεχνική εξόρυξης διαδικασιών δεν υποστηρίζει τις πύλες KAI. Σε αυτήν την περίπτωση τα υπόλοιπα δύο BPMN της Εικόνας 6 είναι πιθανοί υποψήφιοι. Το BPMN μοντέλο της Εικόνας 6(β) είναι συμπαγές αλλά επιτρέπει για πλεονάζουσες συμπεριφορές (π.χ. η περίπτωση  $\langle A, B, B, B, E \rangle$  είναι πιθανή σύμφωνα με το μοντέλο, όχι όμως και σύμφωνα με το αρχείο καταγραφής). Το BPMN μοντέλο της Εικόνας 6(γ) επιτρέπει τις περιπτώσεις της L αλλά κωδικοποιεί ρητώς όλες τις αλληλουχίες ροών, οπότε δεν αποτελεί μία συμπαγή αναπαράσταση του αρχείου καταγραφής. Το παράδειγμα υποδεικνύει πως στα μοντέλα

πραγματικών διαδικασιών, που διαθέτουν δεκάδες από πιθανά παράλληλες δραστηριότητες, τα εκτιμώμενα μοντέλα διακρίνονται από σοβαρή υποπροσαρμογή (δηλ. επιτρέπουν πλεονάζουσες συμπεριφορές) και/ή από εξαιρετική πολυπλοκότητα, όταν η παράλληλη δρομολόγηση δραστηριοτήτων δεν υποστηρίζεται. Όπως παρουσιάζεται στην Εικόνα 6, είναι σημαντικό να υποστηρίζονται τουλάχιστον τα βασικά μοτίβα ροών εργασιών (workflow patterns). Εκτός από τα βασικά μοτίβα που προαναφέρθηκαν, είναι επίσης ευκατὰ να υποστηρίζονται πύλες εναλλακτικής απόφασης (OR), επειδή αυτές παρέχουν μία συμπαγή αναπαράσταση των εναλλακτικών δυνατοτήτων απόφασης και των μερικών συγχρονισμών δραστηριοτήτων.

#### ΚΟ4: Τα γεγονότα πρέπει να συνδέονται με στοιχεία του μοντέλου

Όπως σημειώθηκε στο κεφάλαιο 2, η αντίληψη πως η εξόρυξη διαδικασιών περιορίζεται στην ανακάλυψη μοντέλων ροής των εργασιών οφείλεται σε παρανόηση. Όπως φαίνεται στην Εικόνα 1, τα μοντέλα των διαδικασιών που ανακαλύπτονται μπορούν να καλύπτουν διάφορες διαστάσεις (οργανωσιακή, διάσταση του χρόνου, κτλ.). Επιπλέον, η ανακάλυψη είναι μόνο μία από τις τρεις λειτουργίες της εξόρυξης διαδικασιών, όπως φαίνεται στην Εικόνα 3. Οι άλλες δύο λειτουργίες (έλεγχος συμμόρφωσης και εμπλουτισμός) στηρίζονται σε μεγάλο βαθμό στη σχέση μεταξύ των στοιχείων του μοντέλου και των γεγονότων στο αρχείο καταγραφών. Αυτή η σχέση μπορεί να χρησιμοποιηθεί για να «παιχτεί σε επανάληψη» το αρχείο καταγραφών πάνω στο μοντέλο. Η επανάληψη (replay) μπορεί να χρησιμοποιηθεί για να αποκαλύψει ανακολουθίες μεταξύ του αρχείου καταγραφών των γεγονότων και του μοντέλου, π.χ. κάποια γεγονότα που υπάρχουν στο αρχείο να μην μπορούν να αναπαραχθούν μέσα από το μοντέλο. Οι τεχνικές του ελέγχου συμμόρφωσης διαγνώνουν και ποσοτικοποιούν τέτοιες ανακολουθίες. Τα χρονόσημα στο αρχείο καταγραφών μπορούν να χρησιμοποιηθούν για να αναλύσουν τη χρονική συμπεριφορά κατά τη διάρκεια της επανάληψης. Οι διαφορές στο χρόνο μεταξύ συχνά συνδεδεμένων δραστηριοτήτων μπορούν να χρησιμοποιηθούν για να προστεθούν στο μοντέλο κάποιοι εκτιμώμενοι χρόνοι αναμονής. Τέτοια παραδείγματα δείχνουν πως η σχέση

μεταξύ των γεγονότων στο αρχείο καταγραφών και των στοιχείων του μοντέλου μπορεί να λειτουργήσει ως αφετηρία για διάφορες αναλύσεις.

Σε κάποιες περιπτώσεις δεν είναι καθόλου τετριμμένο να εδραιωθεί μία τέτοια σχέση. Για παράδειγμα, ένα γεγονός μπορεί να αναφέρεται σε δύο διαφορετικές δραστηριότητες ή μπορεί να μην είναι σαφές σε ποιά δραστηριότητα αναφέρεται. Τέτοιου είδους αμφισημίες πρέπει να απομακρυνθούν προκειμένου να ερμηνευτούν κατάλληλα τα αποτελέσματα της εξόρυξης. Πέρα από το πρόβλημα του να συσχετιστούν τα γεγονότα με τις δραστηριότητες, υπάρχει και το πρόβλημα να συσχετιστούν τα γεγονότα με τα στιγμιότυπα των διαδικασιών. Αυτό αναφέρεται συχνά ως συσχέτιση γεγονότων.

#### ΚΟ5: Τα μοντέλα πρέπει να αντιμετωπίζονται ως στοχοθετημένες αφαιρέσεις της πραγματικότητας

Τα μοντέλα που παράγονται από δεδομένα γεγονότων προσφέρουν όψεις της πραγματικότητας. Μία τέτοια όψη θα πρέπει να προσφέρει μία στοχοθετημένη αφαίρεση (abstraction) της συμπεριφοράς που καταγράφηκε στο αρχείο γεγονότων. Δεδομένου λοιπόν ενός αρχείου καταγραφών, είναι δυνατόν να υπάρχουν πολλαπλές όψεις που είναι χρήσιμες. Επιπρόσθετα, οι διάφοροι ενδιαφερόμενοι μπορεί να απαιτούν διαφορετικές όψεις. Πράγματι, τα μοντέλα που ανακαλύπτονται από τα αρχεία καταγραφών θα πρέπει να θεωρούνται ως «χάρτες» (όμοια με τους γεωγραφικούς χάρτες). Αυτή η κατευθυντήρια οδηγία παρέχει σημαντικές δυνατότητες αντίληψης, δύο από τις οποίες περιγράφονται στη συνέχεια.

Πρώτα από όλα, είναι σημαντικό να σημειωθεί πως δεν υπάρχει αυτό που θα αποκαλούσαμε «ο τέλειος χάρτης» για μία συγκεκριμένη γεωγραφική περιοχή. Ανάλογα με την προσδοκώμενη χρήση, υπάρχουν και διαφορετικοί χάρτες: Οδικόι χάρτες, ορειβατικοί χάρτες, ποδηλατικοί χάρτες, κτλ. Όλοι αυτοί οι χάρτες απεικονίζουν μία όψη της ίδιας πραγματικότητας και θα ήταν αφελές να υποθέτουμε πως είναι δυνατόν να υπάρξει ένας «τέλειος χάρτης». Το ίδιο ισχύει και για τα μοντέλα διαδικασιών: το κάθε μοντέλο θα πρέπει να τονίζει τα σημεία που έχουν ενδιαφέρον για έναν συγκεκριμένο είδος χρήστη. Τα μοντέλα που θα ανακαλυφθούν

μπορούν να επικεντρώνονται σε διαφορετικές διαστάσεις (ροή, πόροι, χρόνος, κόστος, κτλ.) και να τις αποτυπώνουν με διαφορετικά επίπεδα λεπτομέρειας και ακρίβειας, π.χ. ένα στέλεχος μπορεί να επιθυμεί να δει ένα γενικό (αδρό) και άτυπο μοντέλο της διαδικασίας που να επικεντρώνει στο κόστος ενώ ένας αναλυτής διαδικασιών είναι πιθανόν να επιθυμεί να δει ένα λεπτομερές μοντέλο της διαδικασίας που να τονίζει τις αποκλίσεις από την κανονική ροή. Αξίζει επίσης να σημειωθεί πως οι διάφοροι ενδιαφερόμενοι μπορούν να θέλουν να θεωρήσουν τη διαδικασία από διάφορα επίπεδα: Στρατηγικό επίπεδο (οι αποφάσεις σε αυτό το επίπεδο έχουν μακροπρόθεσμα αποτελέσματα και βασίζονται σε δεδομένα γεγονότων συσσωρευμένων επί μία μακρά περίοδο), τακτικό επίπεδο (οι αποφάσεις αυτού του επιπέδου έχουν μεσοπρόθεσμα αποτελέσματα και βασίζονται κυρίως σε πρόσφατα δεδομένα), και λειτουργικό επίπεδο (οι αποφάσεις αυτού του επιπέδου έχουν άμεσα αποτελέσματα και βασίζονται σε δεδομένα γεγονότων που σχετίζονται με τις διαδικασίες που βρίσκονται σε εξέλιξη).

Δεύτερον, όταν πρόκειται για την παραγωγή κατανοητών χαρτών, είναι χρήσιμο να υιοθετηθούν ιδέες από τη χαρτογραφία. Για παράδειγμα, οι οδοικοί χάρτες αναπαριστούν αφαιρετικά τους λιγότερο σημαντικούς δρόμους και πόλεις. Τα λιγότερο σημαντικά πράγματα είτε δεν αναπαριστώνται, ή ομαδοποιούνται σε συνολικά σχήματα (π.χ. οδοί και προάστια συγχωνεύονται στα σχήματα των πόλεων). Οι χαρτογράφοι όχι μόνο εξαλείφουν τις μη σχετικές λεπτομέρειες, αλλά επιπλέον χρησιμοποιούν και χρώματα για να επισημάνουν τα σημαντικά στοιχεία. Επιπρόσθετα, στα γραφικά στοιχεία αποδίδεται ένα συγκεκριμένο μέγεθος για να καταδείξει τη σημαντικότητα τους (π.χ. τα μεγέθη των γραμμών και των σημείων μπορεί να ποικίλουν). Οι γεωγραφικοί χάρτες έχουν επίσης μία ξεκάθαρη έννοια για τους άξονες συντεταγμένων, δηλ. η διαρρύθμιση ενός χάρτη δεν είναι αυθαίρετη αφού οι συντεταγμένες των στοιχείων έχουν συγκεκριμένο νόημα. Όλα αυτά βρίσκονται σε έντονη αντίθεση με τα επικρατούντα μοντέλα διαδικασιών τα οποία συνήθως δεν χρησιμοποιούν χρώμα, μεταβολές μεγέθους ή στοιχεία γραφικής διαρρύθμισης για να γίνουν πιο κατανοητά. Ωστόσο, οι ιδέες από τη χαρτογραφία μπορούν εύκολα να ενσωματωθούν στην κατασκευή των «χαρτών» των διαδικασιών που ανακαλύπτονται. Για παράδειγμα, το μέγεθος μιας δραστηριότητας μπορεί

να χρησιμοποιηθεί για να αντικατοπτρίσει τη συχνότητα της ή κάποια άλλη ιδιότητα δηλωτική της σημαντικότητας της (π.χ. το κόστος ή τη χρήση πόρων). Το πλάτος μιας ακμής μπορεί να αντικατοπτρίζει τη σημαντικότητα της αντίστοιχης σχέσης αιτιότητας, ή ο χρωματισμός των ακμών μπορεί να χρησιμοποιηθεί για να καταδείξει στενωπούς.

Οι παραπάνω παρατηρήσεις δείχνουν πως είναι σημαντικό να επιλέγεται η σωστή αναπαράσταση και να προσαρμόζεται ανάλογα με το προσδοκώμενο κοινό. Είναι σημαντικό τόσο για να οπτικοποιούνται τα αποτελέσματα στους τελικούς χρήστες όσο και για οδηγούνται οι αλγόριθμοι ανακάλυψης προς τα κατάλληλα μοντέλα (βλέπε επίσης και πρόκληση P5).

## ΚΟ6: Η εξόρυξη διαδικασιών πρέπει να είναι μία διαρκής διαδικασία

Η εξόρυξη διαδικασιών μπορεί να βοηθήσει την παραγωγή κατανοητών «χάρτων» που να είναι απευθείας συνδεδεμένοι με τα δεδομένα γεγονότων. Τόσο τα ιστορικά δεδομένα όσο και τα τρέχοντα δεδομένα για τα γεγονότα μπορούν να προβληθούν πάνω σε τέτοια μοντέλα. Επιπλέον, οι διαδικασίες μπορούν να μεταβάλλονται ενόσω αναλύονται. Με δεδομένη τη δυναμική φύση των διαδικασιών, δεν συνιστάται να αντιμετωπιστεί η εξόρυξη διαδικασιών σαν μία εφάπαξ δραστηριότητα. Ο στόχος δεν μπορεί να είναι η δημιουργία ενός σταθερού μοντέλου, αλλά το να εμψυχηθεί ζωή στα μοντέλα διαδικασιών ώστε οι χρήστες και οι αναλυτές να παρακινούνται να ανατρέχουν σε αυτά καθημερινά.

Αυτό μπορεί να συγκριθεί με τη χρήση των mashup που χρησιμοποιούν γεωγραφικές ετικέτες. Υπάρχουν χιλιάδες mashup που χρησιμοποιούν το Google Maps (π.χ. κυκλοφοριακές συμφορήσεις προβάλλονται πάνω στο χάρτη και ο χρήστης μπορεί να επιλέξει ένα συγκεκριμένο πρόβλημα για να δει λεπτομέρειες). Θα πρέπει να είναι επίσης δυνατόν να διεξάγεται η εξόρυξη διαδικασιών με βάση δεδομένα γεγονότων πραγματικού χρόνου. Συνεχίζοντας τις παρομοιώσεις με τους χάρτες, μπορούμε να σκεφτούμε πως τα γεγονότα διαθέτουν GPS και πως οι συντεταγμένες τους μπορούν να προβληθούν πάνω στο χάρτη σε πραγματικό χρόνο. Σε αναλογία με τα συστήματα πλοήγησης οχημάτων, τα εργαλεία εξόρυξης διαδικασιών μπορούν να βοηθήσουν τους χρήστες α) να πλοηγηθούν μέσα στις διαδικασίες, β) να προβάλλουν

δυναμική πληροφορία στους «χάρτες» της διαδικασίας (π.χ. να αναδείξουν κυκλοφοριακές συμφορήσεις μέσα στις επιχειρηματικές διαδικασίες), και γ) να παρέχουν προβλέψεις όσο αφορά στις περιπτώσεις που βρίσκονται σε εξέλιξη (π.χ. εκτιμώντας το χρόνο περάτωσης μίας περίπτωσης που έχει καθυστερήσει). Τα παραδείγματα αυτά επιδεικνύουν ότι είναι κρίμα για τα μοντέλα διαδικασιών να μην χρησιμοποιούνται περισσότερο ενεργά. Για αυτούς τους λόγους, η εξόρυξη διαδικασιών πρέπει να ιδωθεί σαν μία διαρκής διαδικασία που προσφέρει πληροφορίες ικανές να προκαλέσουν ενέργειες σε διάφορα χρονικά επίπεδα (λεπτά, ώρες, ημέρες, εβδομάδες ή μήνες).

## 4 Προκλήσεις

Η εξόρυξη διαδικασιών είναι ένα σημαντικό εργαλείο για τους σύγχρονους οργανισμούς οι οποίοι έχουν την ανάγκη να διαχειριστούν μη-τετρισμένες λειτουργικές διαδικασίες. Από τη μία, υπάρχει μία εντονότατη ανάπτυξη των δεδομένων γεγονότων. Από την άλλη, διαδικασίες και πληροφορίες πρέπει να είναι απόλυτα ευθυγραμμισμένες προκειμένου να καλύπτονται οι απαιτήσεις που συνδέονται με τον έλεγχο συμμόρφωσης, την αποδοτικότητα και την εξυπηρέτηση των πελατών. Παρά το εύρος των εφαρμογών της εξόρυξης διαδικασιών υπάρχουν ακόμη σημαντικές προκλήσεις οι οποίες πρέπει να αναληφθούν και οι οποίες δείχνουν πως η εξόρυξη διαδικασιών είναι ένα αναδυόμενο πεδίο. Στη συνέχεια του κεφαλαίου, απαριθμούνται ορισμένες από αυτές τις προκλήσεις. Η απαρίθμηση αυτή δεν στοχεύει στο να δημιουργήσει μία εξαντλητική λίστα και με τον καιρό νέες προκλήσεις μπορούν να προκύψουν ή κάποιες από τις υπάρχουσες προκλήσεις να εκλείψουν καθώς νέα επιτεύγματα θα σημειώνονται στο πεδίο.

### Π1: Εύρεση, συγχώνευση και εκκαθάριση δεδομένων γεγονότων

Χρειάζονται ακόμα σημαντικές προσπάθειες για να εξαχθούν δεδομένα γεγονότων κατάλληλα για την εξόρυξη διαδικασιών. Συνήθως, αρκετά εμπόδια πρέπει να ξεπεραστούν:

- Τα δεδομένα μπορεί να είναι διασκορπισμένα σε διάφορες πηγές. Οι διεσπαρμένες πληροφορίες πρέπει να συγχωνευτούν. Αυτό δημιουργεί προβλήματα όταν διαφορετικά αναγνωριστικά χρησιμοποιούνται

## Προκλήσεις:

**Π1:** Εύρεση, συγχώνευση και εκκαθάριση δεδομένων γεγονότων

**Π2:** Διαχείριση πολύπλοκων αρχείων καταγραφών με ποικίλα χαρακτηριστικά

**Π3:** Δημιουργία αντιπροσωπευτικών προγραμμάτων μέτρησης των επιδόσεων

**Π4:** Αντιμετωπίζοντας την Εννοιολογική Απόκλιση

**Π5:** Άμβλυνση της διαστρέβλωσης της αναπαράστασης που χρησιμοποιείται για την ανακάλυψη διαδικασιών

**Π6:** Ισορροπία μεταξύ κριτηρίων ποιότητας όπως προσαρμογή, απλότητα, ακρίβεια και γενίκευση

**Π7:** Διαεπιχειρησιακή εξόρυξη

**Π8:** Παροχή Λειτουργικής Υποστήριξης

**Π9:** Συνδυασμός της εξόρυξης διαδικασιών με άλλους τύπους ανάλυσης

**Π10:** Βελτίωση της Χρησιμότητας για τους μη-ειδικούς

**Π11:** Βελτίωση της ικανότητας κατανόησης από του μη-ειδικούς

στις διάφορες πηγές. Για παράδειγμα, ένα σύστημα μπορεί να χρησιμοποιεί το όνομα και την ημερομηνία γέννησης ενός ατόμου ενώ κάποιο άλλο σύστημα μπορεί να χρησιμοποιεί τον αριθμό κοινωνικής ασφάλισης (ΑΜΚΑ) του ατόμου.

- Τα δεδομένα των γεγονότων είναι συχνά «αντικειμενοστραφή» παρά «διαδικασιο-στραφή». Για παράδειγμα μεμονωμένα προϊόντα, παλέτες και κοντέινερ μπορούν να φέρουν ετικέτες RFID και τα καταγεγραμμένα γεγονότα να αναφέρονται σε αυτές τις ετικέτες. Ωστόσο, για να μπορέσει να γίνει η παρακολούθηση μιας συγκεκριμένης παραγγελίας ενός πελάτη, τέτοια αντικειμενοστραφή

γεγονότα πρέπει να συγχωνευτούν και να προ-επεξεργαστούν.

- Τα δεδομένα των γεγονότων μπορεί να είναι ελλιπή. Ένα κοινό πρόβλημα είναι πως τα γεγονότα δεν αναφέρονται ρητά σε στιγμιότυπα διαδικασιών. Συχνά, είναι πιθανόν να αντληθεί αυτή η πληροφορία αλλά θα χρειαστεί σημαντικός κόπος. Επιπλέον, οι πληροφορίες σχετικές με το χρόνο μπορούν να λείπουν από ορισμένα γεγονότα. Σε αυτήν την περίπτωση θα χρειαστεί να υπολογιστούν τα χρονόσημα μέσω παρεμβολής προκειμένου να αξιοποιηθεί η διαθέσιμη πληροφορία για τους χρόνους.
- Ένα αρχείο καταγραφών γεγονότων μπορεί να περιέχει ασυνήθιστα σημεία (έκτοπα), δηλ. ασυνήθιστη συμπεριφορά η οποία καλείται και θόρυβος. Πώς ορίζονται τα έκτοπα σημεία; Πώς είναι δυνατόν να ανακαλυφθούν τέτοια σημεία; Τέτοιου είδους ερωτήσεις πρέπει να απαντηθούν για να εκκαθαριστούν τα δεδομένα των γεγονότων.
- Οι καταγραφές μπορούν να περιέχουν γεγονότα σε διάφορα επίπεδα λεπτομέρειας. Στο αρχείο καταγραφών γεγονότων ενός νοσοκομείου μπορεί να βρεθούν γεγονότα που αντιστοιχούν σε απλές εξετάσεις αίματος ή σε πολύπλοκες χειρουργικές επεμβάσεις. Επιπρόσθετα, τα χρονόσημα μπορούν να διατυπώνονται με διαφορετικό επίπεδο λεπτομέρειας που μπορεί να ποικίλει από ακρίβεια χιλιοστού δευτερολέπτου (28-9-2011:Ωρα 11 Λεπτά 28 Δευτερόλεπτα 32 Χιλ. Δευτερολέπτου 342) μέχρι σε αδρές πληροφορίες ημερομηνίας (28-9-2011).
- Τα γεγονότα λαμβάνουν χώρα σε ένα συγκεκριμένο πλαίσιο (καιρικό, φόρτου εργασίας, ημέρα της εβδομάδας, κτλ.). Αυτό το πλαίσιο μπορεί να εξηγηθεί ορισμένα φαινόμενα, π.χ., ο χρόνος απόκρισης είναι μεγαλύτερος από τον συνηθισμένο εξαιτίας των εργασιών σε εξέλιξη ή της περιόδου των διακοπών. Για την ανάλυση, είναι επιθυμητό να σκιαγραφηθεί αυτό το πλαίσιο. Αυτό συνεπάγεται τη συγχώνευση των δεδομένων των γεγονότων με τα δεδομένα του γενικότερου πλαισίου. Σε αυτό το σημείο παραδοκεί η «κατάρα της πολυπλοκότητας» καθώς η ανάλυση δυσχεραίνει όσο προστίθεται πλήθος μεταβλητών.

Χρειάζονται καλύτερα εργαλεία και μεθοδολογίες για να αντιμετωπιστούν τα παραπάνω προβλήματα. Επιπλέον, όπως επισημάνθηκε νωρίτερα, οι οργανισμοί πρέπει να αντιμετωπίζουν τα αρχεία καταγραφών των γεγονότων

σαν διακεκριμένους πολίτες και όχι σαν κάποιο παραπροϊόν. Ο στόχος είναι να αποκτηθούν αρχεία καταγραφών επιπέδου (βλέπε Πίνακα 1). Σε αυτό το σημείο, οι εμπειρίες που έχουν αποκτηθεί από τις εφαρμογές data warehouse είναι χρήσιμες για να διασφαλιστεί η υψηλή ποιότητα των αρχείων καταγραφών. Για παράδειγμα, απλοί έλεγχοι κατά την εισαγωγή των δεδομένων μπορούν να βοηθήσουν σημαντικά στη μείωση του ποσοστού των εσφαλμένων δεδομένων.

## Π2: Διαχείριση πολύπλοκων αρχείων καταγραφών με ποικίλα χαρακτηριστικά

Τα αρχεία καταγραφών γεγονότων μπορούν να έχουν πολύ διαφορετικά χαρακτηριστικά. Κάποιες καταγραφές μπορούν να είναι υπερβολικά εκτενείς, καθιστώντας δύσκολο το χειρισμό τους, ενώ άλλες μπορούν να είναι τόσο μικρές ώστε να μην υπάρχουν επαρκή δεδομένα για τη διατύπωση αξιόπιστων συμπερασμάτων.

Σε κάποια πεδία καταγράφονται τεράστιες ποσότητες γεγονότων. Χρειάζονται επομένως επιπλέον προσπάθειες για τη βελτίωση των επιδόσεων και της δυνατότητας κλιμάκωσης. Για παράδειγμα, η ASML επιθεωρεί συνεχώς όλους τους σαρωτές πλακετών της. Αυτοί οι σαρωτές χρησιμοποιούνται από διάφορους οργανισμούς (π.χ. Samsung και Texas Instruments) για να παράγουν ολοκληρωμένα κυκλώματα (τσιπ). Περίπου το 70% των τσιπ παράγονται από τους σαρωτές της ASML. Τα υπάρχοντα εργαλεία συναντούν δυσκολίες όταν έρχονται αντιμέτωπα με τα petabyte των δεδομένων που συλλέγονται σε τέτοια πεδία. Εκτός από το πλήθος των γεγονότων που καταγράφονται, υπάρχουν και άλλα χαρακτηριστικά όπως ο μέσος αριθμός των γεγονότων ανά περίπτωση, η ομοιογένεια των περιπτώσεων, το πλήθος των μοναδικών γεγονότων και το πλήθος των μοναδικών διαδρομών. Ας θεωρήσουμε ένα αρχείο καταγραφής γεγονότων L1 με τα ακόλουθα χαρακτηριστικά: 1000 περιπτώσεις, 10 γεγονότα κατά μέσο όρο ανά περίπτωση, και μικρή διακύμανση (π.χ. αρκετές περιπτώσεις ακολουθούν τις ίδιες ή αρκετά παρόμοιες διαδρομές). Ένα άλλο αρχείο καταγραφών L2 περιέχει μόνο 100 περιπτώσεις, αλλά κατά μέσο όρο υπάρχουν 100 γεγονότα ανά περίπτωση και όλες οι περιπτώσεις ακολουθούν μια ξεχωριστή διαδρομή. Ξεκάθαρα, το L2 είναι πολύ πιο δύσκολο να αναλυθεί από το L1 παρόλο το παραπλήσιο μέγεθος των

δύο αρχείων (περίπου 10.000 γεγονότα).

Καθώς τα αρχεία καταγραφών περιέχουν μόνο ένα δείγμα της συμπεριφοράς, δεν πρέπει να θεωρούνται ως πλήρη. Οι τεχνικές της εξόρυξης διαδικασιών πρέπει να μπορούν να αντιμετωπίσουν την ελλιπή γνώση χρησιμοποιώντας μία «υπόθεση ανοικτού κόσμου»: το γεγονός πως κάτι δεν συνέβη δε συνεπάγεται πως δεν μπορεί αυτό να συμβεί. Αυτό καθιστά ιδιαίτερα δύσκολη την αντιμετώπιση μικρών αρχείων καταγραφών που διακρίνονται από μεγάλη μεταβλητότητα.

Όπως αναφέρθηκε και προηγουμένως, κάποια αρχεία καταγραφών περιέχουν γεγονότα με πολύ μεγάλο βαθμό λεπτομέρειας. Αυτά τα αρχεία τείνουν να είναι ιδιαίτερα εκτενή και τα ατομικά, χαμηλού επιπέδου γεγονότα δεν ενδιαφέρουν τους εμπλεκόμενους φορείς. Επομένως, χρειάζεται η συνάθροιση των γεγονότων χαμηλού επιπέδου σε υψηλότερου επιπέδου γεγονότα. Για παράδειγμα, όταν αναλύεται η διαδικασία διάγνωσης και θεραπείας ενός συγκεκριμένου γκρουπ ασθενών, ίσως δεν έχουν ενδιαφέρον οι ατομικές εξετάσεις που καταγράφηκαν στο πληροφοριακό σύστημα του εργαστηρίου του νοσοκομείου.

Προς το παρόν, οι οργανισμοί πρέπει να ακολουθούν μία προσέγγιση δοκιμής και σφάλματος για να ελέγξουν αν ένα αρχείο καταγραφών είναι κατάλληλο για εξόρυξη διαδικασιών. Επομένως, τα εργαλεία θα πρέπει να επιτρέπουν έναν σύντομο έλεγχο εφικτότητας για ένα συγκεκριμένο σύνολο δεδομένων. Ένας τέτοιος έλεγχος θα πρέπει να υποδεικνύει πιθανά προβλήματα στην επίδοση και να προειδοποιεί στην περίπτωση που εντοπίζονται αρχεία τα οποία είτε απέχουν πολύ από το να είναι πλήρη ή είναι πολύ λεπτομερειακά.

## Π3: Δημιουργία αντιπροσωπευτικών προγραμμάτων μέτρησης των επιδόσεων

Η εξόρυξη διαδικασιών είναι μία αναδυόμενη τεχνολογία. Αυτό εξηγεί γιατί λείπουν καλά προγράμματα μέτρησης και σύγκρισης των επιδόσεων. Για παράδειγμα, δεκάδες τεχνικές ανακάλυψης διαδικασιών υπάρχουν διαθέσιμες και διάφορες εταιρείες λογισμικού προσφέρουν διάφορα προϊόντα, αλλά δεν υπάρχει καμία γενική συναίνεση για την εκτίμηση της ποιότητας τους. Παρα τις μεγάλες διαφορές σε λειτουργικότητα

και επιδόσεις, είναι δύσκολο να συγκριθούν οι διάφορες τεχνικές και τα εργαλεία. Επομένως, πρέπει να αναπτυχθούν καλά προγράμματα μέτρησης και σύγκρισης των επιδόσεων (benchmark) που να περιέχουν πρότυπα παραδείγματα δεδομένων και αντιπροσωπευτικά κριτήρια ποιότητας.

Για τις κλασσικές τεχνικές της εξόρυξης δεδομένων, υπάρχουν πολλά καλά προγράμματα μέτρησης των επιδόσεων. Τα προγράμματα αυτά έχουν παρακινήσει τους παρόχους εργαλείων και τους ερευνητές να βελτιώνουν τις επιδόσεις των τεχνικών τους. Για την περίπτωση της εξόρυξης διαδικασιών, η πρόκληση είναι μεγαλύτερη. Για παράδειγμα, το σχεσιακό μοντέλο που παρουσιάστηκε από τον Codd το 1969 είναι απλό και υποστηρίζεται ευρέως. Σαν αποτέλεσμα, χρειάζεται πολύ μικρή προσπάθεια για τη μετατροπή / μεταφορά των δεδομένων από μία βάση δεδομένων σε μία άλλη και δεν υπάρχουν προβλήματα ερμηνείας. Το ανάλογο απλό μοντέλο για τις διαδικασίες απουσιάζει. Τα στάνταρ που έχουν προταθεί για τη μοντελοποίηση των διαδικασιών είναι πολύ πιο πολύπλοκα και ελάχιστες εταιρείες λογισμικού υποστηρίζουν ακριβώς το ίδιο σύνολο εννοιών. Οι διαδικασίες είναι απλά πιο σύνθετες από τα δεδομένα ενός πίνακα.

Παρόλα αυτά, είναι σημαντικό να δημιουργηθούν αντιπροσωπευτικά προγράμματα μέτρησης των επιδόσεων για την εξόρυξη διαδικασιών. Κάποια αρχική δουλειά είναι μάλιστα διαθέσιμη. Για παράδειγμα, υπάρχουν διάφορες μετρικές για την εκτίμηση της ποιότητας των αποτελεσμάτων της εξόρυξης διαδικασιών (προσαρμογή, απλότητα, ακρίβεια και γενίκευση). Επιπλέον, αρκετά αρχεία καταγραφών γεγονότων είναι διαθέσιμα στο κοινό (βλ. [www.processmining.org](http://www.processmining.org)). Παράδειγμα το αρχείο καταγραφών που χρησιμοποιήθηκε για τον πρώτο διαγωνισμό ευφυίας επιχειρηματικών διαδικασιών (BPIC '11) που διοργανώθηκε από την ομάδα κρούσης (βλ. doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc544).

Αφενός, πρέπει να υπάρχουν προγράμματα μέτρησης των επιδόσεων βασισμένα σε πραγματικά δεδομένα. Αφετέρου, πρέπει να δημιουργηθούν συνθέσεις δεδομένων που να διαθέτουν ιδιαίτερα χαρακτηριστικά. Τέτοιες συνθέσεις δεδομένων θα βοηθήσουν την ανάπτυξη τεχνικών εξόρυξης διαδικασιών οι οποίες θα είναι προσαρμοσμένες σε ελλιπή αρχεία γεγονότων, σε αρχεία καταγραφών με θόρυβο, ή σε ιδιαίτερους πληθυσμούς διαδικασιών. Εκτός από τη δημιουργία

αντιπροσωπευτικών προγραμμάτων μετρήσεων, υπάρχει επίσης η ανάγκη για μεγαλύτερη συναίνεση στα κριτήρια αξιολόγησης της ποιότητας των αποτελεσμάτων της εξόρυξης διαδικασιών (βλέπε επίσης πρόκληση P6). Επιπλέον, οι τεχνικές διασταυρωμένης επικύρωσης (cross validation) από το χώρο της εξόρυξης δεδομένων μπορούν να υιοθετηθούν για να αξιολογήσουν τα αποτελέσματα. Ας πάρουμε το παράδειγμα του ελέγχου k-fold. Είναι δυνατόν να διαχωριστεί το αρχείο καταγραφών σε k μέρη. k-1 μέρη μπορούν να χρησιμοποιηθούν για την εκμάθηση του μοντέλου της διαδικασίας ενώ τεχνικές ελέγχου συμμόρφωσης μπορούν να χρησιμοποιηθούν για να αξιολογήσουν το αποτέλεσμα στο μέρος που απόμεινε. Αυτό μπορεί να επαναληφθεί k φορές, και επομένως να προσφέρει βαθύτερη γνώση για την ποιότητα του μοντέλου.

## P4: Αντιμετωπίζοντας την Εννοιολογική Απόκλιση

Ο όρος «εννοιολογική απόκλιση» (concept drift) αναφέρεται στην κατάσταση κατά την οποία μία διαδικασία αλλάζει ενόσω αναλύεται. Για παράδειγμα, στο ξεκίνημα των καταγραφών ενός αρχείου γεγονότων δύο δραστηριότητες είναι παράλληλες, ενώ στη συνέχεια αυτές οι δραστηριότητες παρατηρούνται να λαμβάνουν χώρα εν σειρά. Οι διαδικασίες μπορούν να αλλάζουν εξαιτίας περιοδικών / εποχικών αλλαγών (π.χ. «η ζήτηση είναι αυξημένη το Δεκέμβριο» ή «το απόγευμα της Παρασκευής υπάρχουν διαθέσιμοι λιγότεροι υπάλληλοι») ή λόγω μεταβαλλόμενων συνθηκών (π.χ. «η αγορά γίνεται πιο ανταγωνιστική»). Τέτοιες μεταβολές επηρεάζουν τις διαδικασίες και είναι κρίσιμο να ανακαλυφθούν και να αναλυθούν. Η εννοιολογική απόκλιση σε μία διαδικασία μπορεί να ανακαλυφθεί μέσω της διαίρεσης του αρχείου καταγραφών σε μικρότερα αρχεία και της ανάλυσης των αποτυπωμάτων των μικρότερων αρχείων. Μία τέτοια ανάλυση «δεύτερου επιπέδου» απαιτεί πολύ περισσότερα δεδομένα για τα γεγονότα. Εντούτοις, λίγες διαδικασίες βρίσκονται σε σταθερή κατάσταση και η κατανόηση της εννοιολογικής απόκλισης είναι πρωτεύουσας σημασίας για τη διοίκηση των διαδικασιών. Επομένως, χρειάζεται επιπλέον έρευνα και εργαλεία υποστήριξη προκειμένου να αναλυθεί επαρκώς η εννοιολογική απόκλιση.

## P5: Άμβλυνση της διαστρέβλωσης της αναπαράστασης που χρησιμοποιείται για την ανακάλυψη διαδικασιών

Μία τεχνική ανακάλυψης διαδικασιών παράγει ένα μοντέλο χρησιμοποιώντας μία συγκεκριμένη γλώσσα ή σημειογραφία (π.χ. BPMN ή Δίκτυα Petri). Ωστόσο είναι σημαντικό να γίνει η διάκριση μεταξύ του γραφικού (οπτικού) αποτελέσματος και της αναπαράστασης που έχει χρησιμοποιηθεί κατά τη διάρκεια της ανακάλυψης. Η επιλογή μιας σημειογραφίας ως στόχο συχνά επιφέρει αρκετές έμμεσες παραδοχές: περιορισμός του χώρου αναζήτησης, διαδικασίες που δεν μπορούν να αναπαρασταθούν με τη δεδομένη σημειογραφία δεν μπορούν να ανακαλυφθούν. Αυτή η αποκαλούμενη «διαστρέβλωση της αναπαράστασης» (representational bias) που χρησιμοποιείται κατά τη διάρκεια της ανακάλυψης διαδικασιών πρέπει να είναι μία συνειδητή πράξη και δεν πρέπει να καθοδηγείται (μόνο) από την προτιμώμενη σημειογραφία.

Ας θεωρήσουμε το παράδειγμα της Εικόνας 6: το αν η γλώσσα-στόχος επιτρέπει ή όχι την παραλληλία μπορεί να επηρεάσει τόσο το οπτικό αποτέλεσμα του μοντέλου που ανακαλύφθηκε όσο και την κατηγορία των μοντέλων που θα εξεταστούν από τον αλγόριθμο. Αν η διαστρέβλωση της αναπαράστασης δεν επιτρέπει την παραλληλία (η περίπτωση της Εικόνας 6(α) δεν είναι δυνατή) και δεν επιτρέπει μία ονομασία να εμφανίζεται σε πολλαπλές δραστηριότητες (η περίπτωση της Εικόνας 6(γ) δεν είναι δυνατή), τότε μόνο προβληματικά μοντέλα όπως αυτό στην περίπτωση της Εικόνας 6(β) είναι δυνατά. Το παράδειγμα αυτό καταδεικνύει πως χρειάζεται μία πιο προσεκτική και εκλεπτυσμένη αντιμετώπιση της διαστρέβλωσης της αναπαράστασης.

## P6: Ισορροπία μεταξύ κριτηρίων ποιότητας όπως προσαρμογή, απλότητα, ακρίβεια και γενίκευση

Τα αρχεία καταγραφών γεγονότων απέχουν συχνά από το να είναι πλήρη, δηλ. περιέχουν μόνο ενδεικτικές συμπεριφορές. Τα μοντέλα διαδικασιών συνήθως αφήνουν έναν εκθετικό ή ακόμη και άπειρο (στην περίπτωση των επαναληπτικών βρόγχων) αριθμό διαφορετικών ιχνών. Επιπλέον, κάποια ίχνη μπορεί να έχουν σημαντικά μικρότερη πιθανότητα από κάποια άλλα. Επομένως, δεν είναι ρεαλιστικό να υποθέτουμε πως κάθε πιθανό ίχνος

βρίσκεται παρόν στο αρχείο καταγραφών. Για να γίνει πιο κατανοητό πως είναι άσκοπο να θεωρείται δεδομένη η πληρότητα σε ένα αρχείο καταγραφών, ας θεωρήσουμε μία διαδικασία που αποτελείται από 10 δραστηριότητες, οι οποίες μπορούν να εκτελεστούν παράλληλα, και το αντίστοιχο αρχείο καταγραφών το οποίο περιέχει πληροφορία για 10.000 περιπτώσεις. Ο συνολικός αριθμός των πιθανών εναλλακτικών του μοντέλου με 10 παράλληλες δραστηριότητες είναι  $10! = 3628800$ . Επομένως είναι απίθανο κάθε μία από αυτές τις εναλλακτικές να είναι παρούσα στο αρχείο καταγραφών καθώς υπάρχουν λιγότερες περιπτώσεις καταγεγραμμένες (10.000) από ότι πιθανά ίχνη (3.628.800). Ακόμη όμως και αν υπήρχαν εκατομμύρια περιπτώσεις στο αρχείο καταγραφών, θα ήταν εξαιρετικά απίθανο να είναι παρούσες όλες οι πιθανές εκδοχές. Μία πρόσθετη περιπλοκή είναι πως κάποιες εναλλακτικές είναι λιγότερο συχνές από κάποιες άλλες. Αυτό μπορεί να θεωρηθεί ένα είδος «θορύβου». Είναι απίθανο να κατασκευαστεί ένα λογικό μοντέλο για τέτοιες συμπεριφορές θορύβου. Το μοντέλο που θα ανακαλυφθεί πρέπει να γενικεύει αυτές τις περιπτώσεις. Προτιμάται οι συμπεριφορές χαμηλής συχνότητας να ερευνώνται μέσα από τον έλεγχο συμμόρφωσης.

Ο θόρυβος και η μη-πληρότητα κάνουν την ανακάλυψη διαδικασιών ένα δύσκολο πρόβλημα. Πράγματι, υπάρχουν τέσσερις ανταγωνιστικές διαστάσεις ποιότητας: α) η προσαρμογή (fitness), β) η απλότητα (simplicity), γ) η ακρίβεια (precision), και δ) η γενίκευση (generalization). Ένα μοντέλο με καλή προσαρμογή επιτρέπει τις περισσότερες από τις συμπεριφορές που απαντώνται στο αρχείο καταγραφών. Ένα μοντέλο έχει τέλεια προσαρμογή αν όλα τα ίχνη που υπάρχουν στο αρχείο καταγραφών μπορούν να αναπαραχθούν από το μοντέλο από την αρχή ως το τέλος. Όσο πιο απλό είναι το μοντέλο που μπορεί να εξηγήσει τη συμπεριφορά που παρατηρείται στο αρχείο καταγραφών, τόσο καλύτερο μοντέλο είναι. Η αρχή αυτή είναι γνωστή ως Λεπίδα του Όκαμ (Occam's Razor). Η προσαρμογή και η απλότητα από μόνες τους δεν επαρκούν για να κρίνουν την ποιότητα ενός μοντέλου διαδικασίας που ανακαλύπτεται. Για παράδειγμα, είναι πολύ εύκολο να κατασκευαστεί ένα εξαιρετικά απλό δίκτυο Petri (flower model) το οποίο είναι ικανό να αναπαράγει όλα τα ίχνη ενός αρχείου καταγραφών (αλλά και κάθε άλλου αρχείου που θα αναφερόταν στο ίδιο σύνολο δραστηριοτήτων). Ομοίως, είναι ανεπιθύμητο να προκύψει ένα μοντέλο

που επιτρέπει μόνο την ακριβή συμπεριφορά που περιέχεται στο αρχείο καταγραφών. Θυμίζουμε πως το αρχείο καταγραφών περιέχει μόνο ενδεικτικές συμπεριφορές και πως πολλά ίχνη που είναι πιθανά μπορεί να μην περιέχονται. Ένα μοντέλο είναι ακριβές αν δεν επιτρέπει για πλεονάζουσες συμπεριφορές. Προφανώς το flower model στερείται ακρίβειας. Ένα μοντέλο που δεν είναι ακριβές πάσχει από υποπροσαρμογή (underfitting). Υποπροσαρμογή καλείται το πρόβλημα σύμφωνα με το οποίο το μοντέλο γενικεύει υπέρ του δέοντος τις ενδεικτικές συμπεριφορές του αρχείου καταγραφών (δηλ. το μοντέλο επιτρέπει συμπεριφορές πολύ διαφορετικές από αυτές που απαντώνται στο αρχείο καταγραφών). Ένα μοντέλο θα πρέπει να γενικεύει και μην περιορίζει τις επιτρεπόμενες συμπεριφορές μόνο στα παραδείγματα του αρχείου καταγραφών. Ένα μοντέλο που δεν γενικεύει πάσχει από υπερπροσαρμογή (overfitting). Υπερπροσαρμογή καλείται το πρόβλημα σύμφωνα με το οποίο κατασκευάζεται ένα πολύ συγκεκριμένο μοντέλο ενώ είναι προφανές πως το αρχείο καταγραφών περιέχει μόνο ενδεικτικές συμπεριφορές (δηλ. το μοντέλο εξηγεί το συγκεκριμένο δείγμα του αρχείου, αλλά ένα άλλο δείγμα αρχείου από την ίδια διαδικασία μπορεί να παράγει ένα εντελώς διαφορετικό μοντέλο διαδικασίας).

Η πρόκληση συνίσταται στο να επιτευχθεί μία ισορροπία μεταξύ της προσαρμογής, της απλότητας, της ακρίβειας και της γενίκευσης του μοντέλου. Αυτός είναι ο λόγος για τον οποίο πολλές από τις πιο ισχυρές τεχνικές ανακάλυψης διαδικασιών παρέχουν διάφορες παραμέτρους. Προκειμένου να επιτευχθεί καλύτερη ισορροπία μεταξύ των τεσσάρων ανταγωνιστικών διαστάσεων ποιότητας, βελτιωμένοι αλγόριθμοι θα πρέπει να αναπτυχθούν. Επιπρόσθετα, κάθε παράμετρος θα πρέπει να είναι κατανοητή από τους τελικούς χρήστες.

## Π7: Διαεπιχειρησιακή εξόρυξη

Παραδοσιακά, η εξόρυξη διαδικασιών εφαρμόζεται μέσα στα πλαίσια ενός μόνου οργανισμού. Ωστόσο, καθώς οι εφοδιαστικές αλυσίδες γίνονται πιο ολοκληρωμένες και οι τεχνολογίες των υπηρεσιών εξελίσσονται και η υπολογιστική νέφους (cloud computing) διαδίδεται, προκύπτουν σενάρια κατά τα οποία είναι διαθέσιμα για ανάλυση αρχεία καταγραφών γεγονότων από πολλαπλούς οργανισμούς. Βασικά, υπάρχουν δύο πλαίσια για την

διαεπιχειρησιακή εξόρυξη διαδικασιών.

Πρώτα από όλα, ας θεωρήσουμε το συνεργατικό πλαίσιο κατά το οποίο διάφοροι οργανισμοί δουλεύουν από κοινού για να διαχειριστούν στιγμιότυπα διαδικασιών. Μία τέτοια διαεπιχειρησιακή διαδικασία θα μπορούσε να ιδωθεί σαν ένα παζλ, δηλ. μία συνολική διαδικασία που τεμαχίζεται σε επιμέρους κομμάτια τα οποία διανέμονται στους οργανισμούς, οι οποίοι με τη σειρά τους πρέπει να συνεργαστούν για να επιτύχουν την περάτωση των περιπτώσεων. Η ανάλυση του αρχείου καταγραφών ενός μόνου οργανισμού δεν επαρκεί. Για να ανακαλυφθεί ολόκληρη η διαδικασία, πρέπει να συγχωνευθούν τα αρχεία καταγραφών όλων των οργανισμών. Αυτό δεν είναι μία τετριμμένη εργασία καθώς τα γεγονότα θα πρέπει να συσχετιστούν ξεπερνώντας τα όρια των οργανισμών.

Δεύτερον, μπορούμε να θεωρήσουμε το πλαίσιο κατά το οποίο διάφοροι οργανισμοί εκτελούν στη ουσία την ίδια διαδικασία ενόσω μοιράζονται εμπειρίες, γνώση ή κάποιον κοινό εξοπλισμό. Ας σκεφτούμε το παράδειγμα της Salesforce.com. Η Salesforce υποστηρίζει και διαχειρίζεται τις διαδικασίες των πωλήσεων πολλών οργανισμών. Αφενός, αυτοί οι οργανισμοί μοιράζονται κάποιες υποδομές (διαδικασίες, βάσεις δεδομένων, κτλ.). Αφετέρου, δεν εξαναγκάζονται να ακολουθήσουν ένα αυστηρό μοντέλο διαδικασίας καθώς το σύστημα μπορεί να ρυθμιστεί για να υποστηρίξει διαφορετικές εκδοχές της ίδιας διαδικασίας (πωλήσεις). Σαν ένα άλλο παράδειγμα, ας σκεφτούμε τις βασικές διαδικασίες που εκτελεί μια Δημοτική Αρχή (π.χ. έκδοση οικοδομικών αδειών). Παρόλο που όλοι οι Δήμοι σε μία χώρα πρέπει να υποστηρίζουν το ίδιο βασικό σύνολο διαδικασιών, μπορεί κάλλιστα να υπάρχουν διαφοροποιήσεις. Προφανώς, είναι ενδιαφέρον να αναλυθούν τέτοιου είδους διαφοροποιήσεις ανάμεσα στους διάφορους οργανισμούς. Αυτοί οι οργανισμοί μπορούν να μάθουν ο ένας από τον άλλο και οι πάροχοι των υπηρεσιών μπορούν να βελτιώσουν τις υπηρεσίες τους και να προσφέρουν υπηρεσίες προσιτότητας αξίας με βάση τα αποτελέσματα της διαεπιχειρησιακής εξόρυξης διαδικασιών.

Και για τα δύο πλαίσια, πρέπει να αναπτυχθούν νέες τεχνικές ανάλυσης. Αυτές οι τεχνικές πρέπει βέβαια να λαμβάνουν υπόψη τα ζητήματα για το προσωπικό απόρρητο και την ασφάλεια. Οι οργανισμοί μπορούν να μην επιθυμούν να μοιραστούν πληροφορίες για λόγους ανταγωνισμού

ή λόγω έλλειψης εμπιστοσύνης. Επομένως, είναι απαραίτητο να αναπτυχθούν τεχνικές εξόρυξης που θα διατηρούν το προσωπικό απόρρητο.

## Π8: Παροχή Λειτουργικής Υποστήριξης

Αρχικά, η προσοχή της εξόρυξης διαδικασιών ήταν στραμμένη στην ανάλυση των παρελθόντων δεδομένων. Ωστόσο, σήμερα, πολλές πηγές δεδομένων ανανεώνονται σε (σχεδόν) πραγματικό χρόνο ενώ υπάρχει διαθέσιμη αρκετή υπολογιστική ισχύς για να αναλύσει τα δεδομένα άμα τη γέννηση τους. Επομένως, η εξόρυξη διαδικασιών δεν θα πρέπει να περιορίζεται σε στατική (off-line) ανάλυση, αλλά μπορεί να χρησιμοποιηθεί και για άμεση (online) λειτουργική υποστήριξη. Τρεις δράσεις λειτουργικής υποστήριξης διακρίνονται: η ανίχνευση (detection), η πρόβλεψη (prediction), και η σύσταση (recommendation). Τη στιγμή που μία περίπτωση αποκλίνει από την προκαθορισμένη διαδικασία, αυτό μπορεί να εντοπιστεί ώστε το σύστημα να παράγει μία προειδοποίηση. Συχνά τέτοιες ειδοποιήσεις είναι επιθυμητές σε άμεσο χρόνο (όσο ακόμα είναι εφικτό να επηρεαστούν οι καταστάσεις) και όχι σε μία αδρανής (μη συνδεδεμένη) κατάσταση. Τα παρελθόντα δεδομένα μπορούν να χρησιμοποιηθούν για να κατασκευαστούν μοντέλα πρόβλεψης. Αυτά μπορούν να χρησιμοποιηθούν για να καθοδηγήσουν τα στιγμιότυπα των διαδικασιών που βρίσκονται σε εξέλιξη. Για παράδειγμα, είναι δυνατόν να προβλεφθεί ο εναπομείνας χρόνος επεξεργασίας μιας εργασίας. Με βάση τέτοιες προβλέψεις, θα μπορούσαν να κατασκευαστούν συστήματα συστάσεων που θα πρότειναν συγκεκριμένες δράσεις για τη μείωση του κόστους ή τη συντόμηση του χρόνου ροής. Η εφαρμογή των τεχνικών της εξόρυξης διαδικασιών με ένα τέτοιο άμεσο τρόπο δημιουργεί πρόσθετες δυσκολίες / απαιτήσεις σε όρους υπολογιστικής ισχύς και ποιότητας δεδομένων.

## Π9: Συνδυασμός της εξόρυξης διαδικασιών με άλλους τύπους ανάλυσης

Η διοίκηση των λειτουργιών (operations management) και ειδικά η επιχειρησιακή έρευνα, είναι ένας κλάδος της επιστήμης της διοίκησης ο οποίος στηρίζεται σοβαρά στη μοντελοποίηση. Σε αυτήν χρησιμοποιούνται διάφορα μαθηματικά μοντέλα που ποικίλουν από γραμμικά προγράμματα ως προγραμματισμό έργων και μοντέλα ουρών, αλυσίδες

Markov και προσομοίωση. Η εξόρυξη δεδομένων μπορεί να οριστεί ως «η ανάλυση των (συχνά μεγάλων) συνόλων δεδομένων ώστε να βρεθούν σχέσεις που περνούσαν απαρατήρητες και να συνοψιστούν τα δεδομένα με νέους τρόπους οι οποίοι είναι τόσο κατανοητοί όσο και χρήσιμοι για τον ιδιοκτήτη των δεδομένων». Μία μεγάλη ποικιλία από τεχνικές έχουν αναπτυχθεί: ταξινόμηση (π.χ. εκμάθηση με δένδρα αποφάσεων), παλινδρόμηση, ομαδοποίηση (π.χ. ομαδοποίηση k-means) και ανακάλυψη μοτίβων (π.χ. εκμάθηση με κανόνες συσχέτισης).

Και τα δύο πεδία (διοίκηση των λειτουργιών και εξόρυξη δεδομένων) διαθέτουν χρήσιμες τεχνικές ανάλυσης. Η πρόκληση συνίσταται στο να συνδυαστούν οι τεχνικές αυτών των πεδίων με την εξόρυξη διαδικασιών. Ας θεωρήσουμε για παράδειγμα την προσομοίωση. Οι τεχνικές εξόρυξης διαδικασιών μπορούν να χρησιμοποιηθούν για την εκμάθηση ενός μοντέλου προσομοίωσης από παρελθόντα δεδομένα. Ακολούθως, το μοντέλο της προσομοίωσης μπορεί να χρησιμοποιηθεί να προσφέρει λειτουργική υποστήριξη. Εξαιτίας της στενής σύνδεσης μεταξύ του αρχείου καταγραφών γεγονότων και του μοντέλου, το μοντέλο μπορεί να χρησιμοποιηθεί για την αναπαραγωγή του παρελθόντος, και η προσομοίωση να ξεκινήσει από την παρούσα κατάσταση προσφέροντας ένα «κουπί γρήγορης προώθησης» για τις μελλοντικές καταστάσεις, βασισμένη στα επί του παρόντος γεγονότα.

Παρομοίως, είναι επιθυμητό να συνδυαστεί η εξόρυξη διαδικασιών με οπτικές αναλύσεις (visual analytics). Οι οπτικές αναλύσεις συνδυάζουν την αυτόματη ανάλυση με αλληλεπιδραστικές οπτικοποιήσεις για μία καλύτερη κατανόηση ενός μεγάλου και πολύπλοκου συνόλου δεδομένων. Οι οπτικές αναλύσεις εκμεταλλεύονται τις καταπληκτικές ικανότητες των ανθρώπων να διακρίνουν μοτίβα σε αδόμητα δεδομένα. Συνδυάζοντας τις τεχνικές για αυτόματη εξόρυξη διαδικασιών με τις αλληλεπιδραστικές οπτικές αναλύσεις, είναι δυνατό να εξαγάμε βαθύτερη γνώση και περισσότερα συμπεράσματα από τα αρχεία των γεγονότων.

## Π10: Βελτίωση της χρησιμότητας για τους μη-ειδικούς

Ένας από τους στόχους της εξόρυξης διαδικασιών είναι να δημιουργήσει «ζωντανά μοντέλα διαδικασιών», δηλ. τα μοντέλα των διαδικασιών να χρησιμοποιούνται σε καθημερινή βάση από το να δημιουργηθούν στατικά μοντέλα που

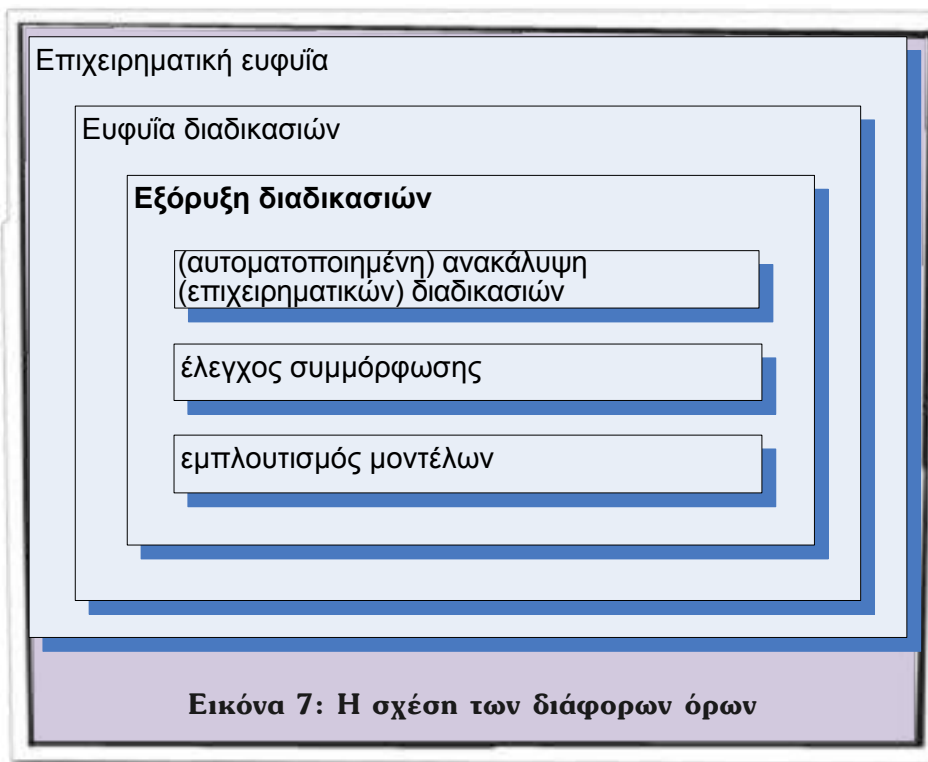
θα καταλήξουν σε κάποιο αρχείο. Τα νέα δεδομένα γεγονότων μπορούν να χρησιμοποιηθούν για να ανακαλύψουν αναδυόμενες συμπεριφορές. Ο σύνδεσμος μεταξύ δεδομένων γεγονότων και μοντέλων διαδικασιών επιτρέπει την προβολή της τρέχουσας κατάστασης και των πρόσφατων δραστηριοτήτων σε πιο επίκαιρα μοντέλα. Επομένως, οι τελικοί χρήστες θα έχουν τη δυνατότητα να αλληλεπιδρούν με τα αποτελέσματα της εξόρυξης διαδικασιών σε καθημερινή βάση. Τέτοιες αλληλεπιδράσεις είναι εξαιρετικά χρήσιμες αλλά απαιτούν διαισθητικές και εύληπτες χρηστικές διεπιφάνειες (user interface). Η πρόκληση είναι να αποκρυφτούν οι πολύπλοκοι αλγόριθμοι της εξόρυξης διαδικασιών πίσω από φιλικές χρηστικές διεπιφάνειες ενώ οι παράμετροι να καθορίζονται αυτόματα και να προτείνονται οι κατάλληλοι τύποι ανάλυσης.

## Π11: Βελτίωση της ικανότητας κατανόησης από του μη-ειδικούς

Ακόμη και αν είναι εύκολη η παραγωγή αποτελεσμάτων εξόρυξης διαδικασιών, αυτό δε σημαίνει πως τα αποτελέσματα αυτά θα είναι όντως χρήσιμα. Είναι δυνατόν ο χρήστης να έχει προβλήματα να κατανοήσει το αποτέλεσμα ή να ολισθήσει σε λανθασμένα συμπεράσματα. Για την αποφυγή τέτοιων προβλημάτων, τα αποτελέσματα πρέπει να παρουσιάζονται με μία κατάλληλη αναπαράσταση (βλ. επίσης και [sub:ΚΟ5:-Τα-μοντέλα]). Επιπλέον, θα πρέπει πάντα να σημειώνεται ευκρινώς το επίπεδο αξιοπιστίας των αποτελεσμάτων. Μπορεί να υπάρχουν πολύ λίγα δεδομένα για να τεκμηριωθούν επαρκώς κάποια συγκεκριμένα συμπεράσματα. Στην πραγματικότητα, οι υπάρχουσες τεχνικές ανακάλυψης διαδικασιών συνήθως δεν προειδοποιούν για μία χαμηλή προσαρμογή ή για την υπερπροσαρμογή. Δείχνουν πάντα ένα μοντέλο ακόμα και αν υπήρξαν πολύ λίγα δεδομένα για να δικαιολογηθεί οποιοδήποτε συμπέρασμα.

## 5 Επίλογος

Η ομάδα κρούσης εξόρυξης διαδικασιών της IEEE στοχεύει α) στην προώθηση της εφαρμογής της εξόρυξης διαδικασιών, β) στο να καθοδηγήσει τους προγραμματιστές, τους συμβούλους, τα διοικητικά στελέχη, και τους τελικούς χρήστες για τη χρήση των πλέον σύγχρονων τεχνικών, και γ) στην παρακίνηση της έρευνας στην εξόρυξη διαδικασιών. Αυτή η διακήρυξη διατυπώνει τις κύριες



λειτουργίες της εξόρυξης διαδικασιών. Η Εικόνα 7 μπορεί να παραπλανεί, με την έννοια πως τα περισσότερα εργαλεία επιχειρηματικής ευφυΐας δεν παρέχουν λειτουργίες εξόρυξης διαδικασιών όπως αυτές έχουν περιγραφεί στο παρόν κείμενο. Ο όρος επιχειρηματική ευφυΐα χρησιμοποιείται συχνά καταχρηστικά ένεκα ενός συγκεκριμένου εργαλείου ή μιας μεθόδου, η οποία καλύπτει μόνο ένα μικρό μέρος του ευρύτερου φάσματος της επιχειρηματικής ευφυΐας.

Υπάρχει το ενδεχόμενο η χρήση εναλλακτικών όρων να οφείλεται σε εμπορικούς σκοπούς. Κάποιοι κατασκευαστές λογισμικού ίσως να επιθυμούν επίσης να τονίσουν κάποιο συγκεκριμένο σημείο (π.χ. την ανακάλυψη ή την ευφυΐα). Ωστόσο, προς αποφυγήν σύγχυσης, για το πεδίο στο οποίο αναφέρεται αυτή η διακήρυξη, είναι καλύτερα να

αρχές και τις προθέσεις της ομάδας κρούσης. Μετά την εισαγωγή στο θέμα της εξόρυξης διαδικασιών, η διακήρυξη απεικονίζει μερικές κατευθυντήριες οδηγίες (Κεφάλαιο 3) και προκλήσεις (Κεφάλαιο 4). Οι κατευθυντήριες οδηγίες μπορούν να χρησιμοποιηθούν προκειμένου να αποφευχθούν προφανή λάθη. Η λίστα με τις προκλήσεις προορίζεται για την καθοδήγηση της έρευνας και των προσπαθειών ανάπτυξης. Και οι δύο στοχεύουν στην αύξηση του επιπέδου ωριμότητας της εξόρυξης διαδικασιών.

Κλείνοντας, λίγη συζήτηση πάνω στην ορολογία. Καταρχήν, στην παρούσα, ελληνική μετάφραση της διακήρυξης παρατίθεται στο τέλος ένα ευρετήριο των λέξεων-κλειδιών με την αγγλική τους (πρωτότυπη) μετάφραση. Οι ακόλουθοι όροι χρησιμοποιούνται στους κύκλους της εξόρυξης διαδικασιών: Εξόρυξη ροών εργασιών (workflow mining), εξόρυξη (επιχειρηματικών) διαδικασιών ((business) process mining), αυτόματη ανακάλυψη (επιχειρηματικών) διαδικασιών (automated (business) process discovery), και ευφυΐα (επιχειρηματικών) διαδικασιών ((business) process intelligence). Διαφορετικοί οργανισμοί φαίνεται να χρησιμοποιούν διαφορετικούς όρους για έννοιες που επικαλύπτονται. Για παράδειγμα, η Gartner προωθεί τον όρο “Automated Business Process Discovery” ενώ η Software AG χρησιμοποιεί τον όρο “Process Intelligence” για να αναφερθεί στην πλατφόρμα ελέγχου της. Ο όρος «εξόρυξη ροών εργασιών» μοιάζει λιγότερο κατάλληλος καθώς η δημιουργία μοντέλων ροών εργασιών

είναι απλά μία από τις πολλές δυνατότητες εφαρμογών της εξόρυξης διαδικασιών. Παρομοίως, η προσθήκη του όρου «επιχειρηματικές» περιορίζει το εύρος των εφαρμογών της εξόρυξης διαδικασιών. Υπάρχει πλήθος εφαρμογών της εξόρυξης διαδικασιών (π.χ. ανάλυση συστημάτων υψηλής τεχνολογίας ή ανάλυση ιστοχώρων) όπου αυτή η προσθήκη μοιάζει ακατάλληλη. Παρόλο που η ανακάλυψη διαδικασιών είναι ένα σημαντικό κομμάτι του φάσματος της εξόρυξης διαδικασιών, είναι μόνο μία από τις πολλές περιπτώσεις χρήσεως. Ο έλεγχος συμμόρφωσης, η πρόβλεψη, η εξόρυξη οργανωτικών δομών, η ανάλυση κοινωνικών δικτύων, κτλ. είναι άλλες περιπτώσεις που εκτείνονται πέρα από την ανακάλυψη διαδικασιών.

Η Εικόνα 7 συσχετίζει κάποιους από τους όρους που αναφέρθηκαν. Όλες οι τεχνολογίες και οι μέθοδοι που αποσκοπούν να παρέχουν πληροφορίες που να υποστηρίζουν τη λήψη αποφάσεων μπορούν να τοποθετηθούν κάτω από το σκέπαστρο της επιχειρηματικής ευφυΐας. Η ευφυΐα (επιχειρηματικών) διαδικασιών μπορεί να ιδωθεί ως ο συνδυασμός της επιχειρηματικής ευφυΐας και της ΔΕΔ, αφού οι τεχνικές της επιχειρηματικής ευφυΐας χρησιμοποιούνται για να αναλύσουν και να βελτιστοποιήσουν τις διαδικασίες και τη διοίκηση τους. Η εξόρυξη διαδικασιών μπορεί να θεωρηθεί ως μία υλοποίηση της ευφυΐας των διαδικασιών η οποία δέχεται αρχεία καταγραφών γεγονότων ως σημείο εκκίνησης. Η (αυτοματοποιημένη) ανακάλυψη (επιχειρηματικών) διαδικασιών είναι απλά μία από τις τρεις βασικές

## Συντομογραφίες

- ΔΕΔ: Διοίκηση Επιχειρηματικών Διαδικασιών
- BPMN: Business Process Modeling Notation
- XES: Extensible Event Stream (πρότυπο βασισμένο σε XML)
- MXML: πρότυπο βασισμένο σε XML που αντικαταστάθηκε από το XES

## Όροι που χρησιμοποιούνται:

- Εξόρυξη ροών εργασιών - (workflow mining)
- Εξόρυξη (επιχειρηματικών) διαδικασιών - (business) process mining
- (Αυτόματη) ανακάλυψη (επιχειρηματικών) διαδικασιών - (automated business) process discovery
- Ευφυΐα (επιχειρηματικών) διαδικασιών - (business) process intelligence

Προτιμάται και συστήνεται ο όρος «εξόρυξη διαδικασιών» - “Process mining”

χρησιμοποιείται ο όρος «εξόρυξη διαδικασιών»

## Λεξικό όρων

### A

- (αυτοματοποιημένη) ανακάλυψη (επιχειρηματικών) διαδικασιών: (Automated business) process discovery
- ακρίβεια: precision
- ανακάλυψη διαδικασιών: process discovery
- (ανα)σχεδιασμός: (re)design
- ανίχνευση: detection
- απλότητα: simplicity
- αποκλειστική απόφαση: XOR
- αρχείο καταγραφών γεγονότων: event log

### Γ

- γεγονός: event
- γενίκευση: generalization

### Δ

- διάσταση της ανά περίπτωση προσέγγισης: case perspective
- διασταυρωμένη επικύρωση: cross-validation
- διασταύρωση (π.χ. AND): split
- διαστρέβλωση της αναπαράστασης: representational bias
- διεκπεραιωτική ικανότητα: throughput
- διεπιχειρησιακός: cross-organizational
- διευθέτηση: adjustment
- διοίκηση επιχειρηματικών διαδικασιών: Business Process Management
- διοίκηση των λειτουργιών: operations management

### E

- έκτοπα σημεία: outliers

- έλεγχος συμμόρφωσης: conformance checking
- εμπλουτισμός: enhancement
- εναλλακτική απόφαση: OR
- εννοιολογική απόκλιση: concept drift
- ένωση (π.χ. AND): join
- επανάληψη: replay
- επαναρύθμιση: reconfiguration
- επιχειρηματική ευφυΐα: business intelligence
- ευφυΐα διαδικασιών: process intelligence
- εξόρυξη δεδομένων: data mining
- εξόρυξη διαδικασιών: process mining

### Z

- ζητήματα ασφάλειας: security issues

### Λ

- λειτουργική υποστήριξη: operational support
- Λεπίδα του Όκαμ: Occam's Razor

### M

- μοτίβα: patterns

### O

- οπτικές αναλύσεις: visual analytics
- οργανωσιακή διάσταση: organizational perspective

### Π

- περίπτωση διαδικασίας: case
- πρόβλεψη: prediction
- πρόγραμμα μέτρησης και σύγκρισης των επιδόσεων: benchmarking
- προσαρμογή: fitness
- προσωπικό απόρρητο: privacy

### P

- ροή αλληλουχίας: sequence

### Σ

- σημασιολογία: semantics
- στατική ανάλυση: off-line analysis

- στιγμιότυπο διαδικασίας: process instance
- στοχοθετημένη αφαίρεση: abstraction
- σύσταση: recommendation
- συσχέτιση γεγονότων: event correlation

### Υ

- υπερπροσαρμογή: overfitting
- υποπροσαρμογή: underfitting,

### X

- χαρτογραφία: cartography
- χρηστική διεπιφάνεια: user interface
- χρονόσημο: timestamp

### Υπεύθυνοι ελληνικής μετάφρασης

Δελιάς Παύλος  
✉ pdelias@teikav.edu.gr  
Μιχαλόπουλος Νίκος

## Συγγραφείς

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

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

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

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