

Grzegorz BLIŹNIUK, Mariusz CHMIELEWSKI, Tomasz GZIK, Jarosław KOSZELA  
Wojskowa Akademia Techniczna, Wydział Cybernetyki

## HURTOWNIE PROCESÓW

**Streszczenie.** W artykule przedstawiono ogólny model architektury hurtowni procesów biznesowych jako systemu analiz procesów biznesowych i wspomagania podejmowania decyzji. Przedstawiono wstępne wyniki realizacji konstrukcji składu danych o procesach, zarówno ich definicji i zrealizowanych instancji, oraz rozszerzonej o wykorzystanie ontologii i metody transformacji i unifikacji procesów pochodzących z różnych źródeł (~ETL). Zaprezentowano również problematykę związaną z konstrukcją procesów dynamicznych (*quasi-ad hoc*) oraz podstawowe metody analiz strukturalnej i semantycznej procesów jako elementów składowych metod analizy procesów.

**Słowa kluczowe:** hurtownia procesów, procesy biznesowe, procesy dynamiczne, metody i techniki analizy procesów, semantyczna unifikacja procesów

## WAREHOUSE OF PROCESSES

**Summary.** The main purpose of this paper is to present an architectural framework for the warehouse of business processes. This concept has been developed as a basis for a business process analysis and a decision support system construction. Following work concentrates on preliminary results gathered from the conducted research, concerning designed analytical approach. Main parts of this method refer to construction of a data store holding process definitions as well as instances of executed processes. We have also managed to solve some secondary issues connected with process unification and transformation tasks using ontology-based mechanisms. Fusing described elements a consistent framework have been defined, which concentrates on a dynamic process construction and a process data flow semantic unification.

**Keywords:** warehouse of processes, business process, dynamic process, methods and techniques process analysis, semantic unified of process

## 1. Wstęp

Zagadnienia przedstawiane w dalszej części mają swoje źródło w projekcie dotyczącym realizacji komputerowych repozytoriów ścieżek klinicznych (ang. *Clinical Pathways*), w którym autorzy artykułumieli swój udział. Jednym z wymagań, na jakie napotyka się w tym miejscu, jest właściwe dostarczanie tych opisów w postaci przewidzianej dla procesów. Wynika to głównie z tego, że ścieżka kliniczna będąca planem leczenia pacjenta jest procesem i składa się z sekwencji powiązanych ze sobą działań medycznych, realizowanych w określonym porządku i w określonym czasie. Takie podejście do przedstawiania ścieżek klinicznych pozwala na ich wykorzystywanie w sposób analogiczny do sposobu wykorzystywania modeli procesów biznesowych. Istotna jest w tym zakresie możliwość:

- usprawniania procesów (ang. *process reengineering*),
- symulowania procesów (ang. *process simulation*),
- automatyzowania procesów (ang. *workflow*).

W niniejszym artykule zostały również poruszone zagadnienia tzw. hurtowni procesów, która umożliwi trwałe przechowywanie odpersonalizowanej informacji o sposobie wykonania planów leczenia w postaci procesów workflow dla poszczególnych ścieżek klinicznych. Na tej podstawie można skutecznie wspomagać prowadzenie indywidualnego procesu leczenia, a także prowadzić badania jakościowe dotyczące procesu leczenia. Autorzy rozważają również problematykę dedykowanego metamodelu instancji procesów gromadzonych w hurtowni oraz metody unifikacji opisu zrealizowanych procesów z wykorzystaniem opracowanej do tego celu ontologii. Jest to konieczne dla utworzenia mechanizmów przechowywania trwałej informacji o przebiegu procesów leczenia, dokonywania różnego rodzaju analiz (w tym m.in. wyszukiwania: podobieństw, wzorców, trendów i anomalii w zrealizowanych procesach leczenia), analiz semantycznych i ontologicznych procesów leczenia, analiz wielowymiarowych przebiegów tych procesów oraz opracowywania wniosków w zakresie efektywności przebiegów tych procesów.

## 2. Ścieżki kliniczne – podejście procesowe

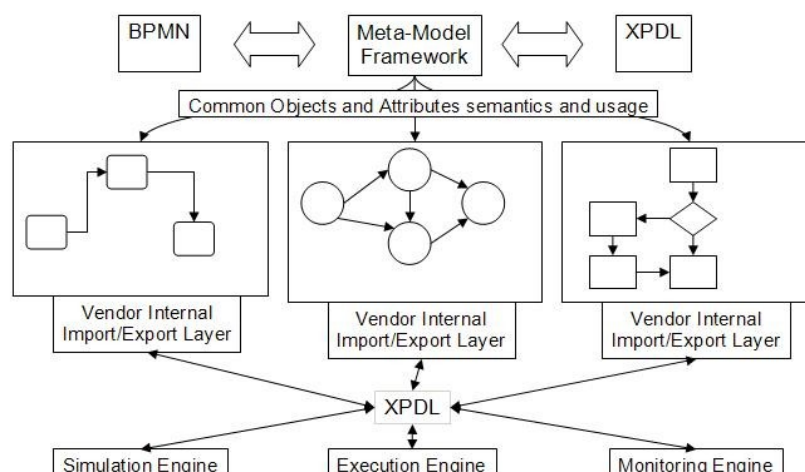
### 2.1. Procesy biznesowe

Podejście procesowe, a dokładnie zarządzanie procesami biznesowymi (ang. *Business Process Management – BPM*), w szczególności ich modelowaniem, definiowaniem i realizacją, wymaga zastosowania odpowiednich metod oraz narzędzi informatycznych. Do niedawna przeważnie każdy system klasy BPM oparty był na odmiennym od pozostałych podejściu

do realizacji przedmiotowych zagadnień, co stanowiło poważne ograniczenie w kontekście szeroko pojętej i aktualnie bezwzględnie „pożądaną” integracji różnych środowisk systemowych. Brak uniwersalnej reprezentacji procesów powodował również istotne ograniczenia w zakresie badania przebiegu poszczególnych instancji procesów, tym samym ograniczał możliwości automatycznego wnioskowania i monitorowania procesów. Aktualnie funkcjonują i stale rozwijane są standardy, które stanowią wytyczne dla sposobu modelowania i definiowania procesów biznesowych, a także budowy platform BPM.

Business Process Model and Notation (BPMN) to jedna z najbardziej popularnych notacji, dedykowana do modelowania procesów. Rozwijana jest przez Object Management Group i jest uwzględniona praktycznie we wszystkich dostępnych na rynku narzędziach.

XML Process Definition Language (XPDL) to opracowany i rozwijany przez Workflow Management Coalition (WfMC) język definiowania procesów, oparty na XML, który podobnie jak BPMN implementowany jest w większości narzędzi przeznaczonych do modelowania i definiowania procesów. Zbieżność ta nie jest przypadkowa, gdyż z założenia język XPDL ma pełnić rolę odpowiednika w relacji 1-1 dla notacji BPMN, tzn. na podstawie modelu BPMN powinno być możliwe wygenerowanie definicji tego modelu wyrażonej w języku XPDL i odwrotnie (ang. *reverse engineering*). Definicja wyrażona w XPDL (i tym samym model stworzony z wykorzystaniem BPMN) powinna być „przenaszalna” pomiędzy różnymi narzędziami informatycznymi (rys. 1).



Rys. 1. Koncepcja wymiany definicji procesów<sup>1</sup>

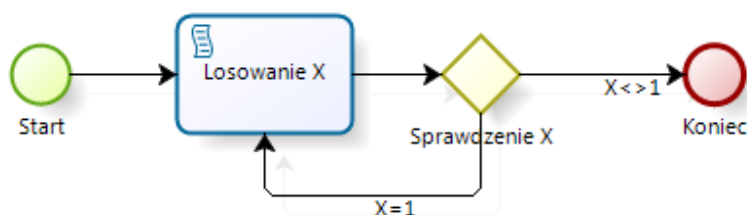
Fig. 1. Example interexchange of processes definition

Proces biznesowy w XPDL rozumiany jest jako przepływ prac/zadań (ang. *workflow*) i opisywany jest na podstawie Meta-Modelu, zawierającego potrzebne do zdefiniowania procesu obiekty (np. aktywność, rola, komunikat, asocjacja), analogicznie do struktury BPMN. Meta-Model określa logikę i sposób opisywania procesów, który jest zgodny z semantyką notacji BPMN. Na podstawie definicji procesu, wyrażonej w XPDL, możliwa jest realizacja

<sup>1</sup> Źródło: Process Definition Interface – XML Process Definition Language, 2005.

procesu w środowisku wykonawczym (ang. *workflow engine* lub *BPM engine*), symulacja procesu (ang. *Business Process Simulation*) i jego monitorowanie (ang. *Business Process Monitoring*). Dodatkowy zakres wykorzystania związany jest z poruszaniem w niniejszym artykule zagadnieniami hurtowni procesów (ang. *Process Warehouse*) i procesów dynamicznych.

Przeanalizujemy prosty przykład mapowania BPMN na XPDŁ. Rysunek 2 przedstawia krótki proces zamodelowany z wykorzystaniem BPMN, którego celem jest wylosowanie liczby różnej od 1 ze zbioru  $\{1,2,3\}$ .



Rys. 2. Przykładowy proces losowania liczby X  
Fig. 2. Example the process random X

Zdarzeniem rozpoczynającym proces jest „Start” (typ BPMN – Start Event), czynność „Losowanie X” jest wykonywana automatycznie (typ BPMN – Script Task) i jej celem jest wylosowanie wartości ze zbioru  $\{1,2,3\}$  i przyporządkowanie zmiennej X. Na etapie bramki logicznej „Sprawdzenie X” (typ BPMN – XOR Gateway) następuje sprawdzenie wartości zmiennej X; jeżeli wynosi 1 – następuje kolejne losowanie (czynność „Losowanie X”), jeżeli jest różna od 1 – proces kończy się (zdarzenie „Koniec”, typ BPMN – End Event). Tabela 1 prezentuje poszczególne elementy procesu, wyrażone w języku XPDŁ.

W uzupełnieniu do zawartości tabeli 1 definicja XPDŁ zawiera również informacje o metadanych procesu, m.in.: nazwie i identyfikatorze procesu, dacie utworzenia, autorze, wykorzystanych narzędziach, wersji.

Podczas modelowania ścieżek klinicznych konieczne jest skuteczne uporanie się z problemem dużej różnorodności decyzji i zdarzeń, jakie mogą mieć miejsce w trakcie trwania leczenia, których liczność powinna być skończona. Należy ponadto dysponować odpowiednim językiem modelowania, pozwalającym na przedstawienie stosunkowo trudno przewidywalnych przebiegów procesów, jakimi są procesy leczenia pacjenta.

Jak wspomniano, na rynku dostępnych jest wiele metod, notacji i języków przeznaczonych do modelowania procesów, przy czym zdecydowanymi „liderami” są BPMN oraz XPDŁ. Każda z metod może zostać zastosowana z mniejszym lub większym powodzeniem do modelowania ścieżek klinicznych. Jako kryterium powodzenia można przyjąć:

1. stopień czytelności ścieżek dla „zwykłego” odbiorcy,
2. możliwość odzwierciedlenia na modelu wszystkich zachowań procesów,
3. możliwość przenoszenia definicji ścieżek,
4. możliwość automatyzacji całych ścieżek lub ich części.

Tabela 1

## Mapowanie BPMN na XPDL

Element modelu BPMN	Definicja XPDL
Start	<pre>&lt;Activity Id="1" Name="start"&gt;   &lt;Description&gt;Zdarzenie inicjujące proces&lt;/Description&gt;   &lt;Event&gt;     &lt;StartEvent Trigger="None" /&gt;   &lt;/Event&gt;   &lt;Documentation&gt;Zdarzenie inicjujące proces&lt;/Documentation&gt;   &lt;ExtendedAttributes /&gt; &lt;/Activity&gt;</pre>
Przepływ Start – Losowanie X	<pre>&lt;Transition Id="10" From="1" To="2" Name=""&gt;   &lt;Condition /&gt;   &lt;Description&gt;Przepływ sekwencyjny pomiędzy "Start" a "Losowanie X" &lt;/Description&gt;   &lt;ExtendedAttributes /&gt; &lt;/Transition&gt;</pre>
Losowanie X	<pre>&lt;Activity Id="2" Name="Losowanie X"&gt;   &lt;Description&gt;Losowanie X ze zbioru {1,2,3}&lt;/Description&gt;   &lt;Implementation&gt;     &lt;Task&gt;       &lt;TaskScript&gt;         &lt;Script&gt;x=random(1:3)&lt;/Script&gt;       &lt;/TaskScript&gt;     &lt;/Task&gt;   &lt;/Implementation&gt;   &lt;Performers /&gt;   &lt;Documentation&gt;Losowanie X ze zbioru {1,2,3}&lt;/Documentation&gt;   &lt;InputSets /&gt;   &lt;OutputSets /&gt;   &lt;Loop LoopType="None" /&gt;   &lt;ExtendedAttributes /&gt; &lt;/Activity&gt;</pre>
Przepływ Losowanie X – Sprawdzenie X	<pre>&lt;Transition Id="11" From="2" To="3" Name=""&gt;   &lt;Condition /&gt;   &lt;Description&gt;Przepływ sekwencyjny pomiędzy "Losowanie X" a "Sprawdzenie x" &lt;/Description&gt;   &lt;ExtendedAttributes /&gt; &lt;/Transition&gt;</pre>
Sprawdzenie X	<pre>&lt;Activity Id="3" Name="Sprawdzenie X"&gt;   &lt;Description&gt;Bramka logiczna typu XOR, sprawdzenie wartości X&lt;/Description&gt;   &lt;Route /&gt;   &lt;Documentation&gt;Bramka logiczna typu XOR, sprawdzenie wartości X &lt;/Documentation&gt;   &lt;ExtendedAttributes /&gt; &lt;/Activity&gt;</pre>
Przepływy Sprawdzenie X – Losowanie X	<pre>&lt;Transition Id="12" From="3" To="2" Name="X=1"&gt;   &lt;Condition Type="CONDITION"&gt;     &lt;Expression&gt;x=1&lt;/Expression&gt;   &lt;/Condition&gt;   &lt;Description&gt;Przepływ sekwencyjny warunkowy pomiędzy "Sprawdzenie X"   a "Losowanie X" - jeżeli X=1&lt;/Description&gt;   &lt;ExtendedAttributes /&gt; &lt;/Transition&gt;</pre>
Przepływy Sprawdzenie X – Koniec	<pre>&lt;Transition Id="13" From="3" To="4" Name="X&lt;1"&gt;   &lt;Condition Type="CONDITION"&gt;     &lt;Expression&gt;x&lt;1&lt;/Expression&gt;   &lt;/Condition&gt;   &lt;Description&gt;Przejście sekwencyjne warunkowe pomiędzy "Sprawdzenie x" a   "Koniec" - jeżeli x&lt;1&lt;/Description&gt;   &lt;ExtendedAttributes /&gt; &lt;/Transition&gt;</pre>
Koniec	<pre>&lt;Activity Id="4" Name="Koniec"&gt;   &lt;Description&gt;Zdarzenie kończące proces&lt;/Description&gt;   &lt;Event&gt;     &lt;EndEvent Result="None" /&gt;   &lt;/Event&gt;   &lt;Documentation&gt;Zdarzenie kończące proces&lt;/Documentation&gt;   &lt;ExtendedAttributes /&gt; &lt;/Activity&gt;</pre>

Czytelność modelu procesów uzależniona jest m.in. od liczby różnych elementów graficznych wykorzystanych do budowy modelu. Im liczba ta jest większa, tym model staje się mniej czytelny i trudny w zrozumieniu. Dlatego też ogólnie przyjęta zasada modelowania, niezależna od stosowanej notacji/języka, mówi, iż powinno stosować się możliwie najmniejszą liczbę różnych symboli graficznych, bez względu na to jak „bogata” jest wykorzystywana notacja/język.

Nie wszystkie dostępne notacje i języki umożliwiają modelowanie podstawowych wzorców procesowych. Często wymagane jest definiowanie dodatkowych atrybutów, które ukrywane są pod elementami graficznymi przyjętej notacji. Notacją, która w zakresie odzwierciedlenia wzorców procesowych zdecydowanie góruje nad konkurentami, jest BPMN – tym samym języki, których konstrukcje odpowiadają tej notacji, np. XPDŁ.

„Przenaszalność” definicji procesów, w tym definicji ścieżek klinicznych, ma bardzo duże znaczenie m.in. ze względu na możliwość ich rozpowszechniania. Model procesów powinien być zbudowany na podstawie notacji/języka zapewniającej możliwość przenoszenia jego definicji między najważniejszymi (z punktu widzenia projektu) narzędziami informatycznymi. Ma to również bardzo duże znaczenie w kontekście automatyzacji procesów i ich interpretowania przez silniki *workflow*. Język XPDŁ stanowi standard w tym zakresie i jest wspierany przez większość liczących się na rynku narzędzi. Wiele z nich umożliwia automatyczne wygenerowanie definicji procesu w XPDŁ bezpośrednio z modelu graficznego BPMN, co m.in. potwierdza przytoczony fakt, iż para BPMN i XPDŁ jest „liderem” w dziedzinie modelowania procesów [1].

Definicja procesu workflow, opisująca konkretną ścieżkę kliniczną dla jednostki chorobowej, wymaga zdefiniowania jej na trzech poziomach [3]. Na najwyższym poziomie (tzw. poziomie krajowym/narodowym) powinny zostać opracowane wytyczne i ścieżki kliniczne dla jednostek chorobowych przez zespoły specjalistów, którzy na podstawie najnowszej dostępnej wiedzy medycznej opracują wzorcową definicję postępowania w przypadku zaistnienia danej jednostki chorobowej. Następnym poziomem opisu wytycznych i ścieżek klinicznych jest modyfikacja wzorcowego procesu postępowania dla określonej jednostki chorobowej na potrzeby określonej placówki medycznej. Modyfikacja ta może być podyktowana uwarunkowaniami np. ekonomicznymi, technicznymi czy też merytorycznymi i jest konieczna, aby dostosować wzorcową definicję procesu z poziomu krajowego na możliwości danej placówki medycznej. Na najniższym poziomie opisu postępowania w ramach jednostki chorobowej jest plan leczenia dla konkretnego pacjenta, bazujący na definicji procesu opisującego ścieżkę kliniczną. Plan leczenia ma charakter dynamiczny, czyli jest definicją procesu, która może być modyfikowana zarówno na etapie przygotowywania planu leczenia (ang. *design time*), jak również może podlegać modyfikacji w trakcie realizacji tego procesu (ang. *run time*).

Modyfikacja planu leczenia jest decyzją lekarza prowadzącego i może być spowodowana koniecznością odejścia od definicji wzorcowej ze względu na wystąpienie dodatkowych czynników, które nie były uwzględniane przy definicji procesu wzorcowego (np. współistnienie wielu jednostek chorobowych, ograniczenia czasowe, zastosowanie różnych metod diagnostyki medycznej itp.). Modyfikacja ta nie powinna być ograniczona tylko do zmian w obrębie definicji wzorcowej ścieżki, ale w sposób dość swobodny powinna umożliwiać modyfikację lub całkowitą zmianę definicji takiej ścieżki w szczególnych procesach leczenia. W tym przypadku konieczne jest posiadanie możliwości opisanie procesu z uwzględnieniem jego dynamiki zarówno na etapie opisu definicji procesu, jak i jego wykonania. Realizacja wielu planów leczenia pozwoli na analizę sposobów leczenia podobnych przypadków i jednostek chorobowych przy uwzględnieniu aspektów statystycznych zgromadzonych i jednocześnie zrealizowanych planów leczenia oraz semantycznych i ontologicznych uwarunkowań tych działań.

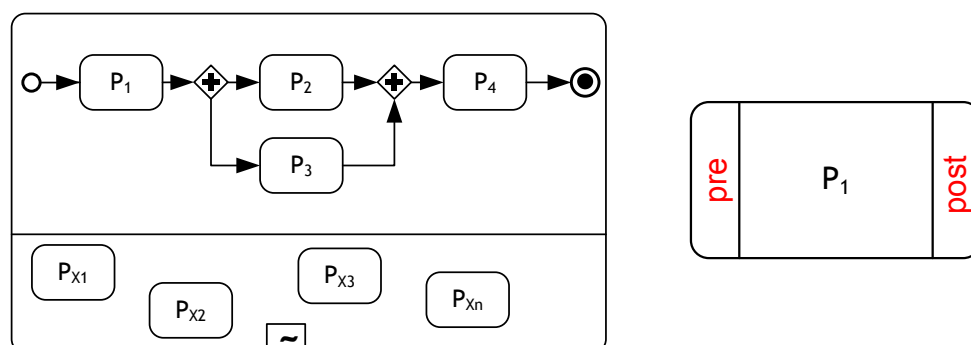
## 2.2. Procesy dynamiczne, adaptacyjne i generyczne

Możliwość definiowania ścieżek klinicznych w postaci procesów biznesowych oraz realizacji planów leczenia (jako ich praktycznych instancji) pozwala na ich gromadzenie i możliwości wykonania analizy zgromadzonych danych o procesach oraz ich realizacjach. Obecnie stosowane podejście w systemach klasy workflow powoduje, że podstawowym nurtem automatyzacji procesów jest ich realizacja zgodnie z definicją procesu. Jest to dość „sztywne” podejście, ale skuteczne w obszarze bardzo dobrze rozpoznany i przebadany. Pewnymi sposobami na zwiększenie elastyczności w stosowaniu podejścia procesowego są tzw. procesy ad-hoc, które pozwalają na wprowadzenie dużego niedeterminizmu zarówno w definicji procesu, jak i jego realizacji. Formą pośrednią pomiędzy „klasyczną” definicją procesu a procesami typu ad-hoc są procesy dynamiczne [4] (Rys. 3), które pozwalają na określenie definicji głównej ścieżki procesu oraz zbioru możliwych lub dopuszczalnych procesów/aktywności wraz z mechanizmami, pozwalającymi na określenie możliwości użycia poszczególnych procesów w realizacji instancji procesu w różnych jej stanach, do czego jest wykorzystywany opis warunków *pre* i *post* procesu.

Zwiększanie elastyczności procesów w obszarach definicji i realizacji jest szczególnie istotne z punktu widzenia realizacji planów leczenia, gdzie definicja postępowania nie jest „sztywno” związana z jego realizacją, tak jak ma to miejsce w przypadku planów indywidualnego leczenia pacjenta, bazującego na ścieżce klinicznej.

Procesy generyczne można inaczej nazwać procesami uniwersalnymi. Można z nich skorzystać przy realizacji dowolnego procesu. Aktualne standardy dotyczące procesów (BPEL, XPDL) nie udostępniają możliwości tworzenia takich procesów, a tworzenie ich za pomocą zadań ad-hoc byłoby bardzo uciążliwe. Również w tym przypadku skorzystanie z procesów

dynamicznych pozwoli na rozwiązanie tego problemu, a także, przy odpowiednich założeniach, na (przynajmniej częściowo) automatyczne tworzenie nowych procesów.



Rys. 3. Przykładowy diagram procesu dynamicznego z opisem warunków *pre* i *post* wykonania procesu

Fig. 3. Example diagram of dynamic process with pre and post definitions of process execution

Procesy adaptujące to odmiana procesów biznesowych, które mogą być adaptowane do nowych warunków, napotkanych w trakcie wykonywania procesu w sposób prosty i niewymagający dużych nakładów pracy lub kosztów. Może to być adaptacja manualnie sterowana (przez człowieka) lub wspomagana automatycznie (w całości lub części). Aktualnie systemy workflow wspomagają takie procesy w sposób bardzo ograniczony z użyciem aktywności ad-hoc, takie rozwiązanie musi jednak zostać przewidziane przez projektującego proces przed jego realizacją. Pełniejszą adaptację do nowych, nieprzewidzianych warunków można zrealizować przy zastosowaniu procesów dynamicznych.

### 3. Hurtownie procesów

Możliwość zgromadzenia wielu definicji procesów i danych o ich realizacji zarówno procesów „klasycznych”, jak i dynamicznych, generycznych, adaptacyjnych czy ad-hoc daje możliwość dokonywania analiz zgromadzonych danych o procesach na etapach ich definicji i realizacji. Takim specjalizowanym rozwiązaniem, pozwalającym na zgromadzenie i analizę procesów, jest hurtownia procesów [5] jako specjalizowana hurtownia danych wraz z metodami i narzędziami do eksploracji procesów. W przypadku hurtowni procesów mamy do czynienia z dwoma obszarami zasobów dotyczących procesów: ich definicji (wzorce) i instancji (realizacje) – Rys. 4 i 6.

Do składowania danych o procesach biznesowych w hurtowni procesów (Rys. 4) zaproponowano wykorzystanie różnych struktur danych. Skład zunifikowanych wzorców procesów bazuje na modelu referencyjnym procesów (WFMC – RDB – model relacyjny). Analityczne składowanie danych o procesach zawarty jest w plikach binarnych, opisujących głów-

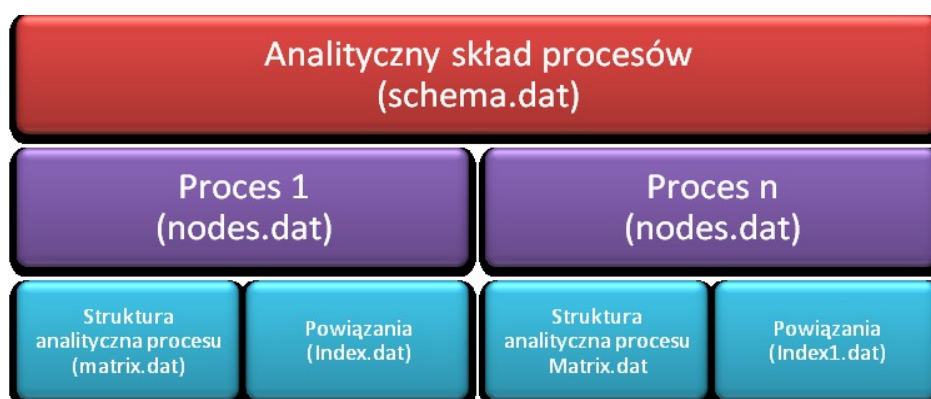


nie procesy w postaci grafów i sieci o postaci macierzowej ze względu na wymagania wydajnościowe przetwarzania analitycznego.



Rys. 4. Schemat składu danych hurtowni procesów

Fig. 4. Data store schema of warehouse of processes



Rys. 5. Schemat składu danych struktury analitycznej procesu

Fig. 5. Data store schema of analytical structure of process

Głównym elementem analitycznego składu procesów (Rys. 5) jest struktura zawierająca dane o gromadzonych procesach biznesowych (schema.dat). Pojedynczy proces biznesowy opisany jest głównie przez zestaw danych podstawowych (nazwa procesu, typ standardu, wersja, data utworzenia, lista elementów składowych procesu, ścieżka do katalogu z plikami zawierającymi dane procesu itp.).

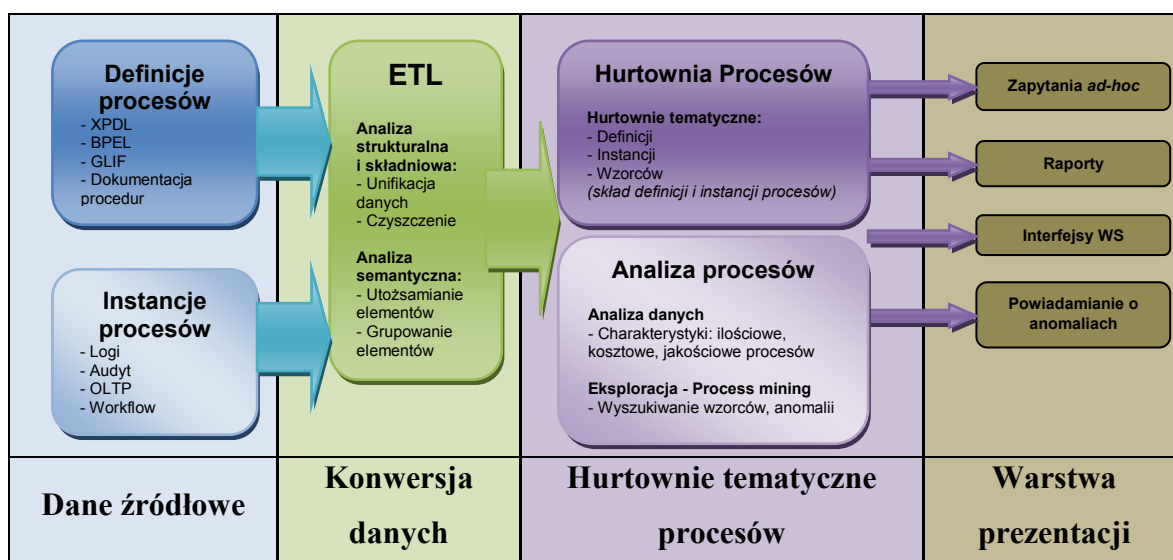
Dane każdego procesu biznesowego przechowywane są w oddzielnych strukturach, które zawierają:

- strukturę analityczną procesu (matrix.dat) zawierającą m.in. macierz sąsiedztwa elementów procesu biznesowego;
- dane o powiązaniach pomiędzy elementami i indeksy zawierające dodatkowe zależności i atrybuty elementów procesu biznesowego.

Głównym zadaniem stawianym przed hurtownią procesów, jako jednym z narzędzi wspomagających podejmowanie decyzji, jest możliwość przeprowadzania analiz zgromadzonych zasobów, takich jak:

- w obszarze analizy definicji procesów:
  - podpowiadanie wzorców definicji procesów podczas modelowania na podstawie użytych już elementów procesu i zbioru wzorców,
  - miary podobieństwa semantycznego definicji procesu lub jego fragmentów,
  - miary podobieństwa strukturalnego definicji procesu lub jego fragmentów,
  - walidacja poprawności definicji procesu metodą analizy grafu utworzonego na podstawie tej definicji,
  - analiza budowy definicji procesu metodą analizy grafu utworzonego na podstawie tej definicji:
    - stopień szczegółowości definicji zdefiniowany jako: czytelność – liczba wierzchołków pojedynczego grafu i podgrafów składowych, złożoność – liczba zasobów przydzielonych do grafu lub grafu składowego,
    - czy może wystąpić zakleszczenie, kiedy dwie aktywności będą nawzajem czekać na swoje zakończenie?
    - czy wszystkie przypadki zostały uwzględnione w definicji?
- w obszarze analizy wykonania procesów (instancji):
  - tzw. process mining:
    - odkrywanie nowych definicji procesów na podstawie instancji wykonania procesów,
    - odchylenia wykonania od definicji procesu,
    - wyszukiwanie wzorców oraz rzadkich i częstych sekwencji zdarzeń,
  - analizy ilościowe:
    - czasów wykonania,
    - opóźnień,
    - wykorzystania zasobów,
    - ile instancji może być powołanych w ciągu godziny?
    - jaki jest średni czas przepływu?
    - ile dodatkowych zasobów może być potrzebnych?
    - jak dużo zasobów określonego typu jest zużywanych w ciągu dni roboczych?
    - itp.,
  - miar podobieństwa sekwencji zdarzeń,
  - adekwatności procesu w stosunku (czy dany proces w efekcie daje oczekiwane wyniki?),
  - identyfikacji problemów (np. zatorów) i ich przyczyn.

Wymienione przypadki są tylko przykładami wielu różnych i możliwych analiz dotyczących procesów, jakie można zrealizować i w efekcie wykorzystać do polepszenia jakości w procesie podejmowania decyzji.



Rys. 6. Ogólny model architektury hurtowni procesów

Fig. 6. Basic architecture model of warehouse processes

Na Rys. 6 został zaprezentowany ogólny schemat architektury hurtowni procesów wraz z narzędziami do analizy i eksploracji procesów. Do głównych komponentów systemu należy zaliczyć: mechanizm ETL, znany z hurtowni danych i rozszerzony o mechanizmy unifikacji strukturalnej i semantycznej procesów i instancji, hurtownię procesów ze składem procesów ukierunkowanym na efektywne przetwarzanie danych o procesach (np. postać macierzowa opisu procesu jako grafu) oraz narzędzia do analizy i eksploracji procesów.

W poprawnej realizacji analiz procesów bardzo istotnym etapem jest ETL, w którym procesowi unifikacji podlegają nie tylko struktura procesu, ale również semantyka opisu poszczególnych procesów. Do unifikacji semantycznej procesów można wykorzystać rozwiązania z obszaru informatycznych zastosowań semantyki i ontologii, gdzie na potrzeby unifikacji semantycznej procesów projektant musiałby przygotować właściwą ontologię [6].

Jednym z problemów napotykanym w podejściu procesowym jest zapewnianie mechanizmów integracji danych przekazywanych pomiędzy poszczególnymi wywołaniami usług. Bardzo często konstruowane środowiska wymagają dostarczania odpowiednich mechanizmów logiki biznesowej, odpowiadających za translację danych pomiędzy poszczególnymi etapami realizacji procesu biznesowego. Prezentowane podejście zakłada wykorzystanie modeli semantycznych do wsparcia opisu, a w konsekwencji – integracji danych współdzielonych w procesowym środowisku wykonawczym. Budowa tego typu modeli zwykle wykorzystuje dedykowane języki deklaratywne, oparte na formalizmach logiki pierwszego rzędu (ang. First-Order Logic) lub logiki opisowej (ang. Description Logic) [7]. Podejście takie jest szczególnie uzasadnione w przypadku integracji środowisk SOA, które w swoim założeniu mają oferować mechanizmy dopasowania semantyki i syntaktyki przetwarzanych danych. O ile transformacje struktur danych są zwykle problemem inżynierskim, o tyle semantyka często przysparza więk-

szych problemów, co jest związane z identyfikacją kontekstu użycia wartości słownikowych, typów wyliczeniowych itd. W większości przypadków specyfikacja modelu semantycznego wykorzystuje język Ontology Web Language (OWL), będący standardem W3C.

### 3.1. Unifikacja procesów – semantyczny ETL

Jednym z ciekawych zagadnień badawczych, związanych z opracowaną metodą, jest zapewnienie mechanizmów ujednoczenia semantyki danych przekazywanych pomiędzy poszczególnymi procesami lub ich składnikami. Zagadnienie to można rozpatrywać na trzech poziomach: technicznym, syntaktycznym i semantycznym. O ile dwa pierwsze poziomy są dość dobrze opisane w literaturze i mają swoje odzwierciedlenie w narzędziach informatycznych, o tyle semantyka bardzo często jest jeszcze poziomem wymagającym ingerencji człowieka i ręcznej budowy procesu z wykorzystaniem narzędzi. Istnieją już pierwsze rozwiązania oferujące semantyczne mechanizmy wywołania usług OWL-S [13] i WSMO [12], jednak rozwiązania te są w formie prototypów i demonstratorów technologii.

W trakcie opracowania opisywanych mechanizmów hurtowni procesów, została wyraźnie zaznaczona potrzeba opracowania automatycznych mechanizmów transformacji danych przekazywanych pomiędzy poszczególnymi procesami. Zakładając, że procesy są realizowane w heterogenicznych środowiskach uruchomieniowych, bardzo użytecznym zagadnieniem staje się integracja danych.

Celem nadrzędnym jest więc zapewnienie takiego mechanizmu, który przez dostarczenie opisu semantyki danych zautomatyzuje przekształcanie danych wyrażonych w semantyce procesu A do semantyki procesu B. Warto przy tym zwrócić uwagę, że w samym procesie nie zależy nam na analizowaniu etykiet danych, a raczej ich związków z pozostałymi elementami modelu semantycznego. Zastosowanie modeli semantycznych wyrażonych w językach ontologicznych zapewnia część poszukiwanych cech takiego rozwiązania, oferując:

- możliwość wyrażania złożonych związków między elementami modelu,
- możliwość wyrażania równoznaczności pojęć i instancji,
- założenia otwartego świata (ang. Open World Assumption) – wprowadzając specyficzny sposób interpretacji niewiedzy w modelu.

Ważną rolę w definicjach modeli semantycznych odgrywają możliwości języka OWL i mechanizmów wnioskujących w ramach automatycznej weryfikacji spójności modelu pojęciowego oraz modelu instancji, klasyfikacji instancji, sprawdzenia tożsamości instancji (badanie, czy dwie instancje są tym samym bytem). Obecny rozwój języków opisu i algorytmów wnioskowania dąży do zapewnienia dużej ekspresyjności wyrażania konstrukcji modelu semantycznego, jednocześnie zapewniając efektywne mechanizmy wnioskujące. Warto w tym momencie zwrócić uwagę, że modele semantyczne wymagają obecnie dużych mocy do przetwarzania, co jest wynikiem dużej pamięciochłonności opisów semantycznych i ich we-

wewnętrznej reprezentacji wewnątrz środowisk przetwarzania tych modeli (JENA, Protege, Neon Toolkit etc.).

Ideologia opisu semantycznego dąży do automatyzacji przetwarzania danych z uwzględnieniem kontekstu, w którym zostały umieszczone. Przez automatyzację rozumie się przede wszystkim budowę automatów przetwarzających zasoby sieci WWW. Środowiska wieloagentowe właśnie są formą takiego przetwarzania, a semantyka jest obecna w komunikacji zarówno pomiędzy agentami, jak i bezpośrednio przy interakcji z zasobami sieci. Wraz z językami opisu powstała potrzeba opracowania standardów języków zapytań operujących na tym opisie. Języki te w swojej idei są językami deklaratywnymi i odzwierciedlają w wyniku struktury grafowe, reprezentujące zawartość pobraną z modelu na podstawie określonych kryteriów. Aplikacja zgodna z ideą Semantic Web potrafi wyszukiwać interesujące użytkownika dane, łącząc je znaczeniowo, a co najważniejsze – wywnioskowując kontekst. Umożliwia to zwiększenie efektywności przeszukiwania zasobów, weryfikację postaci metaopisu, a przede wszystkim zwiększenie trafności zapytań.

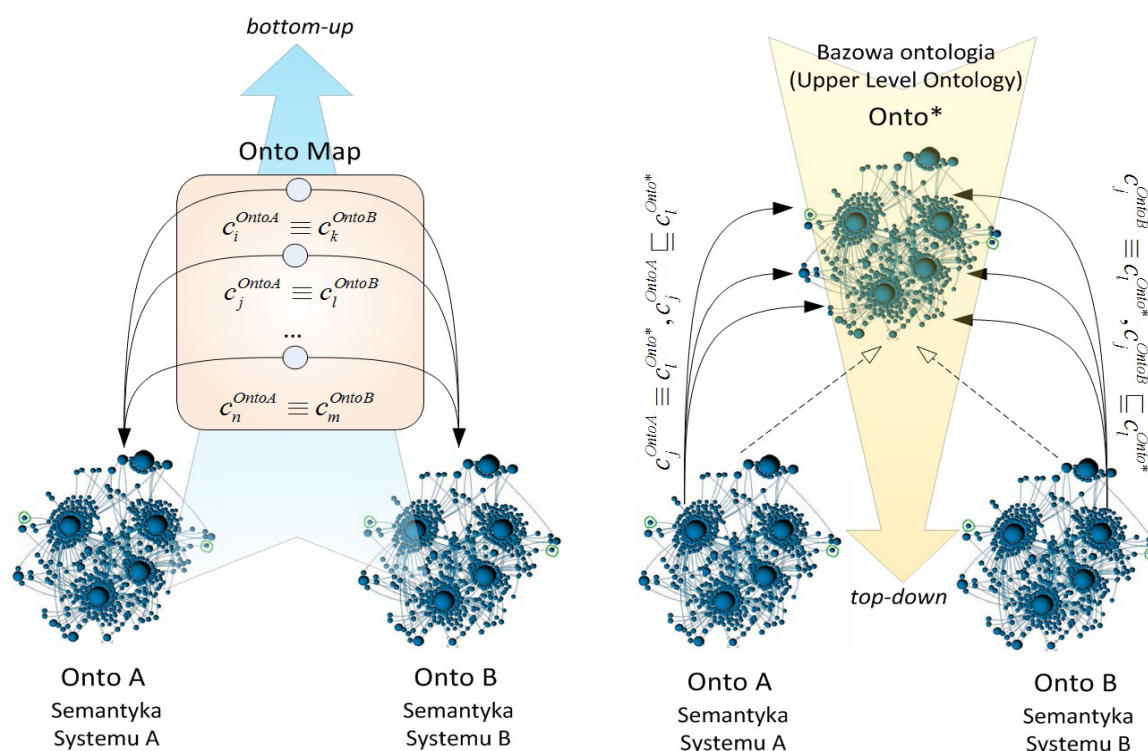
Interoperacyjność procesów często rozpatrywana jest z punktu widzenia zapewniania łączy technicznych i poziomu syntaktycznego danych. Semantyka w zapewnianiu interoperacyjności systemów skojarzona jest z różnym odwzorowaniem znaczenia pojęć wykorzystywanych w kilku systemach. Interoperacyjność na poziomie danych bardzo często zapewniana jest przez dedykowane mechanizmy migracji danych. Trzeba przy tym zwrócić uwagę na to, że zwykle migracja danych przeprowadzana jest w określonych odstępach czasu i nie wykonuje się jej jako procesu ciągłego. Przygotowanie mechanizmów migracji, oprócz wyzwań polegających na zapewnieniu łącza technologicznego i dopasowania struktur danych, wymaga skrupulatnej analizy znaczenia migrowanych danych w systemach docelowych. Semantyka danych często powiązana jest z mechanizmami logiki aplikacji, która interpretuje te dane na potrzeby implementowanej logiki biznesowej. Niniejsze rozważania wskazują na możliwość przesunięcia ciężkości tego zadania na odpowiednio przygotowany opis dostarczany wraz z danymi. Niewątpliwą zaletą opisywanego podejścia jest elastyczność modyfikacji tych elementów właśnie w zakresie dostarczania interpretacji

Wykorzystanie specyficznych konstrukcji języka OWL, w którym są wykonywane opisy zbiorów danych, pozwala na budowę pomostów semantycznych (ang. *semantic bridge*) [8]. Integracja zbiorów danych z wykorzystaniem modeli semantycznych ma na celu zastosowanie generycznych mechanizmów oprogramowania z wykorzystaniem dostarczonych opisów syntaktycznego i semantycznego. Różnica zawarta w tym podejściu jest znacząca w stosunku do dostępnych na rynku rozwiązań, albowiem zbudowany mechanizm integrujący zbiory danych wymaga jedynie modeli i reguł transformacji. Dla ułatwienia procesu wytwarzania takich modeli i mapowań jest często wspierane za pomocą przeznaczonych do tego celu środowisk i narzędzi IDE. Warto przy tym zwrócić uwagę na jeszcze jeden ważny aspekt: jeżeli

w wybranych dwóch systemach przygotowane zostały opisy semantyczne dla danych w nich przechowywanych, to proces obustronnej migracji danych może być przeprowadzony z wykorzystaniem pomostów semantycznych na poziomie modeli dziedzinowych. Schematycznie proces ten może być opisany następująco:

1. Dane systemu A transformowane są do postaci semantycznej z wykorzystaniem warstwy mapującej systemu A.
2. Mechanizmy wnioskujące, w połączeniu z definicjami pomostów semantycznych, transformują dane (instancje) wyrażone w semantyce systemu A do danych wyrażonych w semantyce systemu B.
3. Instancje danych, wyrażone w semantyce systemu B, przy użyciu transformat systemu B są spłaszczane do struktur danych systemu B.

Opisana procedura w konsekwencji pozwala na odwracanie procesu migracji oraz jego dalszą rozbudowę o nowe, heterogeniczne źródła danych. Wymogiem koniecznym jest jednak dostarczenie (wraz z zasobami) danych modelu semantycznego oraz mapowań pomiędzy składowymi warstwami syntaktyki i semantyki.



Rys. 7. Strategie mapowania danych w dwóch, odrębnych semantycznie modelach, reprezentowanych przez ontologie

Fig. 7. Mapping strategies of two distinct, semantically separate models

Pomosty semantyczne są konstrukcjami definiującymi tożsame elementy modeli. W tym celu używa się konstrukcji równoważności pojęć (ang. Concept Equivalence Axiom) oraz równoważności relacji (ról) (ang. Concept Equivalence Role). Konsekwencją użycia tych zapisów w modelu jest przyzwolewanie mechanizmom wnioskującym na postrzeganie instancji danych

pojęć jako równoważnych sobie, a więc dane wyrażone w jednej semantyce automatycznie będą interpretowane jako dane o znaczeniu definiowanym przez drugi model semantyczny.

#### 4. Podsumowanie

Przedstawione w niniejszym artykule aspekty problematyki modelowania, realizacji, gromadzenia, monitorowania i analizy ścieżek klinicznych jako procesów workflow, a w szczególności koncepcji procesów dynamicznych i architektury hurtowni procesów wraz z jej elementami składowymi (m.in. modele, standardy, składy procesów, semantyczny ETL), są przedmiotem prac badawczych prowadzonych w Instytucie Systemów Informatycznych WCY WAT, a zapoczątkowane zostały projektem POIG.01.03.01-00-145/08 pt. „Modelowanie repozytorium i analiza efektywności informacyjnej wytycznych i ścieżek klinicznych w służbie zdrowia”. W kontekście opisanej koncepcji analizie poddane są również zagadnienia interoperacyjności systemów, metod wielowymiarowych badań jakości oraz unifikacji procesów. Celem prowadzonych prac oraz głównym założeniem jest takie powiązanie wszystkich wymienionych elementów, aby wspólnie stanowiły one realne wsparcie dla modelowania i definiowania procesów oraz ich wykorzystania do podejmowania decyzji na różnych poziomach i w zakresie różnych obszarów biznesu, nauki i życia. Przedmiotowa koncepcja w szczególności wspiera pełną i praktyczną implementację idei ścieżek klinicznych, co nie jest możliwe z wykorzystaniem standardowych metod i narzędzi. Warty podkreślenia jest fakt, iż jest ona oparta na powszechnie dostępnych standardach informatycznych (z zakresu zarządzania procesami, hurtowni danych, interoperacyjności), co sprawia, że jej „charakter” jest jak najbardziej praktyczny, a poszczególne elementy koncepcji mogą zostać zaimplementowane z wykorzystaniem funkcjonujących na rynku narzędzi.

#### BIBLIOGRAFIA

1. Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P.: *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.
2. Bliźniuk G.: *Ranking inicjatyw standaryzacyjnych i dobór zestawu standardów, kluczowych dla dalszych etapów realizacji projektu*. Raport z realizacji zadania 1 projektu POIG.01.03.01-00-145/08. Opracowanie wewnętrzne WAT, Warszawa 2009.
3. Todman Ch.: *Projektowanie hurtowni danych. Wspomaganie zarządzania relacjami z klientami*. [tłum.] Paweł Gonera. Helion, Warszawa 2011.

4. Chmielewski M., Gałka A.: Semantic battlespace data mapping using tactical symbols, [in:] Nguyen N. T., Katarzyniak R., Chen S.-M.: *Advances in Intelligent Information and Database Systems*. Springer, 2010.
5. Chmielewski M., Koszela J.: The concept of C4I systems data integration for planning joint military operations, based on JC3 standard. 2008.
6. Ehrig M., Sure Y.: *Ontology mapping – an integrated approach*. ESWS 2004. LNCS, Vol. 3053, Springer-Verlag, Greece 2004.
7. Gruber T. R.: *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. *International Journal of Human-Computer Studies*, Vol. 43, 1995, s. 907÷928.
8. Gzik T.: *Analiza rozwiązań informatycznych wykorzystywanych do wspierania modelowania procesów wspomaganie podejmowania decyzji medycznych*. Raport z realizacji zadania 1 projektu POIG.01.03.01-00-145/08. Opracowanie wewnętrzne WAT, Warszawa 2009.
9. Koszela J.: *Opracowanie oceny przydatności metod standaryzacji opisu planu wykonywania instancji procesów działalności w kontekście wytycznych i ścieżek klinicznych*. Raport z realizacji zadania 1 projektu POIG.01.03.01-00-145/08. Opracowanie wewnętrzne WAT, Warszawa 2010.
10. Gzik T., Kędzierski P., Koszela J.: *Hurtownie procesów i procesy dynamiczne jako narzędzia wspierające wspomaganie podejmowania decyzji medycznych*. Modelowanie i zastosowanie komputerowych systemów medycznych. Vizja Press&IT, Warszawa 2009, s. 50÷57.
11. Tukaj D.: *Ścieżki kliniczne – co to jest i jak je wytyczać?* *Ogólnopolski Przegląd Medyczny*, 2005, s. 9÷10.
12. Fensel D., Lausen H., Polleres A., Bruijn J., de Stollberg M., Roman D., Domingue J.: *Enabling Semantic Web Services: The Web Service Modeling Ontology*. 2007.
13. Liyang Y.: *Introduction to the Semantic Web and Semantic Web Services*. Chapman and Hall/CRC, USA 2007.

Wpłynęło do Redakcji 22 stycznia 2012 r.

## Abstract

Presented concept of warehouse of processes summarises our understanding on possible advances in the area of process modelling. After analysis of available solutions the need of new approach must be recognised as not all cases of information systems can be modelled using known processes-oriented approach. In our understanding introducing dynamic compo-



sition of processes delivers the advantage of execution and modelling flexibility as well as dynamically adapting new perception of given process. This approach has been applied to solve modelling issues in the clinical path problem. Accordingly to our observations and analysis we have identified a need for dynamically altering clinical path execution. In such case an existing process definition should be reorganised or even completely altered to cope with new medical staff requirements meeting new diagnostics and medication procedures. This case directly addresses the need of medical staff, which requires that the conducted medical treatment process for a given individual must be actively modified accordingly to the knowledge that has been gathered about this distinct medical case. Discussed method recognises also the need of unified understanding of data exchanged within the heterogeneous execution environments hosting process definitions. We have proposed a genuine approach to unifying and automatically integrating heterogeneous datasources in order to support data-flow within the distributed SOA environments hosting process execution.

### **Adresy**

Grzegorz BLIŹNIUK: Wojskowa Akademia Techniczna, Wydział Cybernetyki,  
ul. Kaliskiego 2, 00-908 Warszawa 49, Polska, grzegorz.blizniuk@gmail.com.

Mariusz CHMIELEWSKI: Wojskowa Akademia Techniczna, Wydział Cybernetyki,  
ul. Kaliskiego 2, 00-908 Warszawa 49, Polska, mchmielewski@wat.edu.pl.

Tomasz GZIK: Wojskowa Akademia Techniczna, Wydział Cybernetyki,  
ul. Kaliskiego 2, 00-908 Warszawa 49, Polska, tomasz.gzik@gmail.com.

Jarosław KOSZELA: Wojskowa Akademia Techniczna, Wydział Cybernetyki,  
ul. Kaliskiego 2, 00-908 Warszawa 49, Polska, jkoszela@wat.edu.pl.