Business Processes as Artifacts

Jianwen Su
University of California, Santa Barbara
The “Big Data” Report

- **Mckinsey Global Institute, June 2011:**
  Big data: The next frontier for innovation, competition, and productivity

- **MGI:** established in 1990 to develop deeper understanding of the evolving global economy

  **Mission:**
  To provide leaders in the commercial, public, and social sectors with the facts and insights on which to base management and policy decisions
From EXECUTIVE SUMMARY:
“The United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings.”
What big data can generate:

- **US health care**
  - $300 billion value per year
  - ~0.7 percent annual productivity growth

- **Europe public sector administration**
  - €250 billion value per year
  - ~0.5 percent annual productivity growth

- **Global personal location data**
  - $100 billion+ revenue for service providers
  - Up to $700 billion value to end users

- **US retail**
  - 60+% increase in net margin possible
  - 0.5–1.0 percent annual productivity growth

- **Manufacturing**
  - Up to 50 percent decrease in product development, assembly costs
  - Up to 7 percent reduction in working capital

**SOURCE:** McKinsey Global Institute analysis
A biz process is a set of one or more linked activities (automated or manual) that collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships.

Obtaining a Permit
BP Management Systems (BPMSs)

Software systems to manage and support (and control)
- biz models
- data (documents, files, …)
- enactments
- resources (including human)
- others (e.g. auditing)

BP “=” workflow in the wider sense

Traditional meaning of workflow in 80’s to early 90’s means task sequencing
Outline

- Challenges in Business Process Management
- Artifact-centric Modeling Approach
- EZ-Flow and Selected Technical Issues
- Conclusions
Vanda Group

- Developing workflow systems for regional banks, credit unions, provident funds, ...

- Est. 60% of the market excluding national banks

**Key obstacles:**

- Training (engineer liquidity)
- Repetition of work, labor intensive (could make more $ or ¥¥ and be more competitive)
- High maintenance cost

- developed workflow application domains
Hangzhou Housing Management Bureau

- Population: 8.7 millions

One division (~400 SMEs) deals with all real estate licenses, permits, titles, etc.

- 300,000 cases each year,
  ~500 workflow (types), 35% 1 day, 30% 7-9 days

*developing workflow application domains*
Hangzhou Housing Management Bureau

- Population: 8.7m

One division (~400 SMEs) deals with all real estate licenses, permits, titles, etc.

- 300,000 cases each year,
  ~500 workflow (types), 35% 1 day, 30% 7-9 days

[Reference models: 600+]

SAP

Haier CHINA

Suncorp

CNR China CNR Corporation Limited
Hangzhou Housing Management Bureau

- Population: 8.7m

One division (~400 SMEs) deals with all real estate licenses, permits, titles, etc.
- 300,000 cases each year,
  ~500 workflow (types), 35% 1 day, 30% 7-9 days
- Contractor/in-house development of workflow system(s)
  (¥¥ millions for in-house only)

Challenges:
- Manage changes (policy, environment, ...)
- Serious lack of automation for developing workflow application domains
Hospitals: RuiJin & Cottage

- Health care delivery:
  much of the $300 billion could be gained
- Treatment workflows can fundamentally improve health care quality

Falling far behind:
- No workflows, conflicting “workflows”
- “Shaky” IT infrastructures

- RuiJin has the largest IT team (40+FTEs) among all hospitals in Shanghai

new IT divide?

RuiJin Hospital
SHANGHAI JIAO TONG UNIVERSITY SCHOOL OF MEDICINE

Cottage
health system

wishful workflow application domains
Application and Research Challenges

- Lack of clear ways to combine various factors of workflows
- Lack of workflow technology to support a variety of essential functions
- Long tail phenomenon is a “holy grail”
- Application domains work in isolation
- Unifying holistic conceptual models
- Design and runtime support
- Reasoning, business “informatics”, process mining
- Interoperation
Outline

- Challenges in Business Process Management
- Artifact-centric Modeling Approach
- EZ-Flow and Selected Technical Issues
- Conclusions
The Challenge of BPM

- Business Strategy:
  - “Be more green”
  - “Use our differentiators”

- Business Goals:
  - Business Architecture
  - Business Optimization

Roles:
- High Executive
- High Manager
- Business Architect
- Solution Designer
A Business Component Map is a tabular view of the business components in the scope of interest.

### Business Competencies:
- Large biz area with characteristic skills and capabilities

### Accountability Levels:
- **Directing**: Scope and intent of activity and decision-making
- **Controlling**: Part of enterprise that has potential to operate independently
- **Executing**: Business administration

#### Business Component Map

<table>
<thead>
<tr>
<th>Business Administration</th>
<th>New Business Development</th>
<th>Relationship Management</th>
<th>Servicing &amp; Sales</th>
<th>Product Fulfillment</th>
<th>Financial Control and Accounting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Planning</td>
<td>Sector Planning</td>
<td>Account Planning</td>
<td>Sales Planning</td>
<td>Fulfillment Planning</td>
<td>Portfolio Planning</td>
</tr>
<tr>
<td>Business Unit Tracking</td>
<td>Sector Management</td>
<td>Relationship Management</td>
<td>Product Management</td>
<td>Fulfillment Planning</td>
<td>Compliance</td>
</tr>
<tr>
<td>Staff Appraisals</td>
<td>Product Management</td>
<td>Credit Assessment</td>
<td>Sales Management</td>
<td>Fulfillment Planning</td>
<td>Reconciliation</td>
</tr>
<tr>
<td>Staff Administration</td>
<td>Product Directory</td>
<td>Credit Administration</td>
<td></td>
<td></td>
<td>Customer Accounts</td>
</tr>
<tr>
<td>Production Administration</td>
<td>Marketing Campaigns</td>
<td></td>
<td></td>
<td></td>
<td>General Ledger</td>
</tr>
</tbody>
</table>
The Challenge of BPM

Business Strategy
• “Be more green”
• “Use our differentiators”

High Executive

High Manager
Business Architect
Solution Designer

Business Goals
Business Architecture
Business Optimization

Business Operations

Customers

Employees

Partners

Resources

IT
**Common Model at IT Level:**

An **Activity Flow** is a (typically) graph-based specification of how activities/processes are to be sequenced.
The Challenge of BPM

- **Business Strategy**
  - “Be more green”
  - “Use our differentiators”

- **Operations need to be**
  - Faithful
  - Measurable
  - Flexible

- **Speak in terms of**
  - “Functional Decomposition”
  - “Business Components”
  - “Workflow”
  - “Process centric”
  - “Activity-flow”

- **Hard to Communicate!!**
Common Model at IT Level:

An Activity Flow is a (typically) graph-based specification of how activities/processes are to be sequenced

- Data and business objects are typically an afterthought
- Hard for stake-holders to communicate about the big picture
  - People “see the trees but not the forest”
  - Overall process can be chaotic – Cf. “staple yourself to a customer order”
- Hard to manage versions
  - E.g., evolution, re-use, generic workflow with numerous specializations
Typical Biz Process Modeling

- A bookseller example: Traditional control-centric models

Fill Shopping Cart → ID Customer → Shipping Preference → Payment Information → Confirmation → Archive
Typical Biz Process Modeling

- A bookseller example: Traditional control-centric models
- Multiple steps needed for each activity

In practice, 100s to 1000s of nodes

Hard to reason, find useful views: missing data
BP Analytics (Biz Intelligence)

- Extract-Transform-Load

Data Warehouse

- cust_db
- inventory
- catalog

Analysis

Transactions

Biz Process is missing!
Why We Should Look for a Unifying Model

Good models go beyond description – they support action

- Selecting the right model for the job matters

Example: “Game of 15”
Winner: First one to reach exactly 15 with any 3 chips

First model - A is [ ] and B is [ ] - what is B’s move?

Second model - [ ] [ ] [ ] - B’s move is 6!

Can we find a “model” of business operations that is
- Useful & natural for the business level stake-holders to use
- Useful & natural for mapping to the IT infrastructure
Data Management In the Infancy (60’s)

- Driving applications: inventory control, financial data management

- The key to the success: automation

- Query
  - File structures (indexes, …)
  - Labor intensive
  - COBOL program

- Logical data model

- Desirable
  - have to deal with

- By hand
A Fundamental “Theorem” of Databases

- Physical data independence allows us to focus only data management issues.

**Diagram:**

- SQL
- Logical data model
- Physical organization (files, pages, indexes, ...)

**Hierarchy:**

- Conceptual
- Physical
Future of BPM?

- Automate 's

Changes to system

Changes

process model

data model

system (model)
(databases, services, workflows, resources)

Reuse concepts, tools, techniques developed in CS

First step: a single conceptual model for biz processes
  - both data and processes are 1st class citizens
Outline

- Challenges in Business Process Management
- Artifact-centric Modeling Approach
- EZ-Flow and Selected Technical Issues
- Conclusions
BP Modeling: Data Exclusion to Data Centricity

- Data exclusive models focus on activity flow and management
  - WfMC, BPMN, …

- Incorporating data as views complements well (but separate from) activity views
  - UML (object modeling and activity diagrams)

- Executable models integrate data and activities with low level of abstraction
  - BPEL

- Recent data-centric approaches treat both data and activities “equally” in a more uniformed manner
  - Biz artifact-centric, form-based, spreadsheet-based
Business Artifacts

- A business artifact is a key conceptual business entity that is used in guiding the operation of the business
  - fedex package delivery, patient visit, application form, insurance claim, order, financial deal, registration, …
  - both “information carrier” and “road-maps”

- Very natural to business managers and BP modelers

- Includes two parts:
  - Information model: data needed to move through workflow
  - Lifecycle: possible ways to evolve
Example: Restaurant

**Artifacts**

- Guest Check
- Kitchen Order
- Receipt
- Cash Balance

**Activity**

- Create Guest Check
- Open GCs
- Add Item
- Prepare Receipt
- Pending KOs
- Prepare & Test Quality
- Recalculate Receipt
- Closed GCs
- Update Cash Balance
- Ready KOs
- Deliver
- Disagreed Receipts
- Paid Receipts
- Archived Receipts
- Archived GCs
- Closed GCs
- Archived GCs
- Archived KOs
- Closed GCs
- Archived GCs
- Cash Balance
- Archived KOs
- Closed GCs
- Archived GCs
- Cash Balance
- Archived KOs
- Closed GCs
- Archived GCs
- Cash Balance
- Archived KOs
- Closed GCs
- Archived GCs
- Cash Balance
- Archived KOs
Example: Restaurant

**Artifacts**

- **Guest Check**
- **Kitchen Order**
- **Receipt**
- **Cash Balance**

**Activities**

- Create Guest Check
- Open GCs
- Add KOs
- Prepare Receipt
- Pending KOs
- Prepare & Test Quality
- Cash Balance
- Payment
- Closed GCs
- Ready KOs
- Deliver
- Disagreed Receipts
- Recalculate Receipt
- Archived Receipts
- Archived GCs
- Archived KOs

Example: Restaurant Guest Check Kitchen Order Receipt Cash Balance

Artifacts

- GC
- KO
- RC
- CB

Update Cash Balance

Payment

Closed GCs

Pending Receipts

Disagreed Receipts

Recalculate Receipt

Archived Receipts

Archived GCs

Archived KOs

Create Guest Check

Add KOs

Prepare Receipt

Pending KOs

Prepare & Test Quality

Cash Balance

Payment

Closed GCs

Pending Receipts

Disagreed Receipts

Recalculate Receipt

Archived Receipts

Archived GCs

Archived KOs
Case Study: IBM Global Financing

[Chao, Cohn, et al BPM 2009]

- Finance HW, SW & services from IBM & others for clients
- IBM internal financing business w/ global reach
  - World’s largest IT financier w/ $38B asset base
  - Financing >$40B IT assets / year for last 3 years
  - 125K clients across >50 countries (9% of IBM profit)
- Business challenges
  - Operations tailored to mega-deals becoming too costly
  - Efficiency & cost control required global performance metrics
  - Country “silos” inhibited integration & annoyed clients
  - Current methods failed to produce end-to-end “tangible model”
  - Needed globally standard process w/ local variations
How the Artifact-Centric Approach Helped

- In a 3-day workshop with 15 business SMEs from IGF, a preliminary artifact design was created
  - Already useful to stakeholders from different regions as a common vocabulary
- 6 weeks of design refinements lead to final design
  - Enabled visibility into the global process and the regional variations: not possible before
  - A blueprint for transformation of IGF operations
    - VP roles assigned to pieces of top-level artifact model
- Current plan: automate the global-level artifact model
  - Anticipate significant improvement in efficiency
  - Plan to substantially augment the sales staff
Emerging Artifact-Centric BPs

Informal model [Nigam-Caswell IBM Sys J 03]

Systems: BELA (IBM 2005), Siena (IBM 2007), ArtiFlow (Fudan-UCSB 2010), Barcelona (IBM 2010)

Formal models
- State machines [Bhattacharya-Gerede-S. SOCA 07][Gerede-S. ICSOC 07]

Specification of artifact lifecycles

Artifacts (Info models)
Outline

- Challenges in Business Process Management
- Artifact-centric Modeling Approach
- EZ-Flow and Selected Technical Issues
- Conclusions
Artifact-Centric BPMSs

@IBM:
- Declarative models
- Semantics (U Rome)
- Analysis (UCSD)
- Workflow views (lenses)

@UCSB
in collaboration with
IBM, U Rome, Fudan, …
Declarative Biz Processes

Variation of [Bhattacharya-Gerede-Hull-Liu-S. BPM 07]
Artifact Classes

- An artifact class consists of:
  - a finite set of attributes, of type $U$ or artifacts IDs
  - a finite set of states, initial and final states (transitions not defined)

- An artifact is a pair:
  - a mapping from attributes to $U \cup \text{IDs} \cup \{\perp\}$
  - a state

**GuestCheck Artifact**

<table>
<thead>
<tr>
<th>GCID</th>
<th>date</th>
<th>time</th>
<th>Name</th>
<th>KOID</th>
<th>table#</th>
<th>TOTAL</th>
<th>Payment</th>
<th>ptime</th>
</tr>
</thead>
</table>

**States:**
- Waiting for table
- Seated
- Ordered
- Delivered
- Completed
EZ-Flow: Procedural Biz Processes

- Each biz process has a core artifact (class)
  - Business data (object) + enactment
  - Event driven
  - Similar notion in recent GSM model from IBM
EZ-Flow Engine

```
... | e_3 | e_1 | e_2 |

event queue
```

```
exec(T_2, PAF01)
```

```
exec(T_3, PAF05)
```

```
perform T_2
```

```
perform T_4
```

```
perform T_3
```

task performer:

handles data wrapping and service wrapping
EZ-Flow and Research Problems

Biz process modelers, administrators

Dashboard

Runtime Query Engine
Biz Process Optimizer
Resource Registry
Anomaly Handler
Execution Engine
Entity (Artifact) Manager
Worklist Manager

System data store

External databases
Applications
Human performers

Enactment events (new, abort, …)
Workitem & application events
EZ-Flow and Research Problems

Biz process modelers, administrators

Runtime Query Engine

Biz Process Optimizer

Resource Registry

Anomaly Handler

Execution Engine

Entity (Artifact) Manager

Worklist Manager

System data store

runtime monitor

process ICs

System data store

External databases

Applications

Human performers

runtime monitor

process ICs

System data store

External databases

Applications

Human performers

Biz process modelers, administrators

data ICs

doneminace

automated construction

preserve data ICs

dynamic modification

Enactment events (new, abort, …)

Workitem & application events

exec. res. calculation

Biz process modelers, administrators

verifcation

Enactment events (new, abort, …)

EZ-Flow and Research Problems

Biz process modelers, administrators

- Runtime Query Engine
- Biz Process Optimizer
- Resource Registry
- Anomaly Handler
- Execution Engine
- Entity (Artifact) Manager
- Worklist Manager
- System data store
- External databases
- Applications
- Human performers

- Verification
- Runtime monitor
- Process ICs
- Dominance
- Automated construction
- Preserve data ICs
- Dynamic modification
- Exec. res. calculation
- Enactment events (new, abort, …)
- Workitem & application events
Changes in Biz Processes

- Reason for changes:
  - Policy/regulation change
  - Technology change
  - Environment change
  - User demand change
  - …

- The long tail phenomenon:
  - large number of cases of a small number of patterns
  - a small number of cases are mostly different

- BPMSSs must handle the latter more efficiently
Manage Changes

- Modify biz process model: time consuming, big effort

- Anticipate change at design time, and build flexibility in schema, e.g., [Gottschalk-van der Aalst-Jansen-Vullers-La Rosa 2008] [Hallerbach-Bauer-Reichert 2008]
  - limited options

- Declarative models: worklet [Adams-ter Hofstede-Edmond-van der Aalst 2006], LTL-based [van der Aalst-Pesic-Schonenberg 2009]
  - Data not included

- Runtime dynamic execution mechanism based on objects (task wrappers) [Redding-Dumas 2010]
  - Detached from process model, low abstraction

- Our approach: procedural process model with declarative changes, conservative extension [Xu-S.-Yan-Yang-Zhang 2011]
### Technical Approach

**Ingredient 1: artifact-centricity**

- Each biz process has a core artifact (class)
Technical Approach

- Ingredient 2: formal model (semantics) for execution

- Ingredient 3: declarative change specification
  - Four execution altering operators
  - Rules for applying the operators based on conditions

start → fetch → invoke

event_s  started  ready  done

(1)  (2)  (3)  (4)
Natural Disaster Victims on Green Channel

Express-SR: $A$

MAY skip SecondaryReview ON PAF
WHERE projectType="resettled"
New Fee Schedule for Low Incoming Housing

Affordable-Fee:
MUST REPLACE PaymentProcessing
BY AffordablePaymentProcessing ON PAF
WHERE SELF.projectType="affordable"
New Contractor Needs Prequalification

First-Timer:
MUST ADD Prequal BEFORE Prelim_Decision ON PAF
WHERE projectType="affordable" AND developerName NOT IN
SELECT developerName FROM PAF P
WHERE P.artifactId <> SELF.artifactId
AND P.projectType="affordable"
Insufficient Selling Space Need Re-Check

Re-eval:
MUST RETRACT FROM SecondaryReview
TO ReceivingApp-form ON PAF
WHERE SELF.cp.planArea <
( SELECT sum(P.sellingArea)
FROM PAF P
WHERE P.cp=SELF.cp
GROUP BY p.artifactId )
Mixed Procedural and Declarative Pays off

- Biz process = state machine lifecycle + change rules
- Modification rules conservatively extend workflow
  - Could be temporary, non-schematic
- Allows biz process to respond to situations with many more options:
  - # of “trace types” grow exponentially in # rules
- Performance estimates:
  - 9% labor savings for Real Estate Administration of Hangzhou (preliminary study)

[Xu-S.Yan-Yang-Zhang 2011]
EZ-Flow and Research Problems

Biz process modelers, administrators

Dashboard

Runtime Query Engine
Biz Process Optimizer
Resource Registry
Anomaly Handler
Execution Engine
Entity (Artifact) Manager
Worklist Manager
System data store
External databases
Applications
Human performers

Verification
Automated construction
Preserve data ICs
Dynamic modification
Runtime monitor
Internal ICs
Enactment events (new, abort, …)
Workitem & application events
Exec. res. calculation

Nanjing U/2012 Summer School
2012/07/24
Comparing Two Workflows

- Assuming they work on the same input-output (types)
- Can one workflow “simulate” the other

\[
\begin{align*}
W_1 & \text{Artifact (Info model)} \quad S \quad R \quad \text{if } C \text{ enable} \\
A & \quad + \quad S \quad + \quad \text{if } C \text{ enable} \\
W_2 & \text{Semantic Services (IOPEs)} \quad R' \quad \text{if } C' \text{ enable} \\
A' & \quad + \quad S' \quad + \quad \text{if } C' \text{ enable}
\end{align*}
\]
Why Comparison

Many reasons:
- Optimization (similar to comparing queries)
- Replacing part of workflow (reorganization)
- Updating workflow (evolution)
- Reusing workflow
- . . .
Workflow Dominance

If every input-output pair that can be produced by \( W_1 \) can also be produced by \( W_2 \),

\[
W_1 \leq W_2
\]

if every input-output pair that can be produced by \( W_1 \) can also be produced by \( W_2 \)

- **Note:**
  - their temporary data can be very different
  - services are different; rule sets are different
  - services may be done by human
Performance Policies

- A performance policy $\pi$ is a function that assigns each service $\sigma$ a multi-valued function over $U$

$$\pi(\sigma) : x \rightarrow \{x+1, x+2\}$$

- Since the “flow” is fixed, the choice of a performance policy determines how the workflow would perform
  - E.g., given an input, a workflow can execute and generate an output

- Classes $\Pi$ of performance policies $\pi$
  - Absolute (ABS): $\pi(\sigma) = U \times U$
  - Fixed choice: $\pi(\sigma)$ is some single-valued function
Definition of $(k-)$Dominance

Fix a class of performance policies $\Pi$.

If for each performance policy $\pi_1$, there is a performance policy $\pi_2$, such that every input-output pair produced by $W_1[\pi_1]$ in at most $k$ steps can also be produced by $W_2[\pi_2]$ in at most $k$ steps.

$W_1 \leq^k_{\Pi} W_2$
Capturing Workflow Under ABS

Key Lemma:

\( W \) : a workflow with service pre- and post-conditions and rule conditions expressed in FOL with equality

\( k \) : a positive integer

Then

there is a FOL formula \( \varphi(k, W) \) that characterizes the set of all input-output pairs produced by \( W[\text{ABS}] \) in at most \( k \) steps
Results on \((k)\)-Dominance

- Absolute \(k\)-dominance is decidable but dominance is undecidable:
  1. \((\mathbb{Z}, +, <)\), integers with additions
  2. \((\mathbb{Q}, +, <)\), rational numbers with additions
  3. \((\mathbb{R}, +, \cdot, <)\), real numbers with additions and multiplications (the real closed field)

- Absolute dominance is undecidable:
  1. \((\mathbb{Z}, <)\), integers with discrete order
  2. \((\mathbb{Q}, <)\), rational numbers with dense order
  3. \((\mathbb{R}, <)\), real numbers with dense order

[Calvanese-De Giacomo-Hull-S. ICSOC 09]
EZ-Flow and Research Problems

Biz process modelers, administrators

Dashboard

Runtime Query Engine

Biz Process Optimizer

Resource Registry

Anomaly Handler

Execution Engine

Entity (Artifact) Manager

Worklist Manager

System data store

External databases

Applications

Human performers

Enactment events (new, abort, …)

Workitem & application events

Verification

Automated construction

Preserve data ICs

Dynamic modification

Runtime monitor

Process ICs

Runtime

Calculation

Executing resource calculation

Nanjing U/2012 Summer School

2012/07/24
Synthesis Problem

- Given a goal and a set of services, construct a set of rules so that every execution satisfies the goal.

\[
\text{if } C \text{ enable } \ldots
\]

[Artifact (Info model)] + [Semantic services (IOPEs)] + [Goal (FO)]

[Fritz-Hull-S. ICDT 09]
(restricted to single artifact, first-order goals)
Artifact Schema

- An artifact schema is a finite set $A$ of attributes.
  An artifact of $A$ is a mapping from $A$ to $U \cup \{\bot\}$

- Assume a set of initial attributes $A_{init} \subset A$

- An artifact is $B$-completed, $B \subset A$, if it is defined on all attributes in $B$
  - “input” artifacts are $A_{init}$-completed
Semantic Services (Tasks)

- A semantic service over $A$ is a tuple $(\sigma, R, W, \pi, \rho)$, where
  - $\sigma$: service name
  - $R, W$: finite sets of (resp., read, write) attributes
  - $\pi, \rho$: quantifier-free formulas (pre- and post-condition, resp.) over $R, R \cup W$, resp.
- allow $\text{DEF}(A)$ for an attribute $A$

- $o'$ is the result of executing $\sigma$ on $o$, $o \rightarrow o'$, if
  - $(o, o') \models \pi \land \rho$, and
  - frame conditions are satisfied
An Example Semantic Service

\[ 0 \leq A \leq 2 \quad \sigma \quad 0 \leq A < 1 \land 0 \leq B \]
\[ \lor \]
\[ 1 \leq A \leq 2 \land 1 \leq B \]
**Condition-Action Rules**

- A *condition-action rule* is an expression "**if** \( \varphi \) **enable** \( \sigma \)"
  - where
    - \( \varphi \) is a (quantifier-free) formula and
    - \( \sigma \) is a semantic service

- \( o' \) is the result of executing a rule \( r : \text{if } \varphi \text{ invoke } \sigma \text{ on } o, \rightarrow o' \), if
  - \( o \models \varphi \), and
  - \( o \rightarrow o' \)
Workflow Schema

- A workflow schema is a triple $W = (A, S, R)$
  - $A$ : artifact schema
  - $S$ : a finite set of semantic tasks
  - $R$ : a finite set of condition-action rules

- Denote $\longrightarrow$ the closure of $\bigcup_{r \in R} r$
The Synthesis Problem

Given $A$, $A_{\text{init}}$, a finite set $S$ of semantic tasks, a formula $\varphi$ over $A_{\text{final}} \subseteq A$, find a rule set $R$ such that

$$\text{if } o \rightarrow o', \text{ then } o' \models \varphi$$

$A_{\text{init}}$-completed $\quad A_{\text{final}}$-completed

$A$

$S$

Artifact (Info model) + Semantic services (IOPEs) + Goal (FO) $\rightarrow$ ? $R \models \varphi$

Condition-action
A Trivial Solution

- Just let $R = \emptyset$
- Need to revise the problem statement
Maximally Safe Ruleset

- A ruleset enables all executions that guarantee to satisfy the goal.
- Goal: $1 \leq B$

$$0 \leq A \leq 2 \quad 0 \leq A < 1 \land 0 \leq B \lor 1 \leq A \leq 2 \land 1 \leq B$$

- $1 \leq A \leq 2$: definitely good
- $0 \leq A < 1$: possibly good but can’t be sure
- Best we can do

$$\text{if } 1 \leq A \leq 2 \text{ enable } \sigma$$
Maximally Safe Ruleset With Exception

- A ruleset eagerly move dead-end executions to EXCEPTION status
- Goal: $1 \leq B$

$$0 \leq A \leq 2 \quad \sigma \quad 0 \leq A < 1 \land 0 \leq B \land 1 \leq A \leq 2 \land 1 \leq B$$

- $1 \leq A \leq 2$ : definitely good
- $0 \leq A < 1$ : possibly good but can’t be sure
- Be optimistic:

  if $0 \leq A \leq 2$ enable $\sigma$
  if $B < 1$ goto EXCEPTION
Pre-Conditions

- Given a semantic task \((\sigma, R, \mathcal{W}, \pi(x), \rho(\text{xy}))\), and a (sub-goal) condition \(\delta(\text{xy})\)
- A \(\forall\)-precondition of \(\sigma, \delta\) is a formula \(\epsilon(x)\) such that
  - \(\epsilon\) logically implies \(\pi\) and
  - \(\forall x (\epsilon(x) \rightarrow (\forall y \rho(\text{xy}) \rightarrow \delta(\text{xy})))\) holds
  \[\text{WP}^\forall(\sigma, \delta) : \text{weakest } \forall\text{-precondition}\]

- A \(\exists\)-precondition of \(\sigma, \delta\) is a formula \(\epsilon(x)\) such that
  - \(\epsilon\) logically implies \(\pi\) and
  - \(\forall x (\epsilon(x) \rightarrow (\exists y \rho(\text{xy}) \land \delta(\text{xy})))\) holds
  \[\text{WP}^\exists(\sigma, \delta) : \text{weakest } \exists\text{-precondition}\]
Weakest Pre-Conditions

- Given a semantic task \((\sigma, R, W, \pi(x), \rho(xy))\), and a (sub-goal) condition \(\delta(xy)\)

- The weakest \(\forall\)-precondition

\[
WP^\forall(\sigma, \delta) \equiv \pi(x) \land (\forall y \, \rho(xy) \rightarrow \delta(xy))
\]

useful for maximally safe ruleset

- The weakest \(\exists\)-precondition

\[
WP^\exists(\sigma, \delta) \equiv \pi(x) \land (\exists y \, \rho(xy) \land \delta(xy))
\]

useful for maximally safe ruleset with exception
Necessary Condition

Theorem:
If there exists an algorithm to find maximally safe rule sets, the FOL theory is decidable (for the context structure)
The Other Direction

- Invoke-once constraint: each semantic task is allowed to run once

Theorem:

Under the invoke-once constraint, if the FOL theory (of the structure) is decidable and admits quantifier elimination, then the maximally safe rule sets can be computed.
A Special Case: Dense Order \((Q, <)\)

- Goal and task conditions are quantifier free formulas
- Acyclic task invocation dependencies
- Each task writes one attribute

**Theorem:**

Computing Maximal Safe Ruleset is PSPACE-complete

- Key ideas: cell decomposition; reduction from QBF

- Acyclicity condition can be dropped

  [Hull-S. 2009] (in preparation)
Further Restrictions

- A constructive EXPTIME algorithm

- PTIME if \#needed attributes is bounded
Summary of Results

- Synthesis problem is harder than FO logic theory of the underlying structure
- Positive answer for special cases
  - Invoke once
  - Concrete algorithm for dense order domain: PSPACE-complete
EZ-Flow and Research Problems

Process/workflow modelers, workflow administrators

Dashboard

Runtime Query Engine
Workflow Optimizer
Resource Registry
Anomaly Handler
Execution Engine
Worklist Manager
Entity (Artifact) Manager
Workflow data store

Domiance
Automated construction
Preserve data ICs
Dynamic modification

Verification
Runtime monitor

Workflow ICs

Enactment events (new, abort, ...)
Workitem & application events
Exec. res. calculation

External databases
Applications
Human performers
An Example Workflow - EzMart

- Traditional workflow specifications
  - Centered on control flow
  - Data flow is embedded in workflow executions

[X. Liu-S.-Yang, 2011]
Data integrity constraints

In data schema

- key, foreign key, candidate key \texttt{UNIQUE}, not-null

On attribute content

- Order: qty>0; Ship: from \texttt{\neq} addr

Business specific constraints

- Status: order cannot be canceled or returned when there is an associated shipment not finished
- …
GSM: A Declarative Workflow Language

Customer
- Register
- Pay
- Checkout
- Order

Order
- Create
- Pay
- Further action

Ship
- Prepare
- Ship
- Deliver report

Inventory
- Initiate
- Sell
- Update by manager

GSM: A Declarative Workflow Language

Customer
- Register
- Pay
- Checkout
- Order

Order
- Create
- Pay
- Further action

Ship
- Prepare
- Ship
- Deliver report

Inventory
- Initiate
- Sell
- Update by manager
Guard Injection

- Intuition: calculate and inject weakest precondition
- GSM: guard-stage-milestone by IBM

Guard the action by condition: \( \text{checkout.qty} > 0 \)

\[ \kappa_{\text{attr}} = \forall \text{oid}, \ldots \text{Order(oid).qty} > 0 \]

\[ \text{qty} := \text{checkout.qty} \]

...
Conservative Injection

Order

- If there is a shipment associated and is not finished
  - `custsuppreply.ostat = CANCEL`, violated
  - `custsuppreply.ostat = CANCEL`, consistent

- Injection to further_action is FALSE
Result

- The injection is
  - Sound: strong enough to block potential violations
  - Conservative complete: weak enough to allow all possible updates that preserves the constraints in conservative manner

```plaintext
Order

checkout.qty = 0 选址
checkout.qty = 10
checkout.custid = cust001

ostat := custsuppreply.ostat
qty := checkout.qty;
custid := checkout.custid; ...
```

```
invokes custsupp
```
EZ-Flow and Research Problems

Biz process modelers, administrators

Dashboard

Runtime Query Engine
Biz Process Optimizer
Resource Registry
Anomaly Handler

Execution Engine

Entity (Artifact) Manager
Worklist Manager

System data store
External databases
Applications
Human performers

Enactment events (new, abort, …)
Workitem & application events

execution results

dynamic modification

preserve data ICs
dominance

automated construction

runtime
monitor

verification
Given a biz process and a goal, do all executions of the workflow satisfy the goal?

Verification Problem

- Artifacts (Info models)
- Semantic services (IOPEs)
- Condition-action

if \( C \) enable

\[ ? \models \phi \]

[Bhattacharya-Gerede-S. SOCA 07] [Gerede-S. ICSOC 07]
[Bhattacharya-Gerede-Hull-Liu-S. BPM 07]
[Deutsch-Hull-Patrizi-Vianu ICDT 09]
[Vianu ICDT 09]
Summary of Results

- An artifact system $W = (\Gamma, S, R)$
  - artifacts, services, rules

- Ad hoc properties, restricted to defined-ness
  - Completion: Does $W$ allow a complete run of an artifact?
  - Dead-end: Does $W$ have a dead-end path?
  - Attribute redundancy: Does $W$ have a redundant attribute?

Undecidable in general, PSPACE if no artifact creation, intractable for monotonic workflows

[Bhattacharya-Gerede-Hull-Liu-S. BPM 07]

- Temporal properties: LTL(FO) for guarded artifact schema
  - complete in PSPACE

[Deutsch-Hull-Patrizi-Vianu ICDT 09]
**EZ-Flow and Research Problems**

System data store

Runtime Query Engine

Biz Process Optimizer

Resource Registry

Anomaly Handler

Entity (Artifact) Manager

Worklist Manager

Dashboard

**Biz process modelers, administrators**

**Verification**

**Runtime Monitor**

**Process ICs**

**Dynamic Modification**

**Preserve Data ICs**

**Automated Construction**

**Dominance**

**External databases**

**Applications**

**Human performers**

**Enactment events** (new, abort, ...)

**Workitem & application events**

**exec. res. calculation**
Outline

- Challenges in Business Process Management
- Artifact-centric Modeling Approach
- EZ-Flow and Selected Technical Issues
- Conclusions
Conclusions

- Biz process modeling: a foundation for many BPM issues
  - Many challenges: “old” and new
  - Data-centric or data aware approaches promising
- Systematic exploration provides a good setting for the study
  - First step in a long march
- Similar to mySQL, will “myBPM” be on the horizon?
Acknowledgements

- **UCSB**: Cagdas Gerede, Esra Kucukoguz, Yutian Sun
- **IBM**: Rick Hull, Kamal Battacharya, Rong (Emily) Liu
- **Fudan U (China)**: Liang Zhang, Wei Xu, Yanguang Cheng, Haihuan Qin, Jiehui Li, Yi Lu
- **Christian Fritz (U Toronto-USC)**
- **Diego Calvanese (U Bolzano)**
- **Giuseppe De Giacomo (U Rome)**
- **Jian Yang (Macquarie U)**
- **Xi Liu (Nanjing U while visiting UCSB)**