

Challenges in Service Modeling, Composition, and Analysis

Jianwen Su

University of California, Santa Barbara

Web Services: The Big Questions

Simplify and/or automate web service

■ Discovery

- ❖ What properties should be described?
- ❖ How to efficiently query against them?

■ Composition

- ❖ Specifying goals of a composition
- ❖ Specifying constraints on a composition
- ❖ Building a composition
- ❖ Analysis of compositions

■ Invocation

- ❖ Keeping enactments separated
- ❖ Providing transactional guarantees

■ Monitoring

- ❖ How to track enactments
- ❖ Recovering from failed enactments

An old slide from
SIGMOD tutorial
[Hull-S. 04
SIGMOD Rec 05]

Primary focus
of this tutorial

+ Data

Data for Services: A New Frontier

Jianwen Su

University of California, Santa Barbara

Outline

- **Application Needs**
- “Legacy” Services
- “Programmable” Services
- Data Encapsulating Services
- Research Challenges
- Conclusions

Real Estate Property Management



Quick Links

- Property Value Notice
- Marriage License
- Property Search
- FBN Search
- Property Tax Programs
- Polling Place/Sample Ballot Lookup
- Recent Home Sales
- Birth Certificate
- Marriage Certificate
- Death Certificate
- Grantor Grantee Index



Clerk-Recorder

- What we do
- How we can help
- Certificates
- Fictitious Business Name
- Grantor-Grantee Index

[Click here to view more details.](#)



Assessor

- What we do
- How we can help
- Property Value Notice
- Property Tax Programs
- Recent Home Sales

[Click here to view more details.](#)



Elections

- Polling Place/Sample Ballot Lookup
- What we do
- How we can help
- Upcoming Elections
- Voter Registration Lookup

A Housing Management Bureau

The screenshot shows the official website of the Hangzhou Housing Management Bureau (www.hzfc.gov.cn). The page features a navigation bar with categories like 'Home', 'Information Disclosure', 'Online Services', 'Housing Security', 'Housing Information', 'Market Information', 'Contract Filing', 'Trust Archives', 'Publicity', and 'Online Visitation'. A search bar is located at the top right.

The main content area is divided into several sections:

- 政务公开 (Government Openness):** Includes links for government information disclosure regulations, guidelines, and directories.
- 政务动态 (Government News):** Features 'People in Hangzhou' and 'Housing Online' tabs.
- 网上办事大厅 (Online Service Hall):** This section is highlighted with a red border. It contains:
 - 办事进度查询 (Service Progress Query):** Fields for '受理编号' (Application No.) and '证件号码' (ID No.), with a search button. It shows '今日预受理件数: 0件' (Today's pre-accepted cases: 0) and '累计预受理件数: 459783件' (Cumulative pre-accepted cases: 459,783).
 - 面向个人 (For Individuals):** Lists services such as '产权登记' (Property Registration), '房屋拆迁管理' (House Demolition Management), '白蚁防治' (Termite Control), '产权交易' (Property Transaction), '组合业务登记' (Composite Business Registration), '商品房预售' (Commercial Housing Pre-sale), '房改管理' (Housing Reform Management), '危房鉴定' (Dangerous Building Assessment), '房产档案管理' (Real Estate Archiving), '物业管理' (Property Management), '他项权登记' (Other Rights Registration), and '直管公房' (Directly Managed Public Housing).
 - 面向企业 (For Enterprises):** Lists services like '商品房网上合同签订' (Online Contract Signing for Commercial Housing), '二手房网上合同签订' (Online Contract Signing for Second-hand Housing), '商品房预售证办理' (Commercial Housing Pre-sale License Processing), '房地产经纪机构信息核查' (Real Estate Agency Information Check), '商品房网上办证' (Online Certificate for Commercial Housing), '经纪人考试报名' (Broker Exam Registration), '物业管理网上填报' (Online Reporting for Property Management), and '征收上岗培训报名' (Collection Staff Training Registration).
 - 特殊群体 (Special Groups):** Lists services for '廉租房申请' (Affordable Housing Application), '特困家庭白蚁防治' (Termite Control for Special Difficult Families), '经济适用房' (Economic Affordable Housing), and '绿色通道' (Green Channel).
- 各房产办证大厅指南 (Real Estate Certificate Office Guide):** Provides details for the '平海大厦房产办证大厅' (Pinghai Building Real Estate Certificate Office), including its address, working hours, and bus routes.

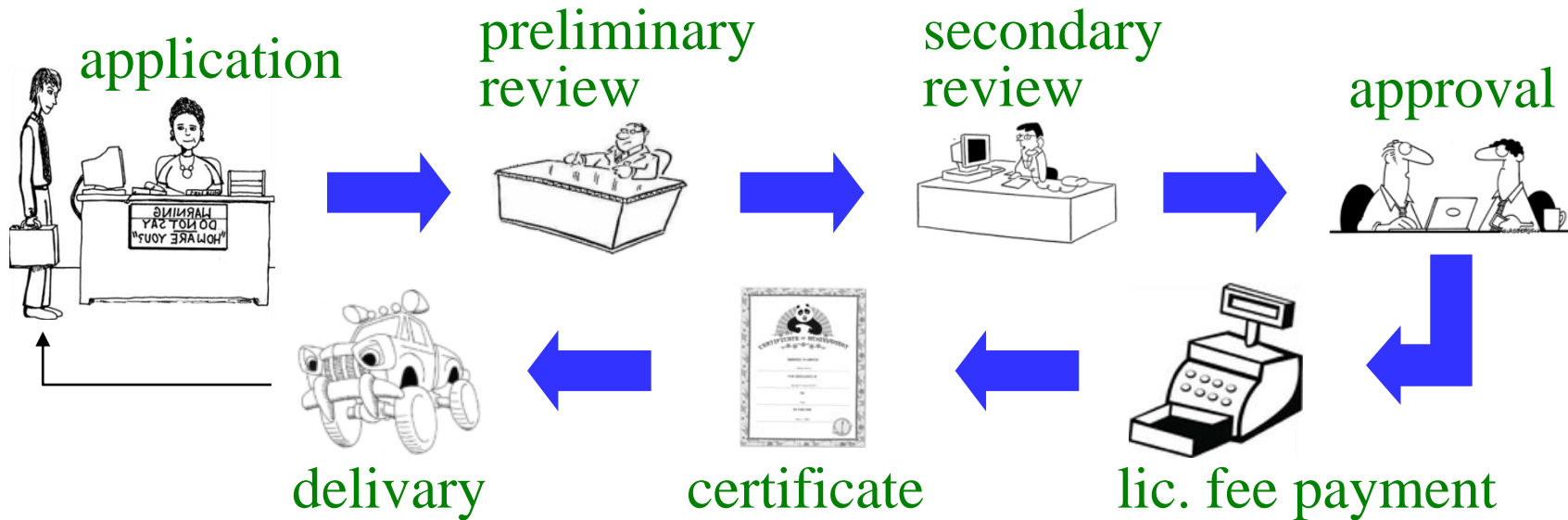
500 workflow models
300,000 cases/year

200,000+ for
 中国北车股份有限公司
 China CNR Corporation Limited
 [Jin et al CoopIS 2011]



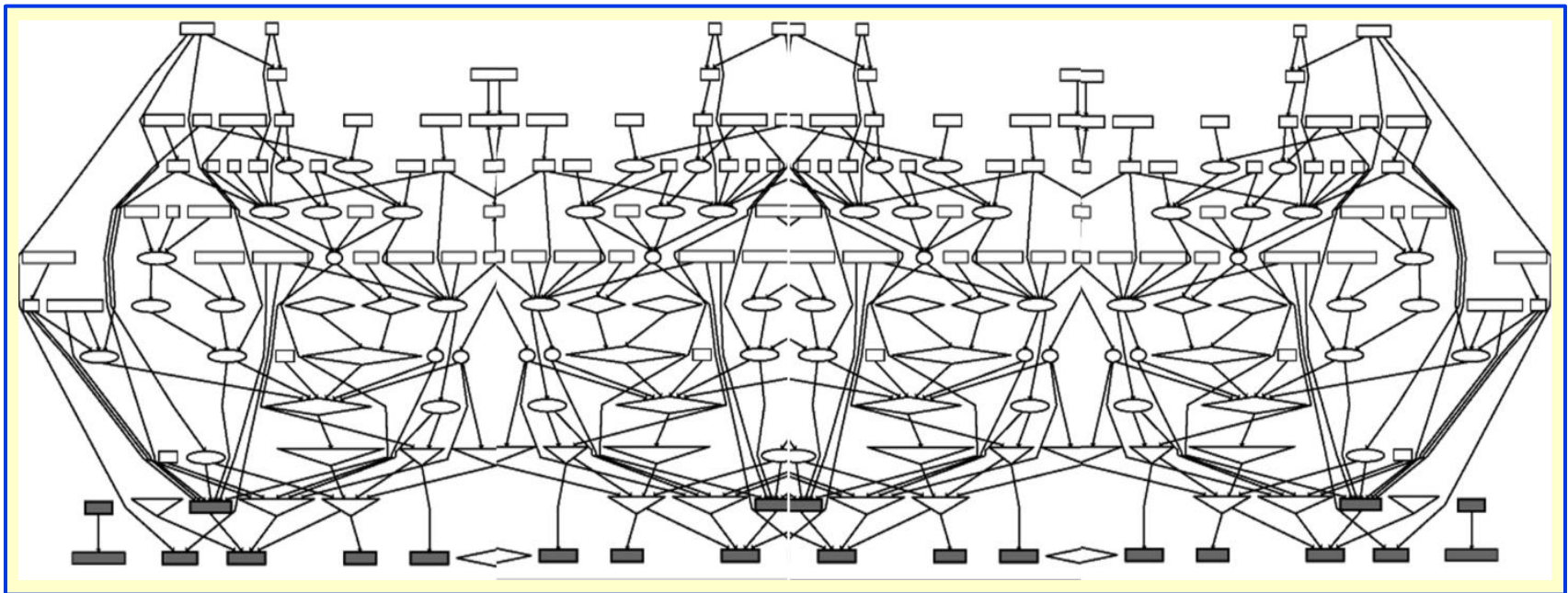
Permit for Selling an Unbuilt Apartment

■ *Obtaining a Permit*



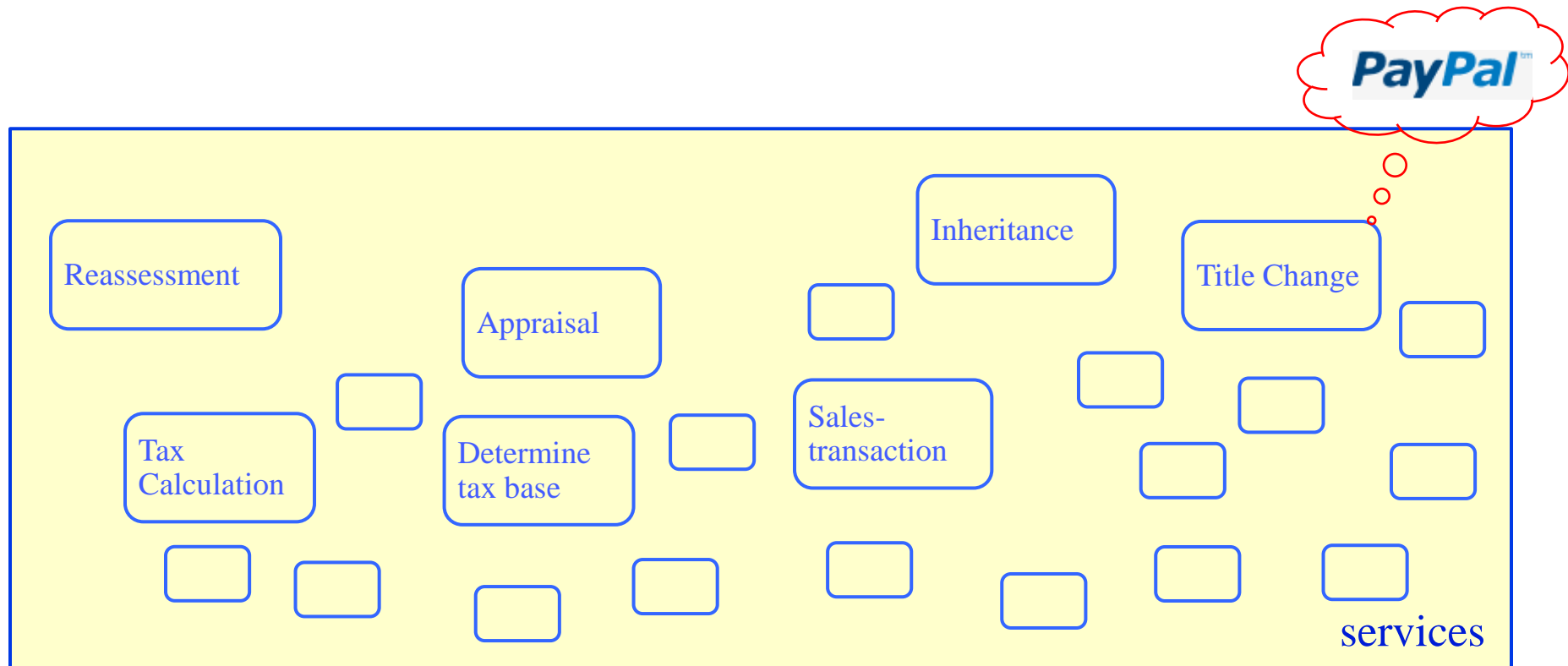
A Housing Management System

- Ad hoc design, developed over time, patches, multiple technologies, ... a typical legacy system
- Problems:
 - ❖ Embedded business logic, hard to learn
 - ❖ hard to maintain, costly to add new functionality
 - ❖ hard to change/evolve



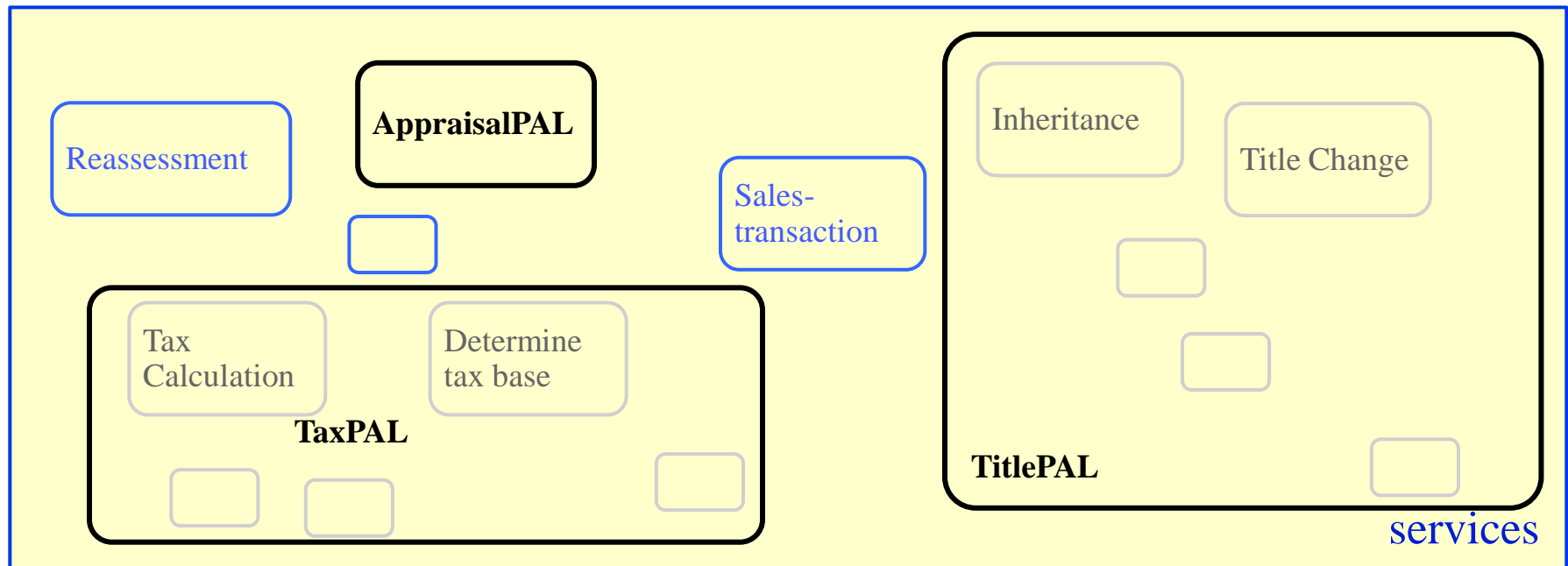
SOA Paints a Bright Picture, Scene 1 ...

- Services encapsulate system details and reflect business logic, easier to learn
- Easier to manage even if not technically
- New functions on top of services

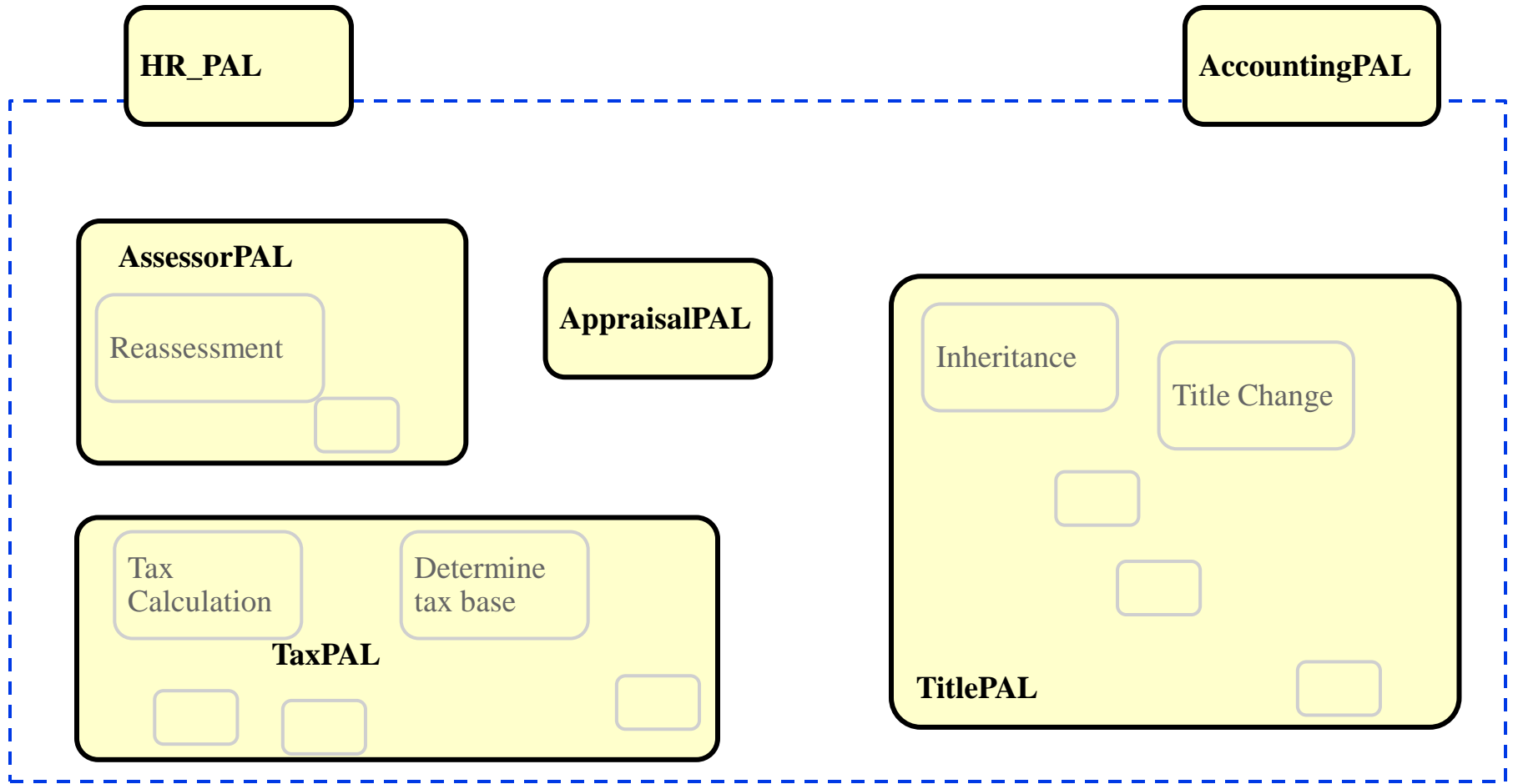


Scene 2: A World of "PAL"s, ...

- Organize into collections of services that may be offered to other cities



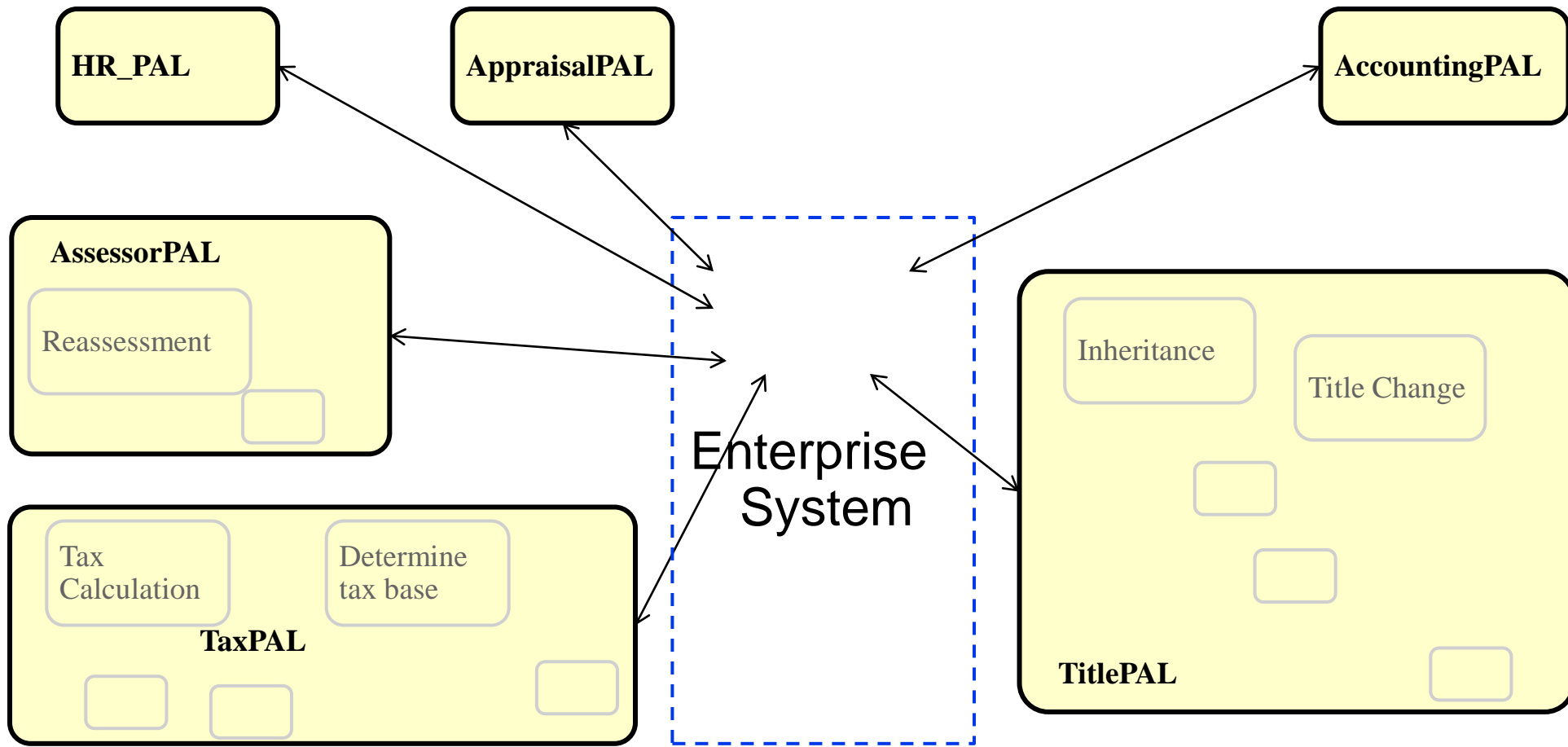
...Their **PALs**, or



Towards a goal of

- **Business Process as a Service (BPaaS)**

Scene 3: Virtual Enterprise Systems



Towards a goal of

- Business Process as a Service (BPaaS)
- Enterprise may run virtual IT systems

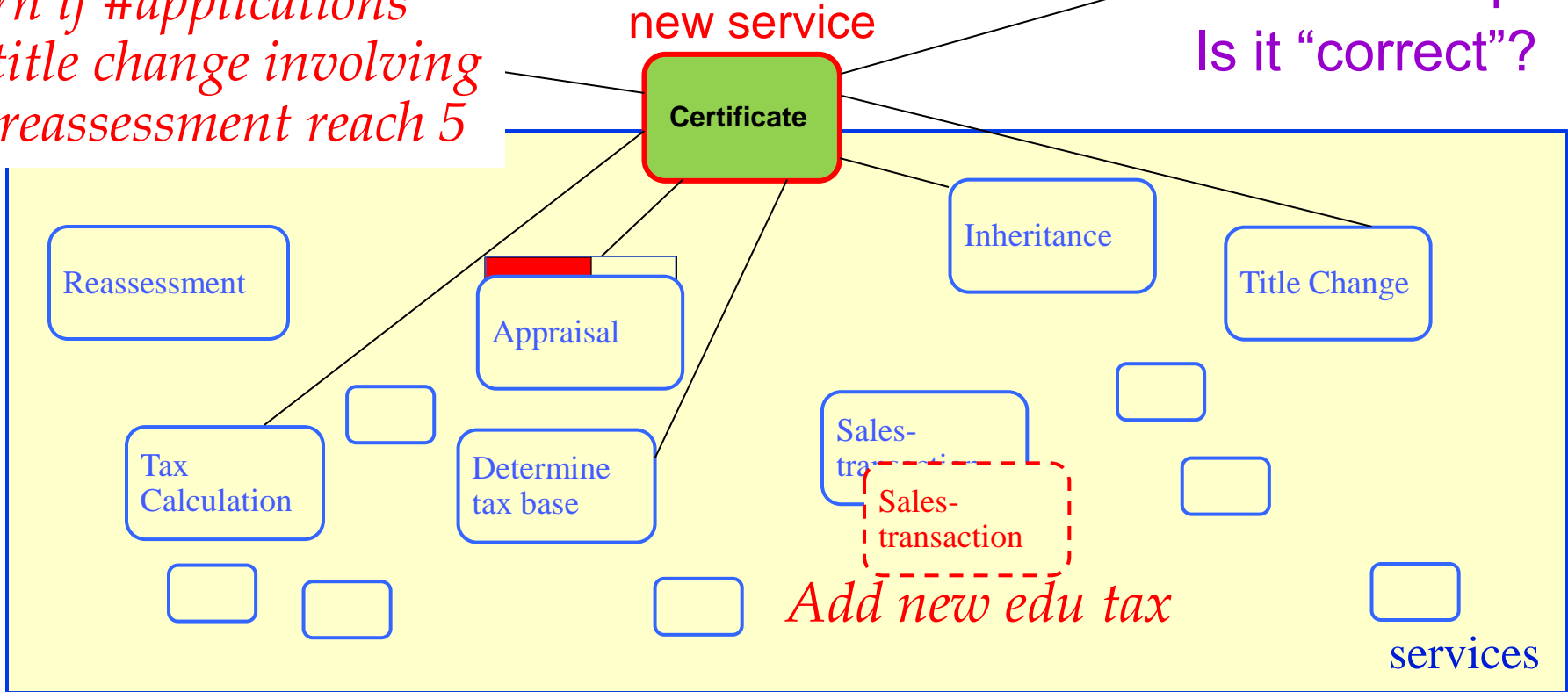
What are technical issues?

Technical Issues

How to query?

Warn if #applications for title change involving tax reassessment reach 5

How to compose?
Is it "correct"?



How to do transactions?

How to change & evolve?

Outline

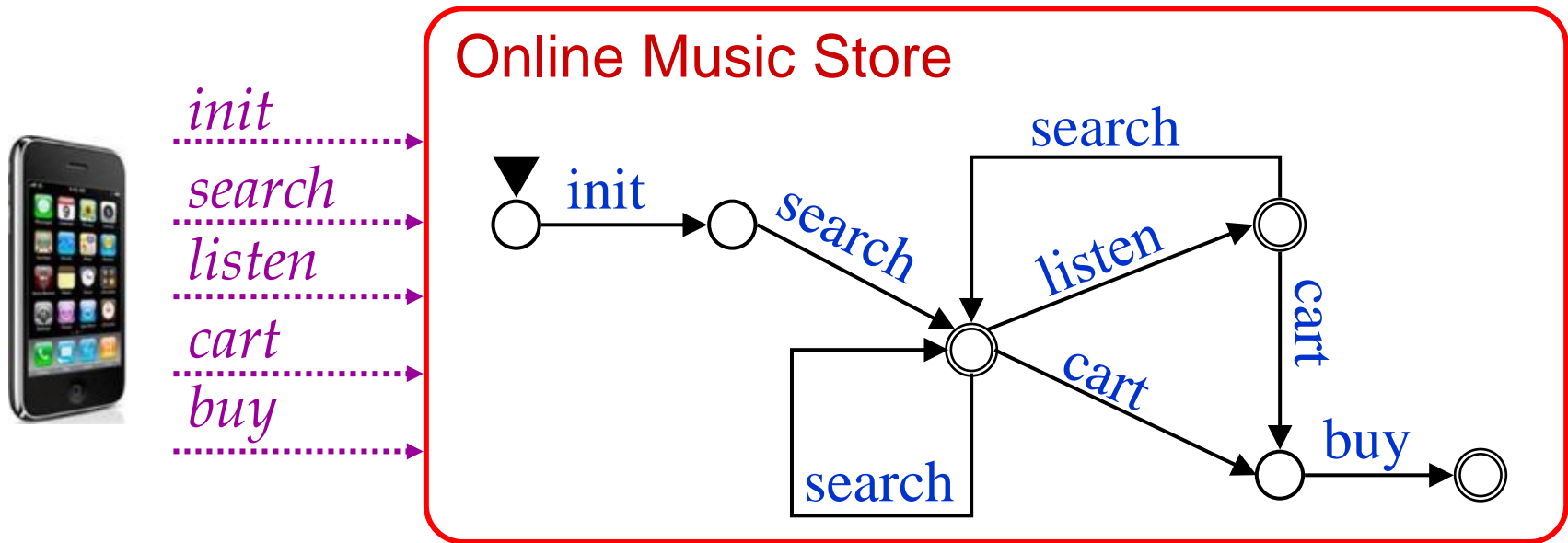
- Application Needs
- “Legacy” Services
- “Programmable” Services
- Data Encapsulating Services
- Research Challenges
- Conclusions

"Legacy" Services

- Services have **states**, but only **finitely many**
- Can be modeled with:
finite state machines, process algebras, workflow nets (Petri nets), activity diagrams, state charts, ...
- Have been used in studying problems related to service composition
 - ❖ **Automated design**, e.g., Roman services
 - ❖ **Verification**
 - ❖ **Optimization** (e.g., QoS based)
- **Wealth of knowledge, rich literature**
[WWW, ICSOC, ICWS, SCC, SOCA, ...]

Roman Services

- A service consists of activities with a finite state control

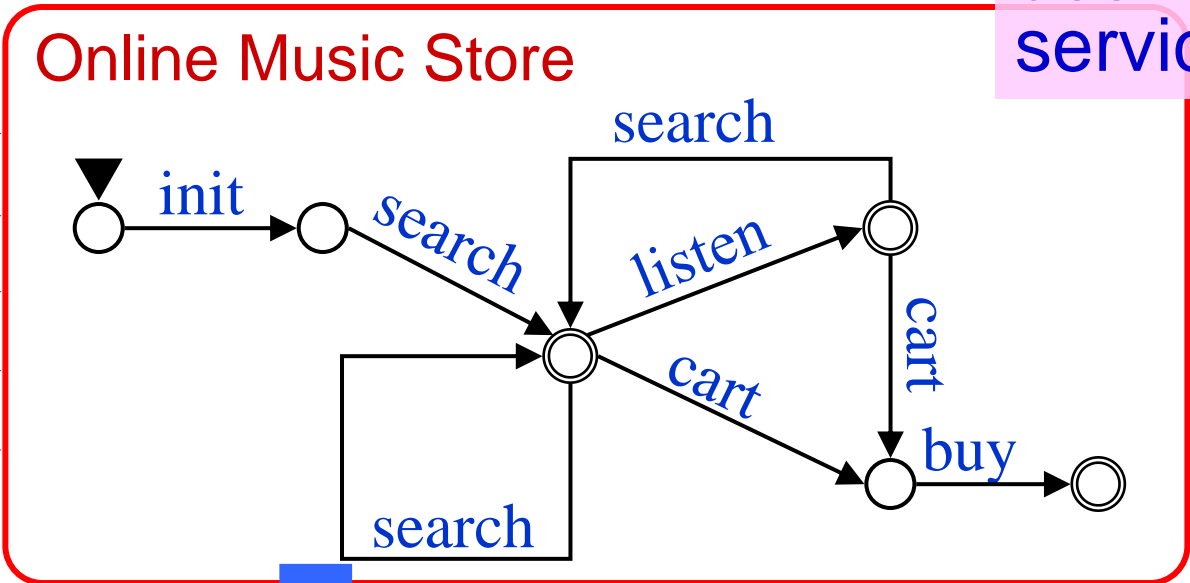


- Transitions labeled by activities
- For a given state, the out-edges represent the set of options that will be presented to the user

The Composition Problem



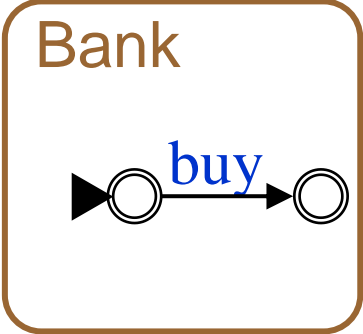
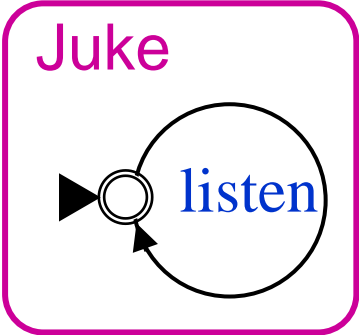
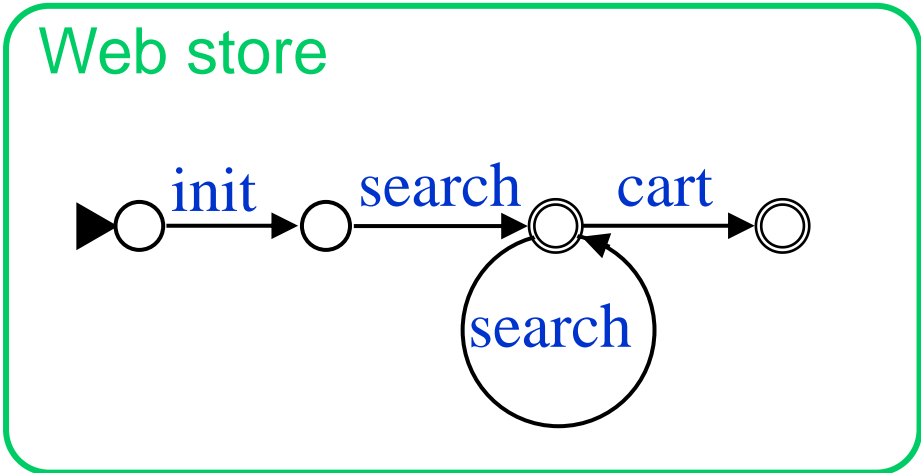
init
search
listen
cart
buy



desired service



available services

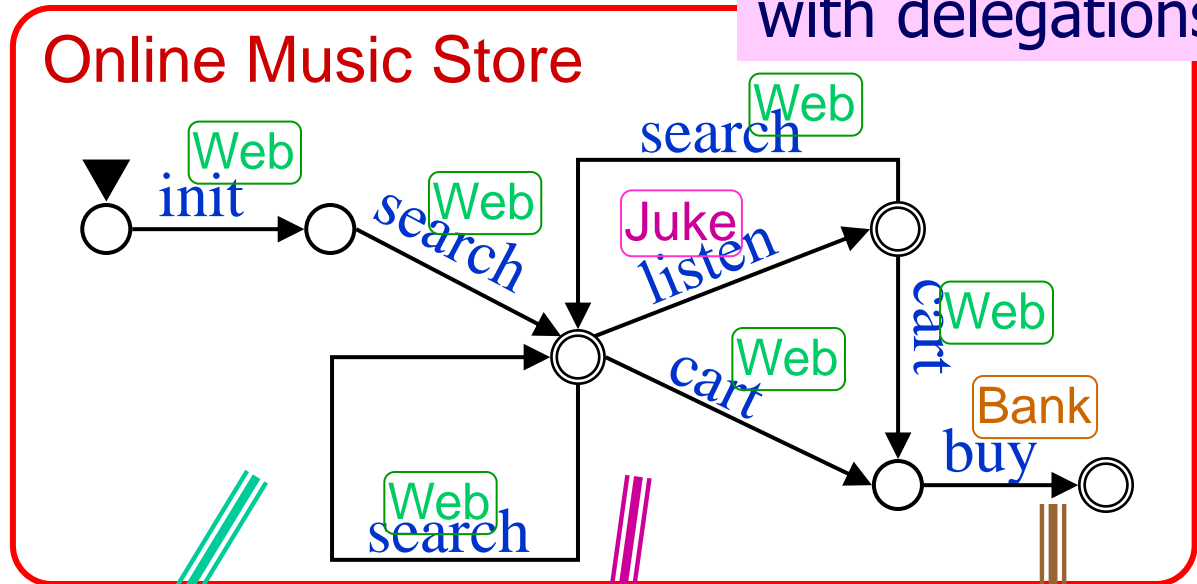


Composition As a Delegator

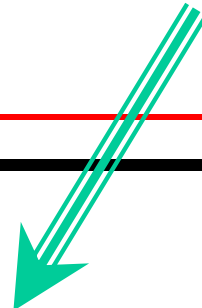
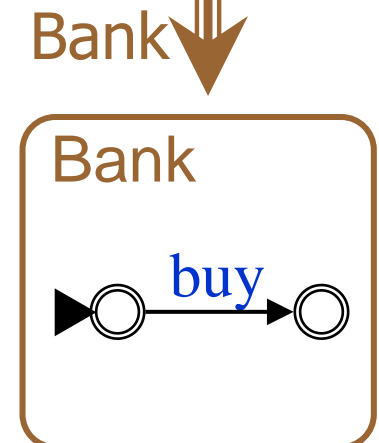
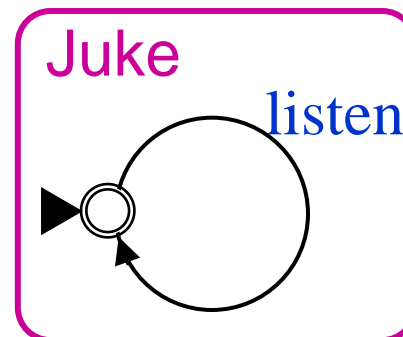
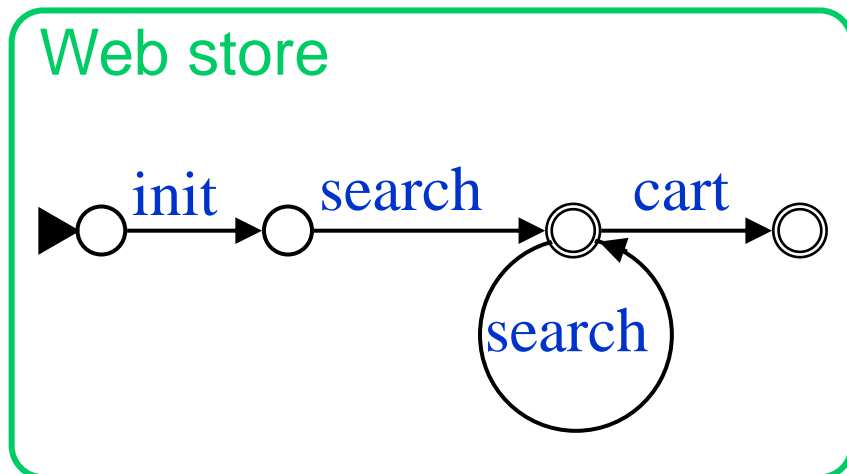
Delegator:
a service annotated with delegations



init
search
listen
cart
buy



Available services



Roman Service Composition

- **Problem:**

Given a target Roman service and a set of Roman services, find a delegator if exists

[Berardi-Calvanese-DeGiacomo-Lenzerini-Mecella ICSOC 03]

- **Deterministic** [Berardi-Calvanese-DeGiacomo-Lenzerini-Mecella ICSOC 03]

- **Nondeterministic & Lookahead (batched)**

[Gerede-Hull-Ibarra-S. ICSOC 04]

- **May delegate more than once** [Berardi et al ICSOC 04]

- **With messages** [Berardi-Calvanese-De Giacomo-Hull-Mecella VLDB 05]

- **Online delegation** [Gerede-Ibarra-Ravikumar-S. TCS 08]

- **Use only a subset of services** [Hassen-Nourine-Toumani ICSOC 08]

Legacy Services in Practice: Limited Applicability

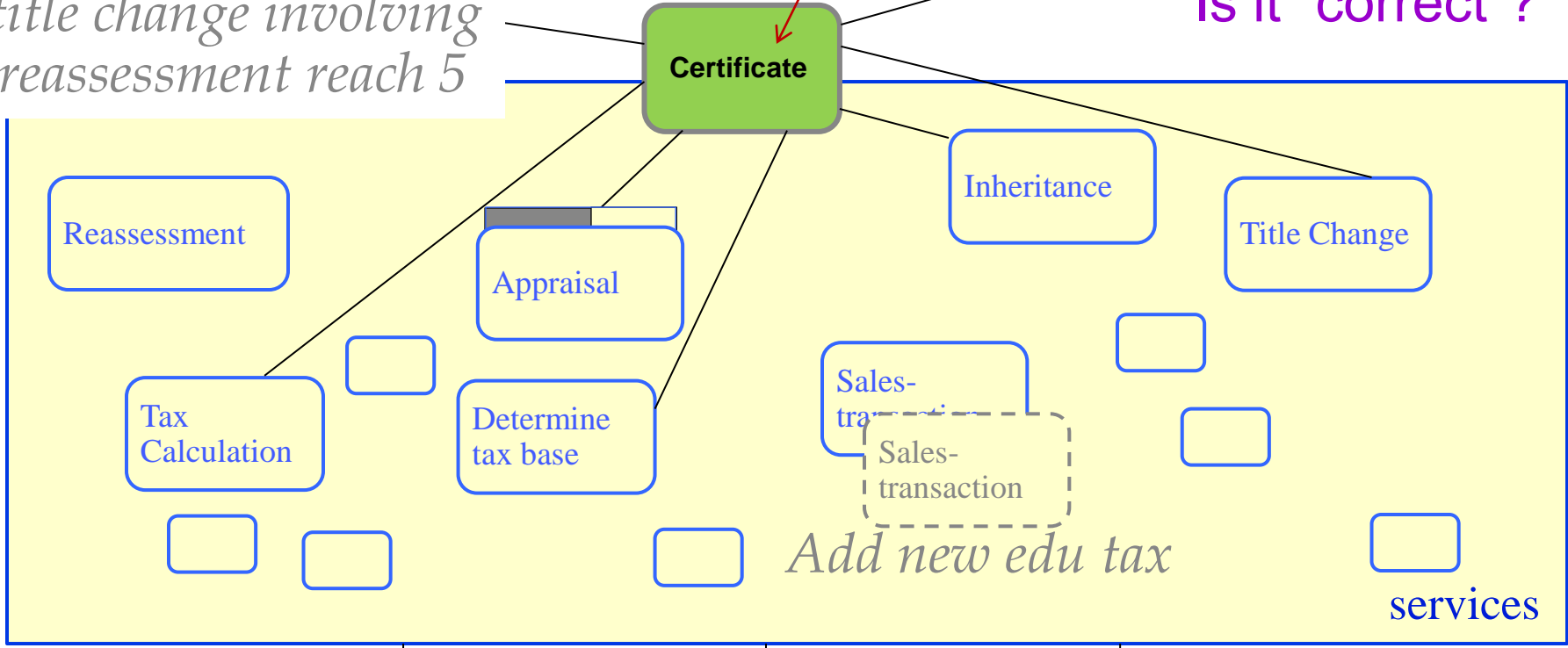
How to query?

Warn if #applications for title change involving tax reassessment reach 5

age > 55 & ...

new service

How to compose?
Is it "correct"?



How to do transactions?

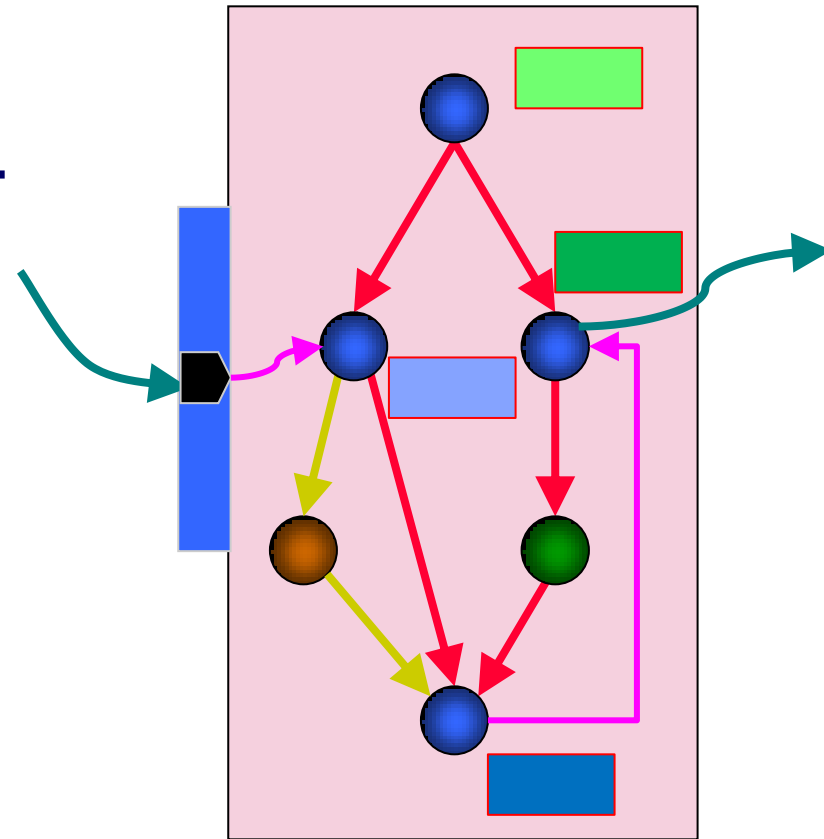
How to change & evolve?

Outline

- Application Needs
- “Legacy” Services
- **“Programmable” Services**
- Data Encapsulating Services
- Research Challenges
- Conclusions

Service Programming: Data as Variables

- Roughly: finite states plus variables
 - ❖ Pragmatic approach
- Examples: BPEL, jBPM, YAWL, ...
- Design:
Programming with services
- Analysis/checking “correctness”:
well, undecidable

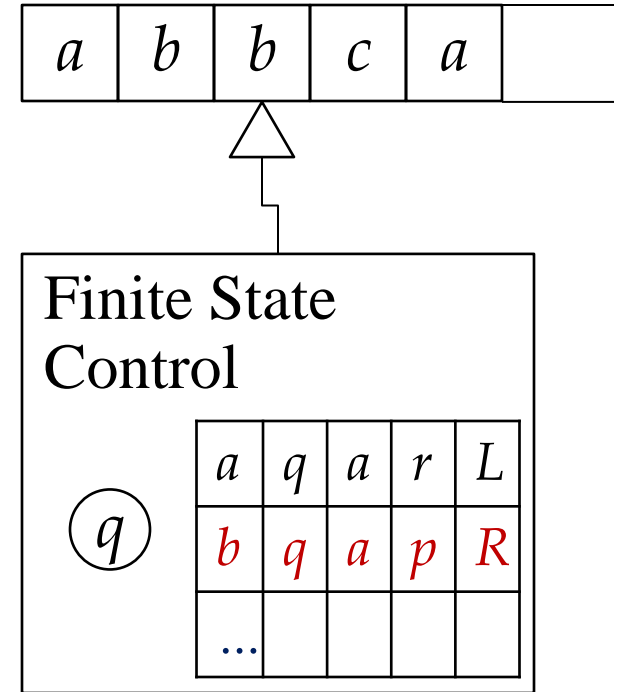
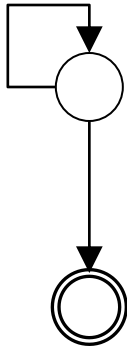


A BPEL Process

BPEL is Turing Complete

Before: ba
 Current: b
 After: $ca \dots$
 State: q

Before: aba
 Current: c
 After: $a \dots$
 State: p

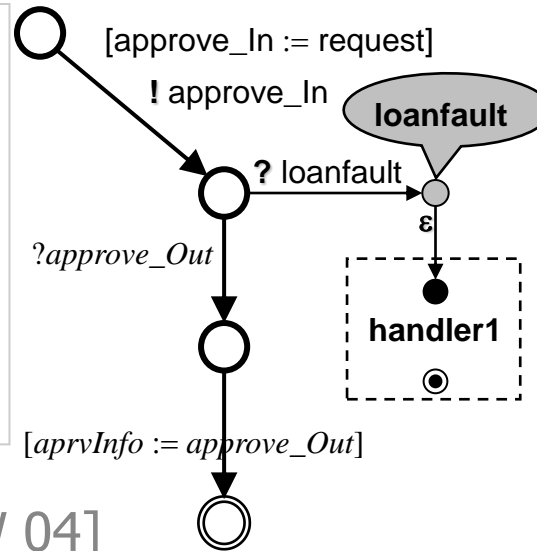


- BPEL can express all Turing computations
- Checking properties for standalone BPEL is not solvable
- Take a step back: finite state variables

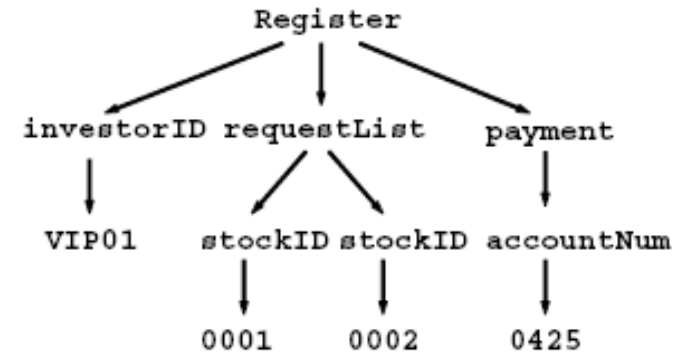
"Finite State" Variables

- BPEL control structure → a finite state machine
- XML Schema typed variables:
 - ❖ primitive types limited to a finite set of values
 - ❖ Structured: finitely many "repeats"

```
<invoke operation="approve",  
  invar="request",  
  outvar="aprInfo" >  
  <catch faultname="loanfault">  
    <... handler1 ... />  
  </catch>  
</invoke>
```



[Fu-Bultan-S. WWW 04]



[Fu-Bultan-S. ISSTA 04]

- Can be improved with the "hyperplane" technique

[Gerede-S. ICSOC 07]

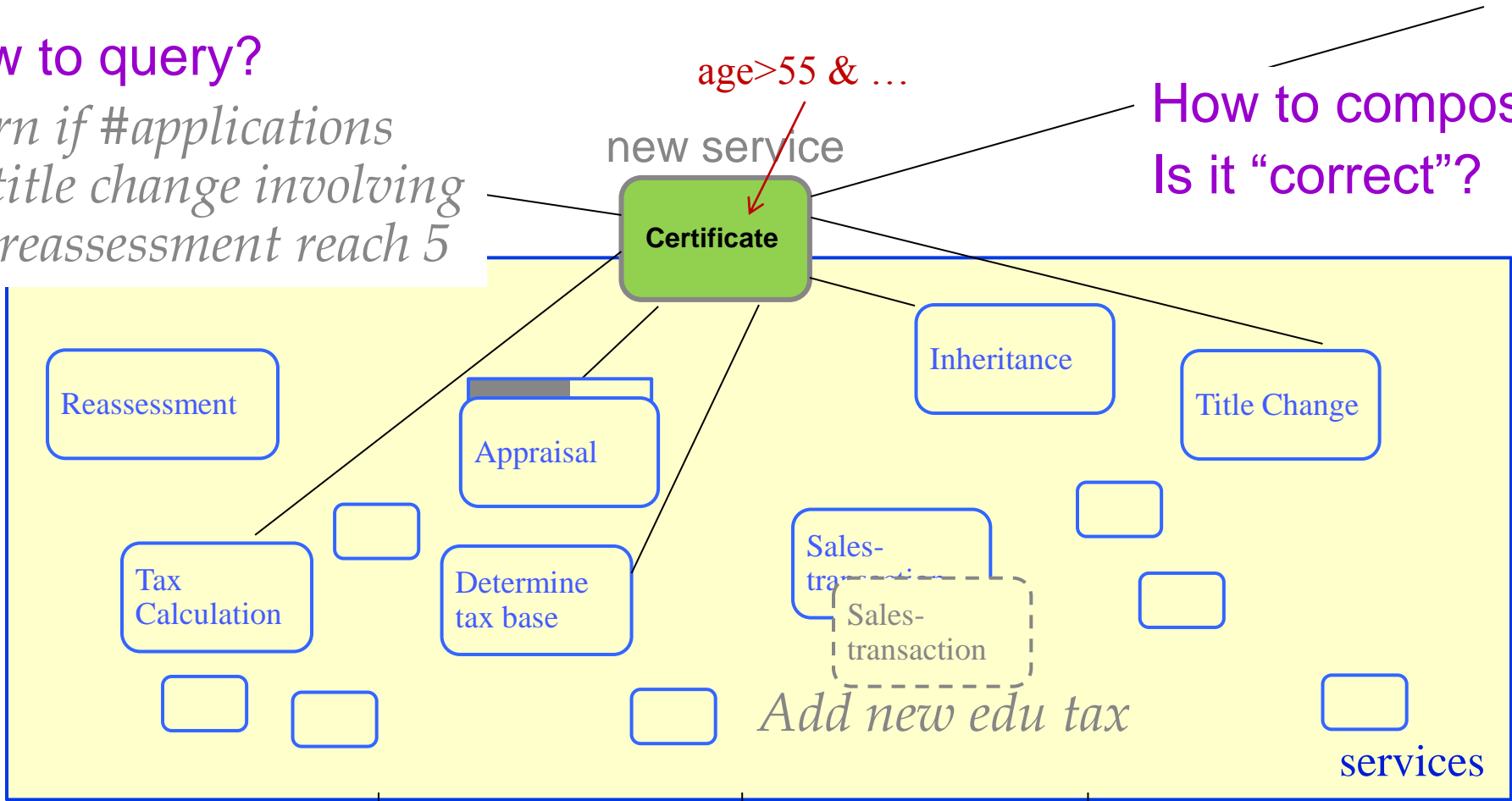
Helps, But Services Are Connected via Data

How to query?

Warn if #applications for title change involving tax reassessment reach 5

age > 55 & ...

How to compose?
Is it "correct"?

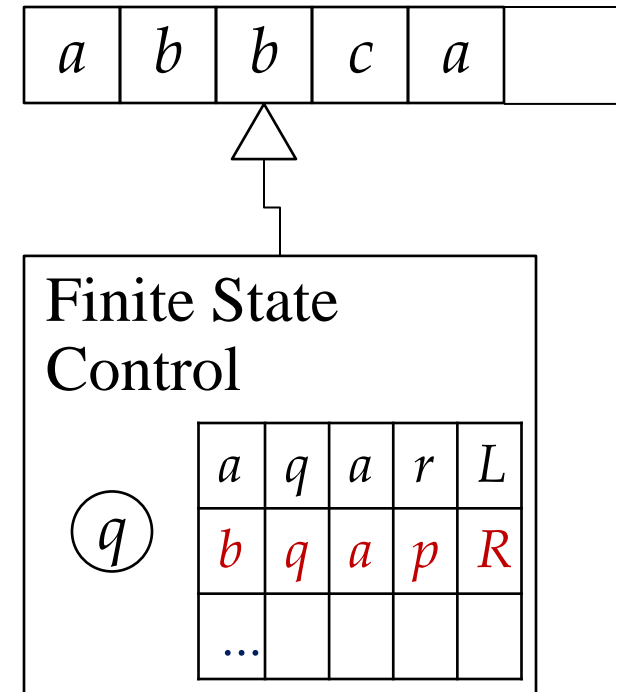
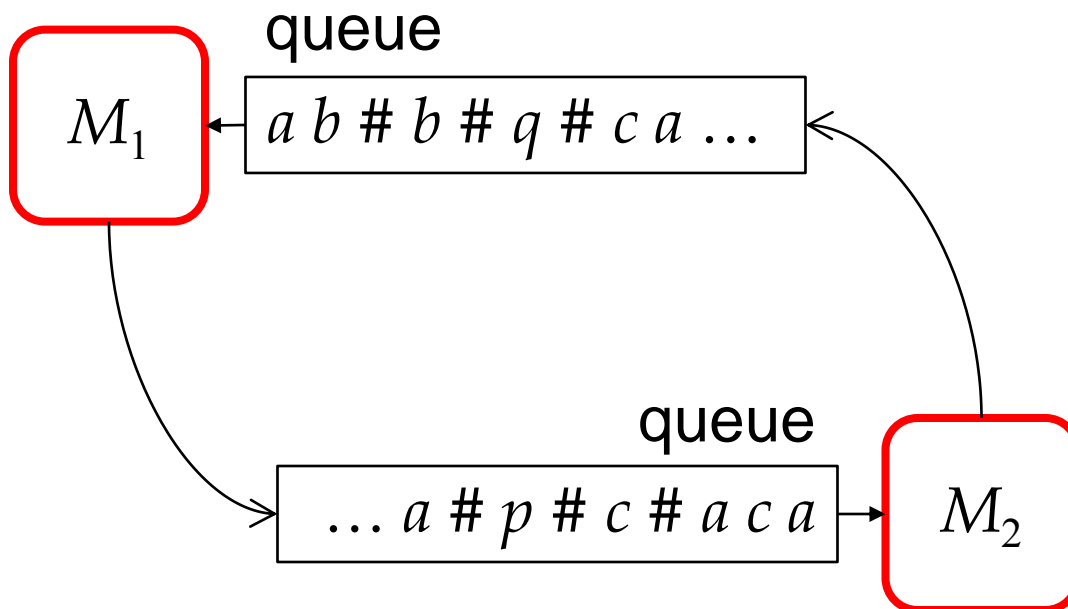


How to do transactions?

How to change & evolve?

Collaborating BPEL Services

- Two finite state machines with input queues can simulate Turing machines [Brand-Zafiropulo JACM 83]
- Model checking BPEL services with queues and finite state variables is not solvable



Service Programming is an Art

How to query?

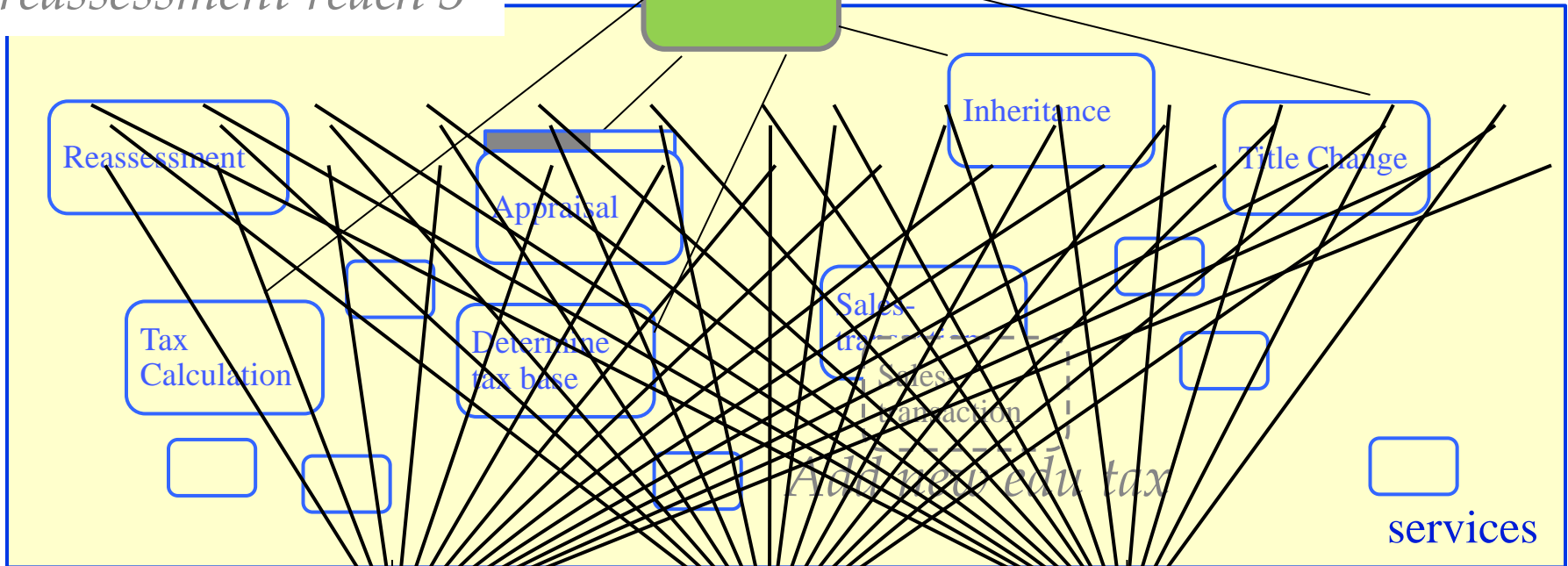
Warn if #applications for title change involving tax reassessment reach 5

age > 55 & ...

new service

How to compose?
Is it "correct"?

Certificate



How to do transactions?

How to change & evolve?

Current Practice

- Data and services are **separately** modeled, designed, managed
- The separation adds difficulties in design, execution, maintenance, changes
- In addition, many issues can't be addressed
 - ❖ Workflow transaction remains an art
 - ❖ Data consistency is a concern of DBMS even though violations are caused by service execution
 - ❖ Biz analytics is an after thought
 - ❖ Long tail phenomenon is a “holy grail”

Big Data—A Gowing Torrent

\$600 to buy a disk drive that can store all of the world's music

5 billion mobile phones in use in 2010

30 billion pieces of content shared on Facebook every month

40% projected growth in global data generated per year vs. **5%** growth in global IT spending

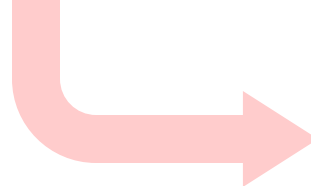
235 terabytes data collected by the US Library of Congress by April 2011

15 out of 17 sectors in the United States have more data stored per company than the US Library of Congress

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

- Availability of “big data” brings opportunities for improving productivity

15 out of 17 sectors in the United States have more data stored per company than the US Library of Congress



Big Data + Biz Processes → Big Potential



US health care

- \$300 billion value per year
- ~0.7 percent annual productivity growth



Manufacturing

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



US retail

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



Europe public sector administration

- €250 billion value per year
- ~0.5 percent annual productivity growth

Two observations

- A significant portion of big data generated by biz processes
- Productivity growth only obtainable via more efficient/effective biz processes



Global personal location data

- \$100 billion+ revenue for service providers
- Up to \$700 billion value to end users

Source:
MGI Analysis

Service Programming is an Art

How to query?

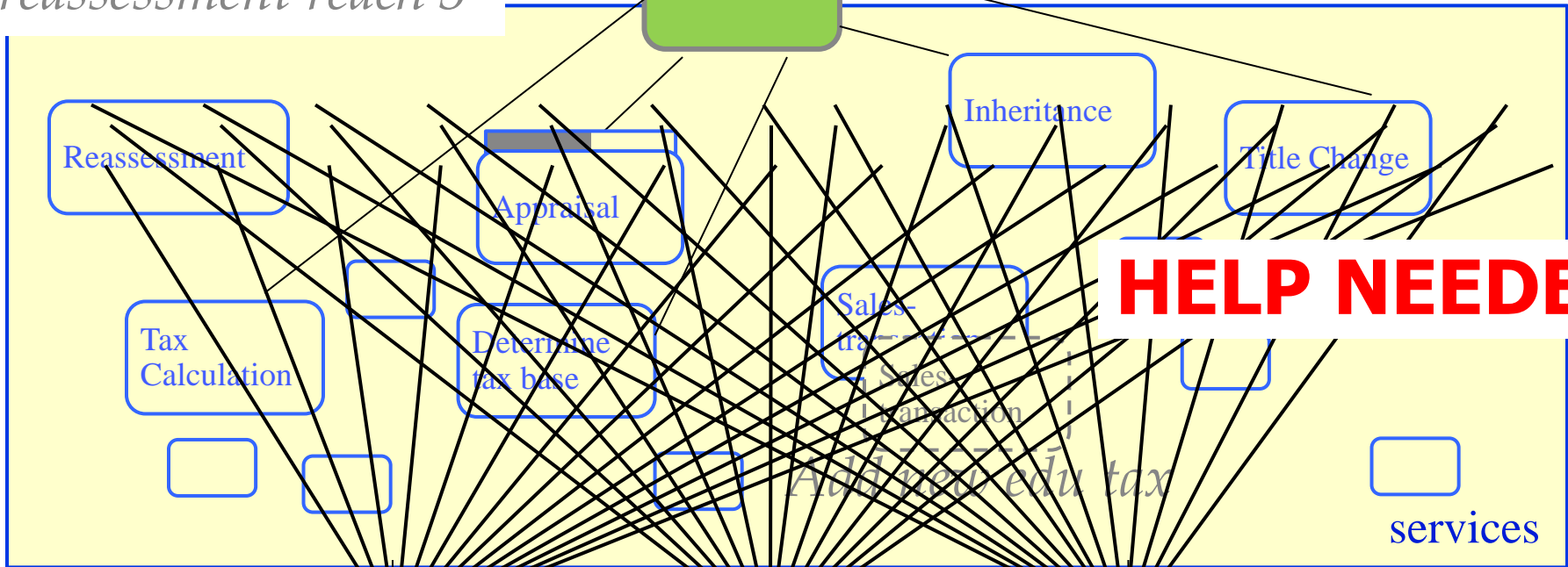
Warn if #applications for title change involving tax reassessment reach 5

age > 55 & ...

new service

Certificate

How to compose?
Is it "correct"?

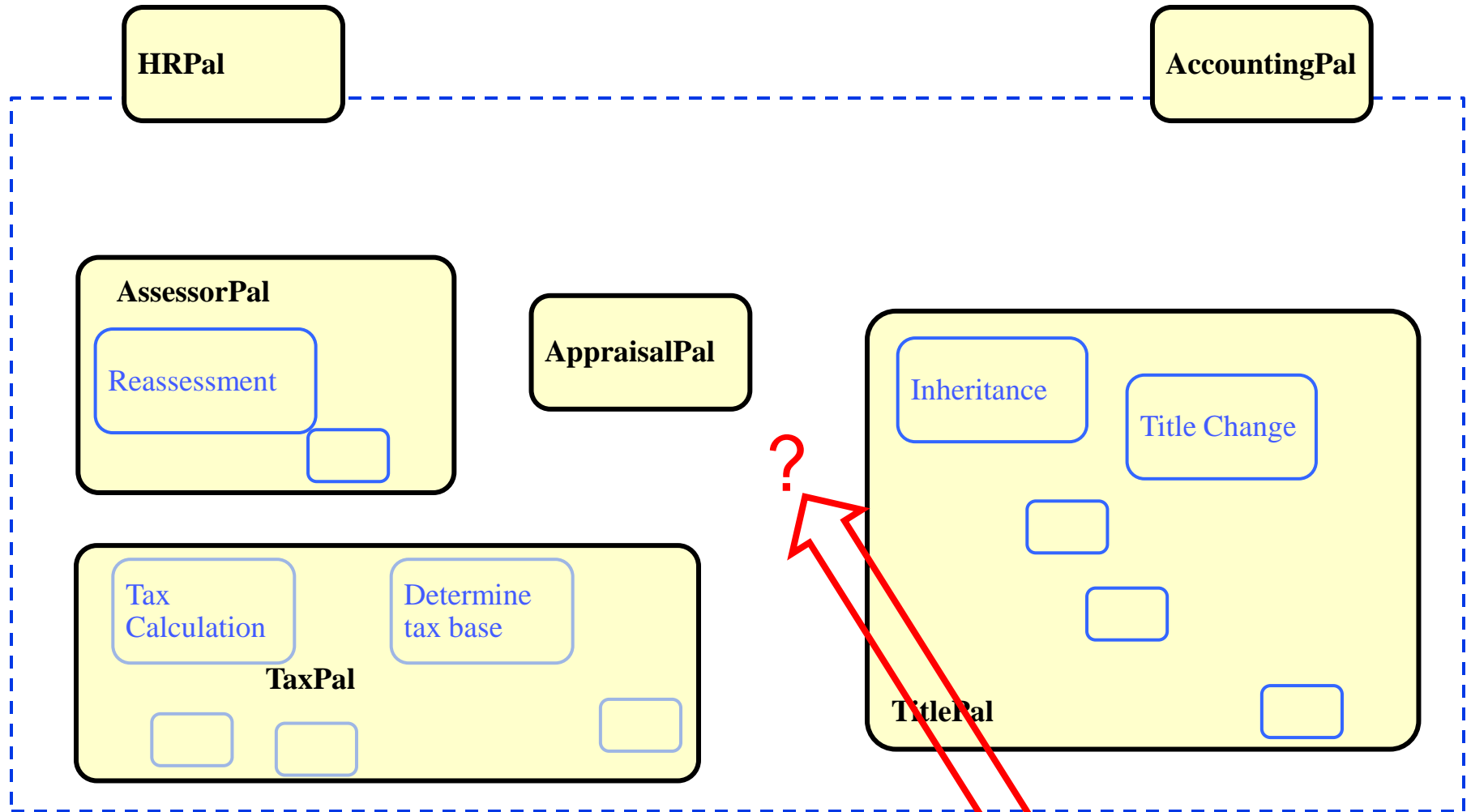


How to do transactions?

How to change & evolve?

The real world is not too kind

Needed: Enterprise Data Management

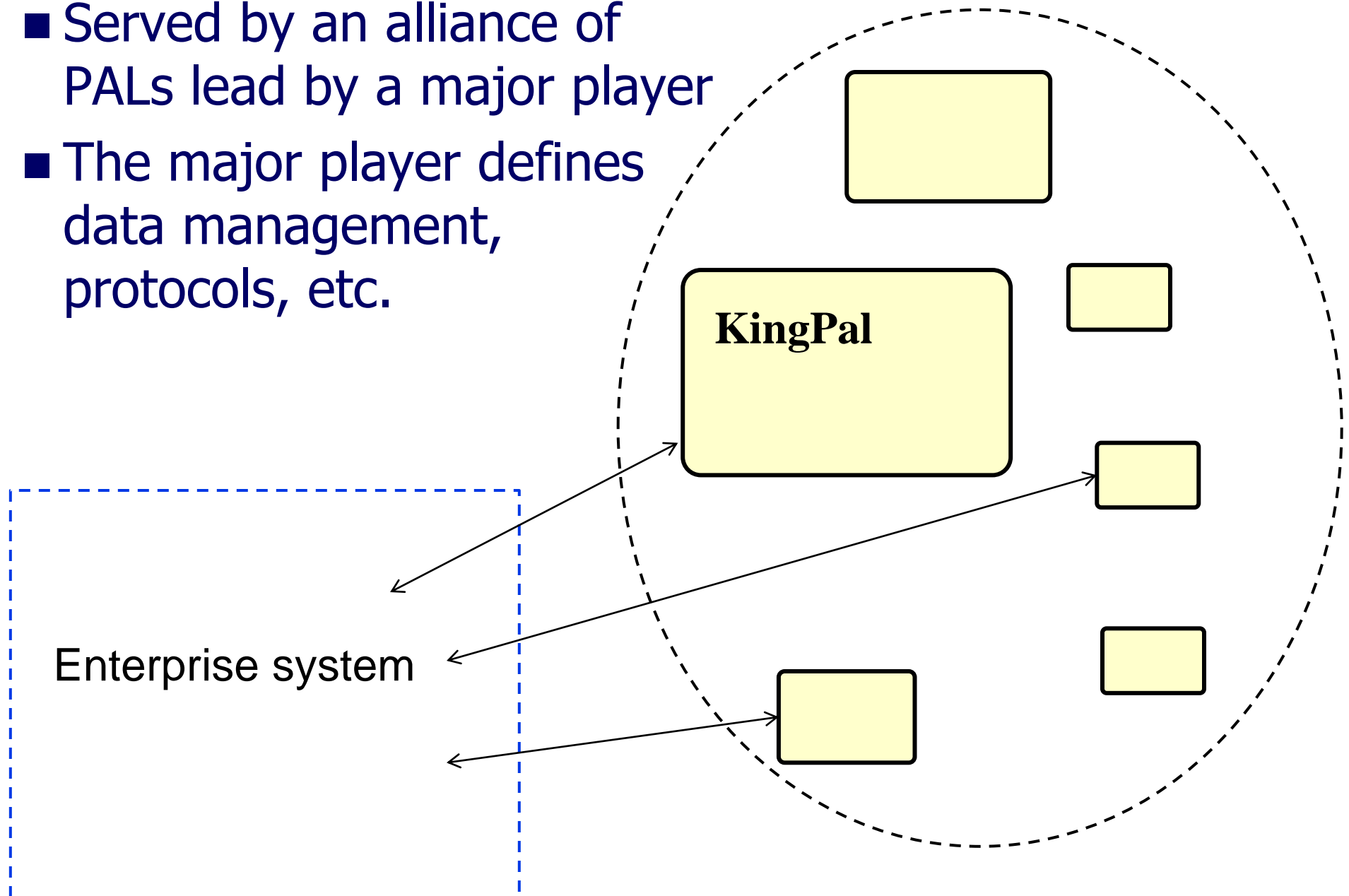


■ Business Process as a Service (BPaaS)



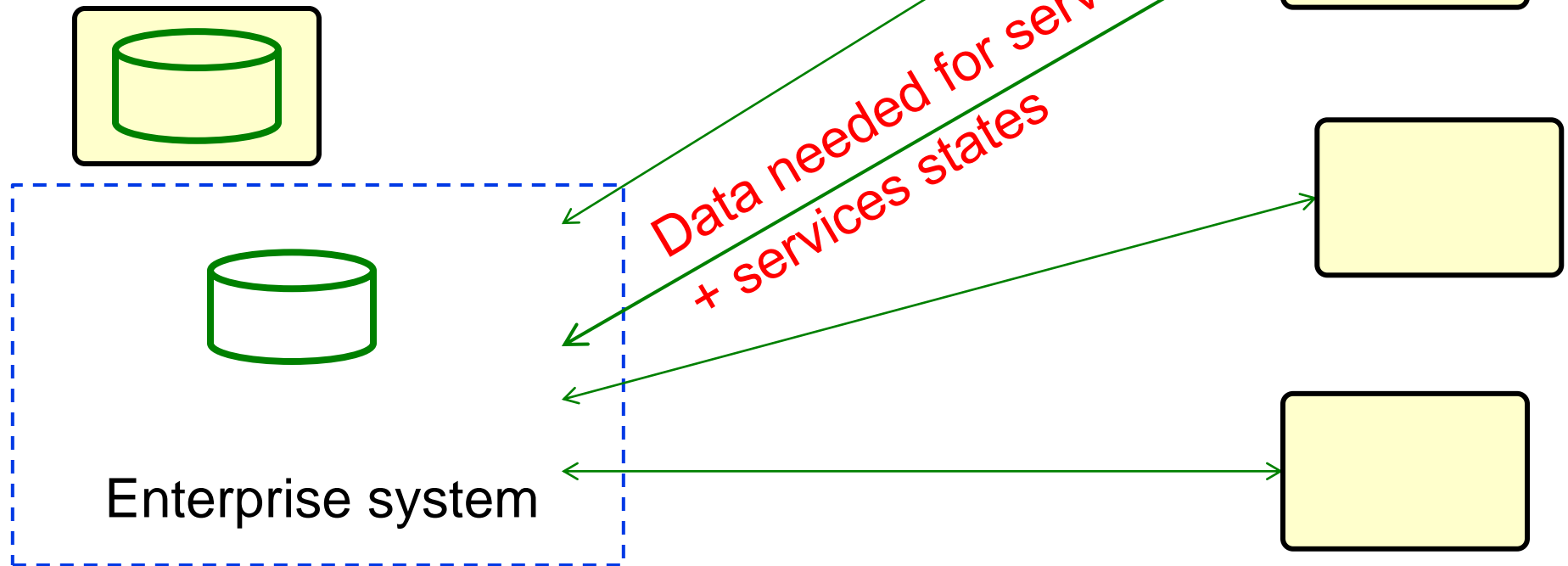
Possibility 1: Monopoly Model

- Served by an alliance of PALs lead by a major player
- The major player defines data management, protocols, etc.



Possibility 2: Open Market Model

- No major players
- Enterprise makes its own data management plan
- **Bring Your Own Data (BYOD)**
- Could also use a “dataPal”



An Application Challenge

- What are appropriate models for both:
 - ❖ Enterprise data and management
 - ❖ Enterprise services inter-related through persistent data

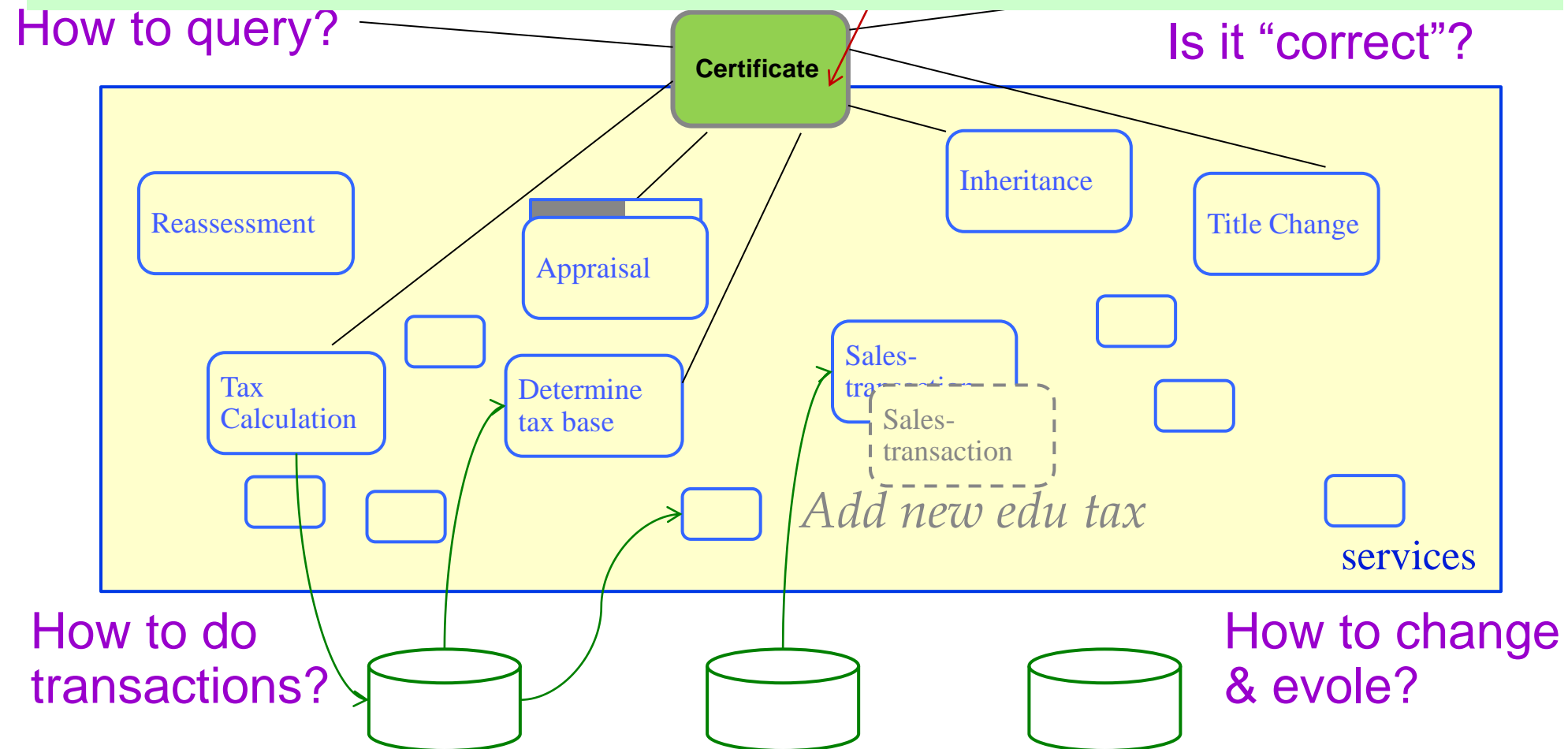
- Must support
 - ❖ Composition design and analysis
 - ❖ Runtime service execution management
 - ❖ Transactions
 - ❖ Process evolution

Outline

- Application Needs
- “Legacy” Services
- “Programmable” Services
- Data Encapsulating Services
- Research Challenges
- Conclusions

Services Aware of Data Management

- First attempt: services + persistent data + mappings
- Verbatim copy of the reality, not much help
- Services and data are not intimately related



Four Kinds of Data

- **Business data** essential for business logic
 - Examples: *items, shipping addresses*
- **Enactment status**: the current execution snapshot
 - Examples: *order sent, shipping request made*
- **Resource usage and state** needed for service execution
 - Examples: *cargo space reserved, truck schedule to be determined*
- **Correlation** between processes instances
 - Example: *3 warehouse fulfillment process instances for Jane's order*
- Need models that include both control and data

Process Models & Data

Four classes of process models:

- **Data abstraction** models: data mostly absent
 - WF (Petri) nets, BPMN, UML Activity Diagrams, ...
- **Data-aware** models: data present (as variables), but storage and management hidden
 - BPEL, YAWL, ...
- **Storage-aware** models: schemas for persistent stores, mappings to/from data in BPs defined and managed manually
 - jBPM, ...
- **Data encapsulating** models: logical data modeling, automated modeling other 3 types, data-storage mapping
 - Business artifact-centric models

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”
 - Technically, it includes two parts:
 - ❖ **Information model:**
 - data needed to move through workflow
 - ❖ **Lifecycle:**
 - possible ways to evolve
- ✓ Very natural to business managers and BP modelers

Example: Restaurant Processes

repository

Activity

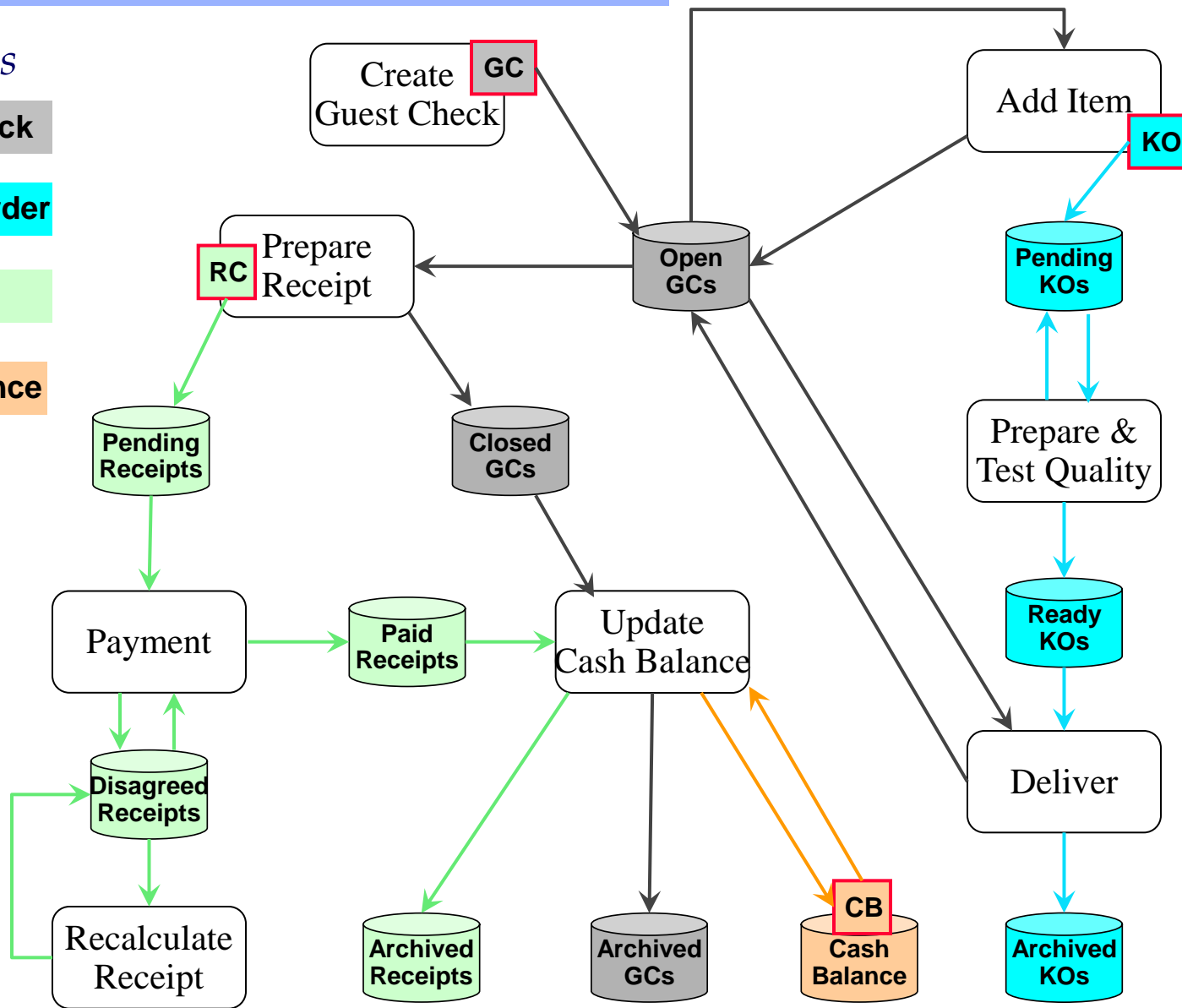
Artifacts

Guest Check

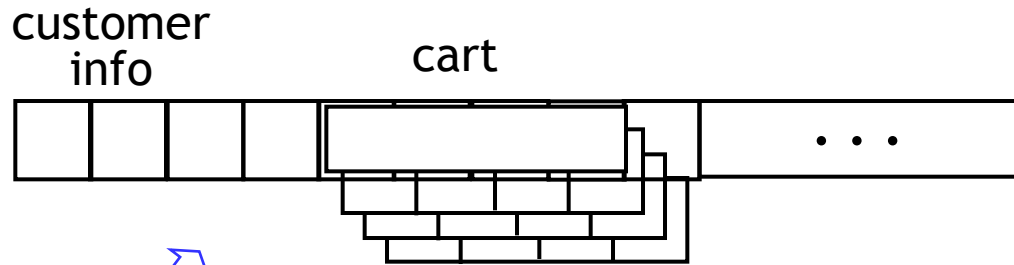
Kitchen Order

Receipt

Cash Balance



Artifact-Centric Biz Process Models



Artifacts (Info models)

+

Specification of artifact lifecycles

- Informal model [Nigam-Caswell IBM Sys J 03]

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

- Formal models

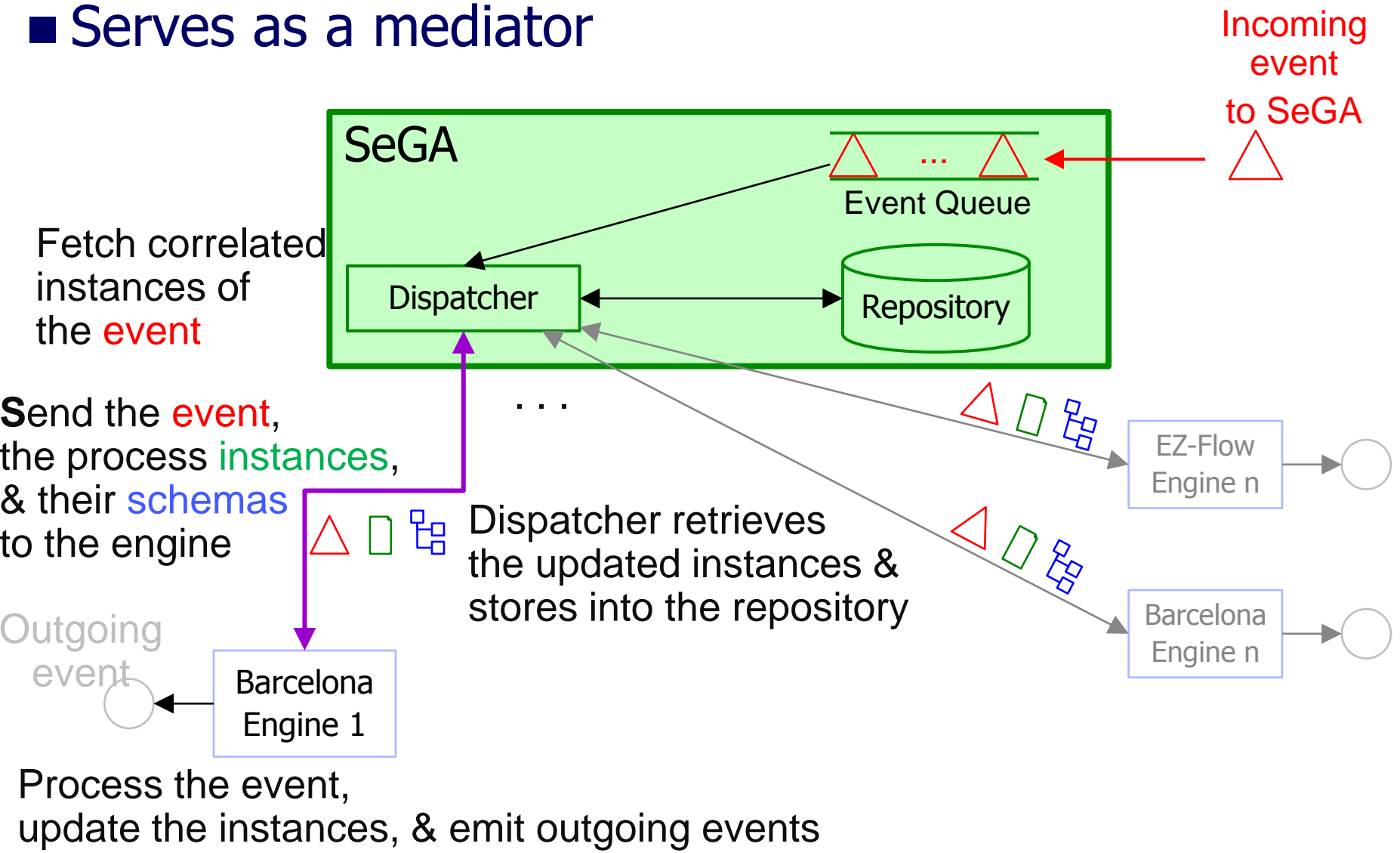
 - ❖ State machines [Gerede-Bhattacharya-S. SOCA 07][Gerede-S. ICSSOC 07]

 - ❖ Rules [Bhattacharya-Gerede-Hull-Liu-S. BPM 07][Hull et al WSFM 2010]

SeGA: A Service Wrapper/Mediator

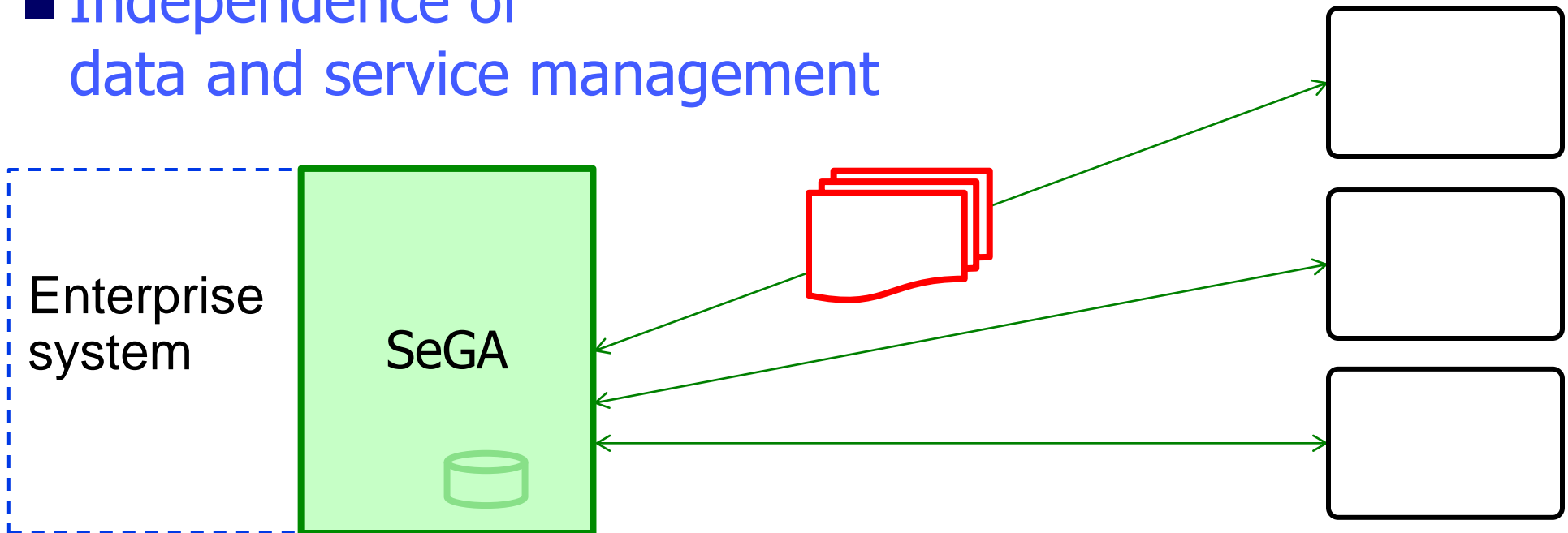
[Sun-Xu-S.-Yang CoopIS 12]

- SeGA separates data from execution engine
- Serves as a mediator



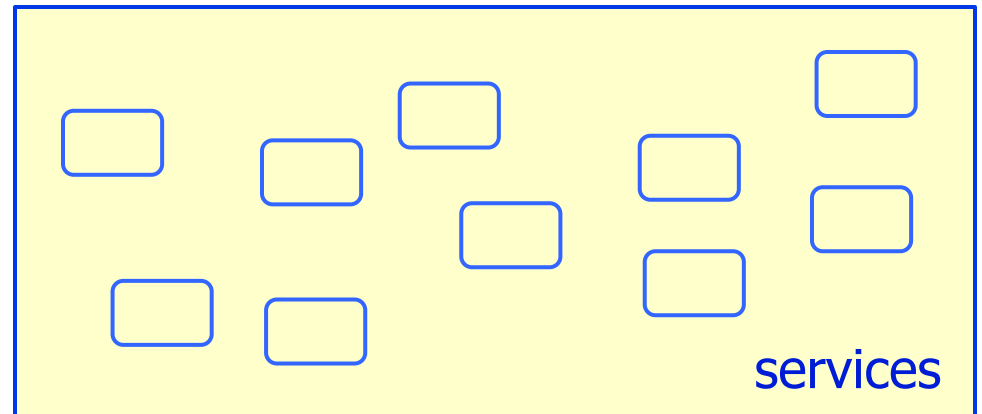
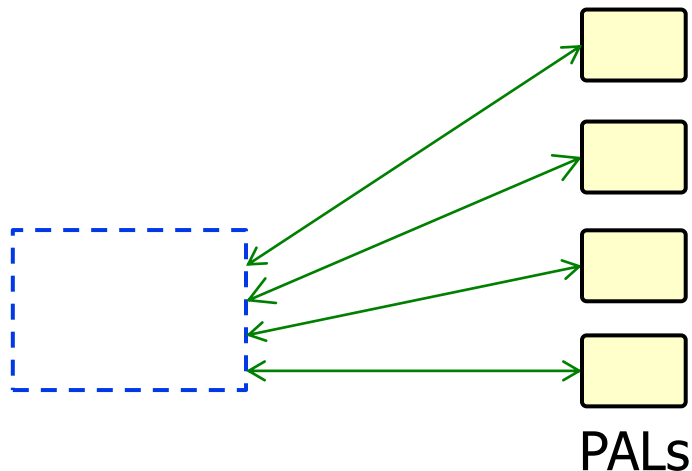
Data Encapsulating Services

- **Data package** between SeGA & services:
 - ❖ Business data, enactment data, resource data, correlation data
- Data encapsulating services: Stateful services but the engine need not maintain state
- Independence of data and service management



Independence of D-S Management

- Freedom to change service and execution without altering data management
- Freedom to change data management without altering services
- The independence hides the differences between the worlds of *services* and *PALs*
 - ❖ Great start for some facinating research!



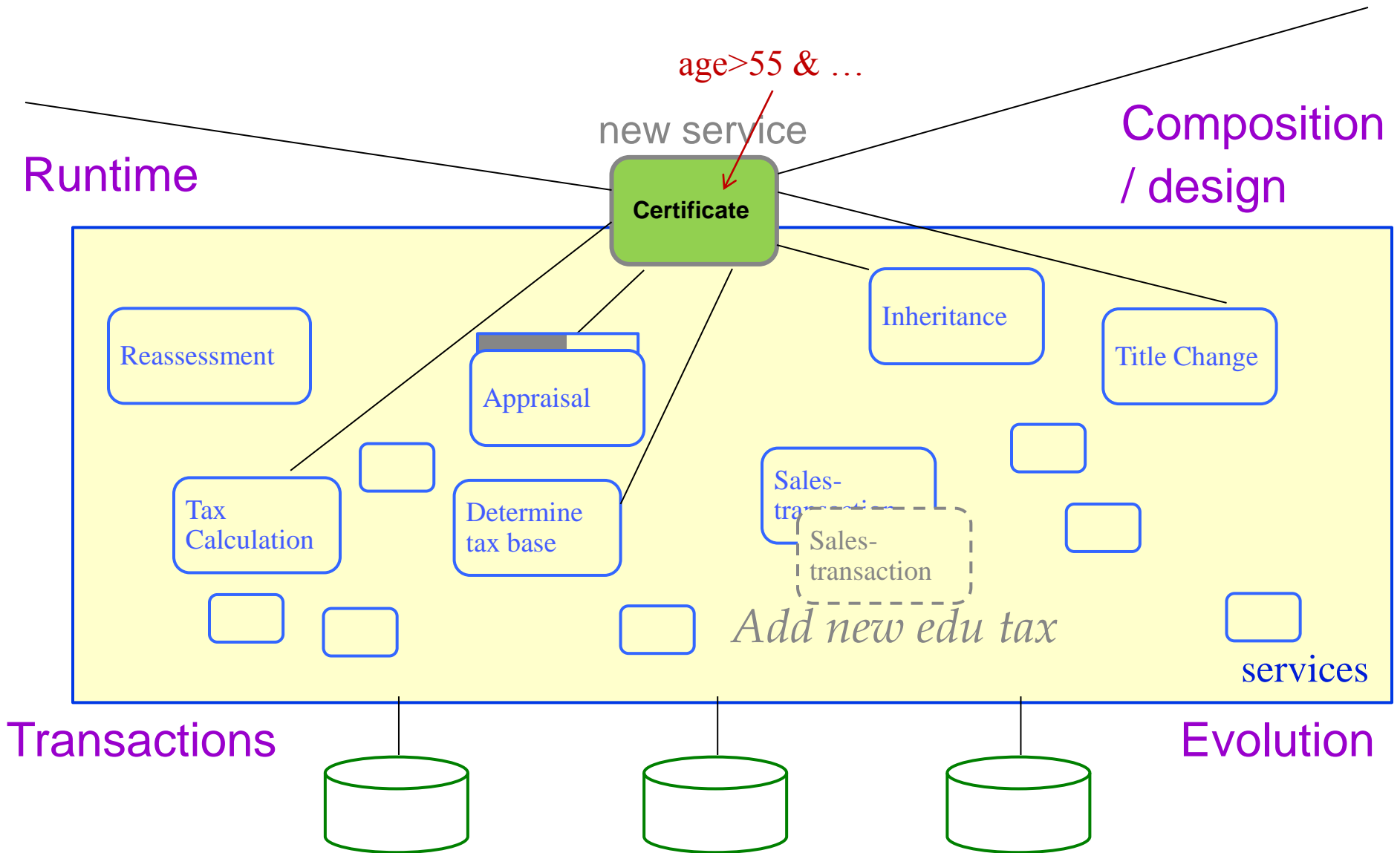
Towards a "Flat World" of Services

- SeGA is a first step but an ad hoc prototype more structured methodology needed
- Conceptual level:
 - ❖ More types of data? Resource models? Transaction issues?
 - ❖ Technical problems—four areas
- System level:
 - ❖ Architecture for data encapsulating services?
 - ❖ APIs for (non-)functional properties?
- Goal: a unified technical framework for services (biz processes and otherwise)

Outline

- Application Needs
- “Legacy” Services
- “Programmable” Services
- Data Encapsulating Services
- Research Challenges
- Conclusions

Research Challenges



Research Challenges

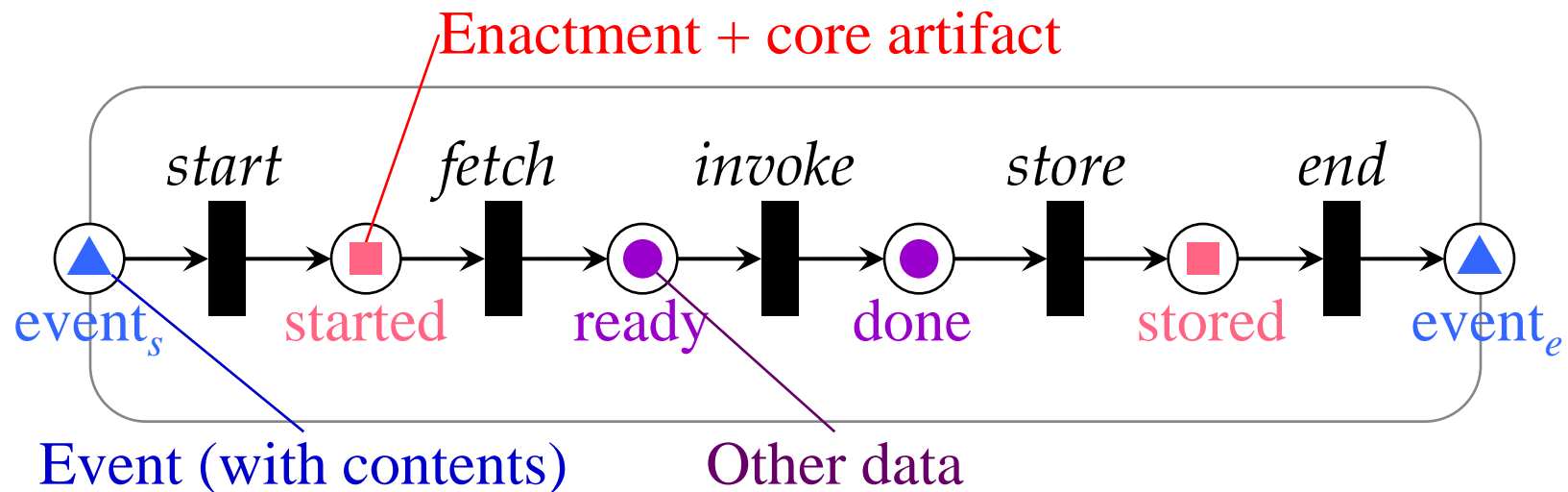
- **Design:** What are appropriate service designs?
Choreography vs orchestration (Part II)? Design aid (analysis/model checking tool), interoperation
- **Runtime:** Enforcement of process/data constraints, KPI/monitoring techniques, resource planning and management
- **Transactions:** What is the notion of workflow transaction?
- **Change/evolution:** Process vs instance changes, long lasting vs temporary, longtail
- **Big data:** monitoring to analytics to change

Choreography For Artifacts

- Participants are processes represented by biz artifacts
 - ❖ Partial information model visible
- Correlations between process instances (not just models)
- Data from artifacts used in specifying sequencing constraints
- A fragment of first-order linear time logic
- Detailed in the afternoon session [Sun-Xu-S. ICSSOC 12]

Execution Semantics

- Formal model (semantics) for task execution based on Petri nets



- Represents data (input/output) requirements and carries enactments

[Xu-S.-Yan-Yang-Zhang CoopIS 2011]

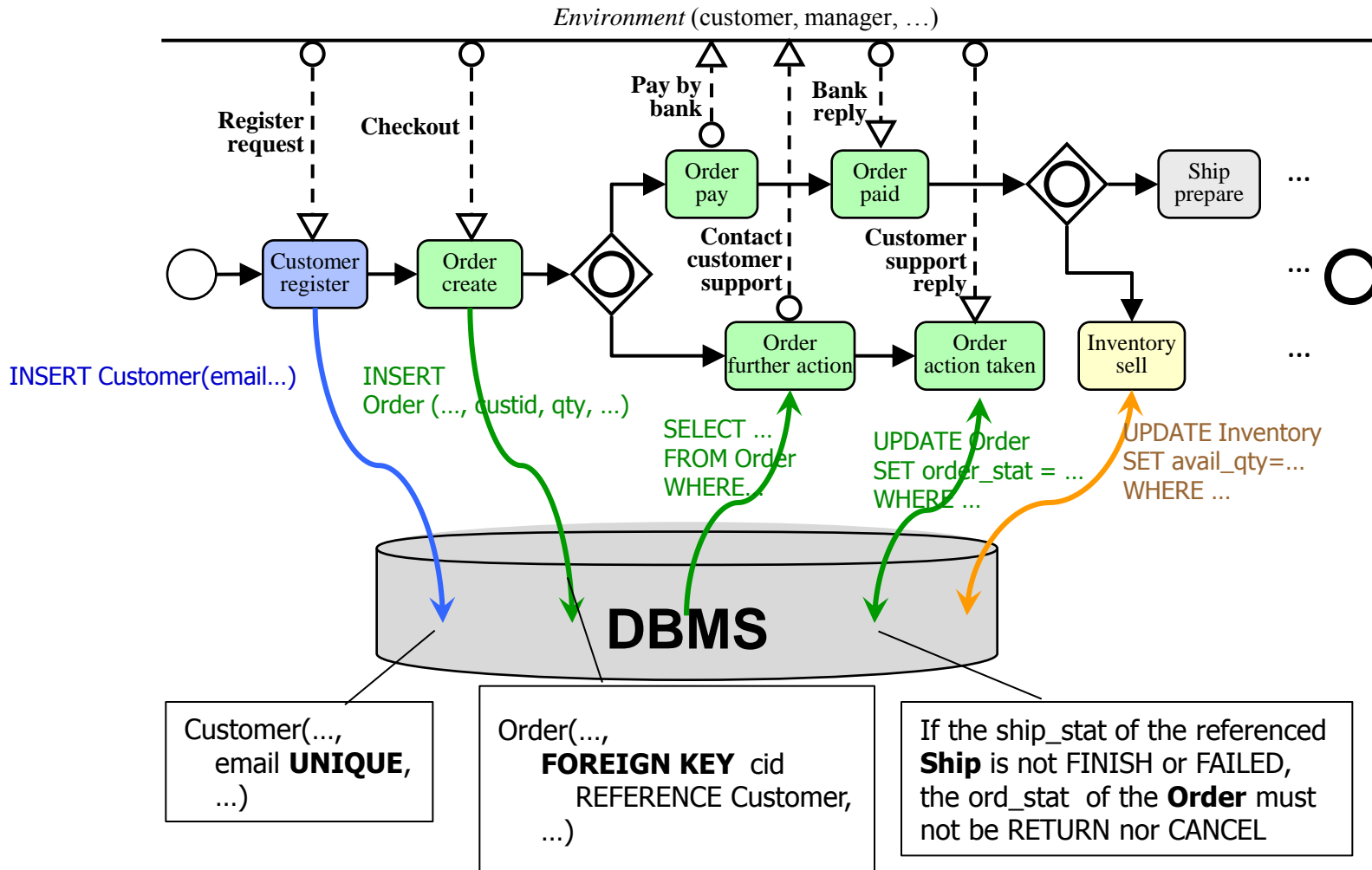
Artifact-Centric BPs are Easier to Change

- Biz process = biz artifacts = state machine lifecycle + **BP change rules**
- **BP change rules** conservatively extend workflow
 - ❖ Could be temporary, non-schematic
- Rules allow biz processes to respond to situations with many more options
- **Estimated labor savings:**
 - ❖ 9% for Hangzhou HMB (preliminary study) or 38 out of 400 FTEs

[Xu-S.-Yan-Yang-Zhang CoopIS 2011]

Workflow and Data Management

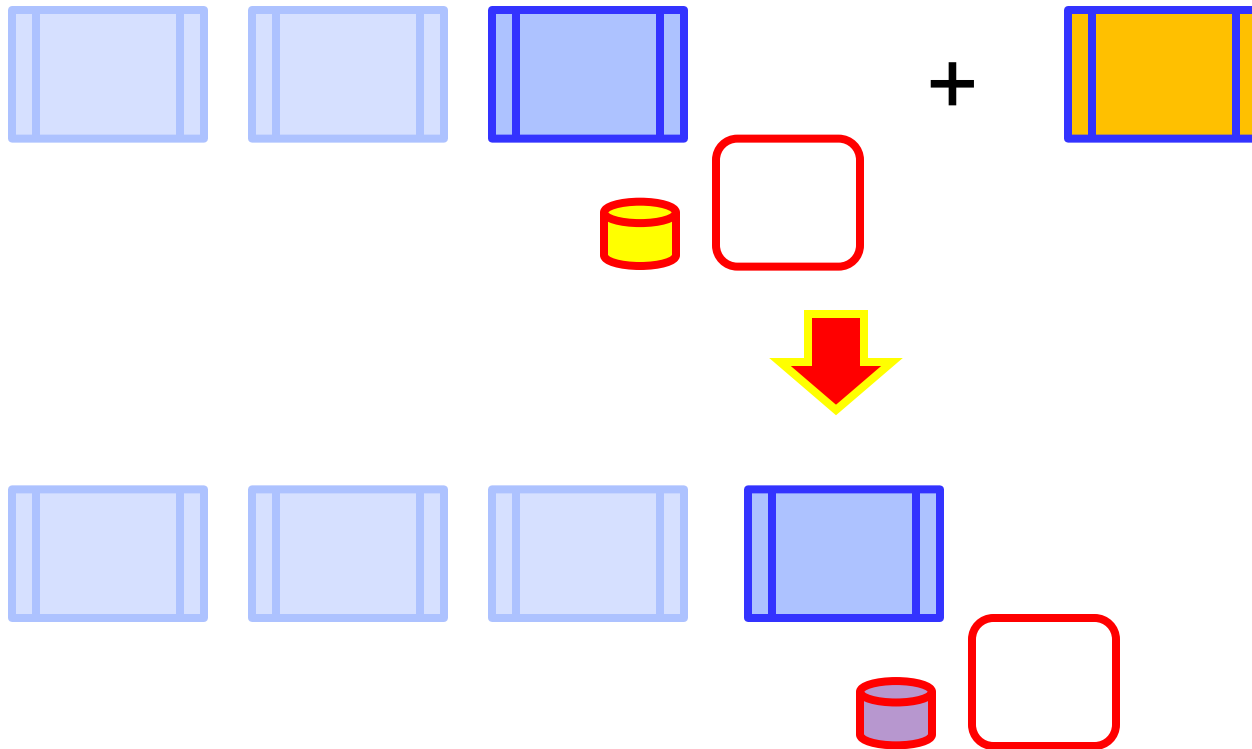
■ Integrity constraints (ICs): key in data management



■ Many possible ways, our approach: Guard injection

Incremental (Runtime) Enforcement

- Logical properties: first order + linear time logic
- Execution snapshot: relational database



[De Masellis-S. 2012]

Research Challenges

- **Design:** What are appropriate service designs?
Choreography vs orchestration (Part II)? Design aid (analysis/model checking tool), interoperation
- **Runtime:** Enforcement of process/data constraints, KPI/monitoring techniques, resource planning and management
- **Transactions:** What is the notion of workflow transaction?
- **Change/evolution:** Process vs instance changes, long lasting vs temporary, longtail
- **Big data:** monitoring to analytics to change

Outline

- Application Needs
- “Legacy” Services
- “Programmable” Services
- Data Encapsulating Services
- Research Challenges
- **Conclusions**

Conclusions

- Inclusion of data is critical to capture business logics into services
 - ❖ Data are not just variables; important to remember “persistence”
- Separation of data and service management is a promising approach (e.g., BYOD)
 - ❖ DSM independence
- Problems are more difficult, demand creative solutions!
 - ❖ Do we have alternatives?
- Scientific principles can and should guide engineering practice, but they don't have to speak the same buzz words

Acknowledgements

Helpful comments from:

- Richard Hull (IBM)
- Yutian Sun (UCSB)
- Jian Yang (Macquarie U)

References

- [Hull-S. 04] R. Hull and J. Su: Tools for Design of Composite Web Services. ACM SIGMOD, 2004, pages 958-961
- [Hull-S. 05] R. Hull and J. Su: Tools for composite web services: A short overview. SIGMOD Record, 34(2):86-95, 2005
- [Jin et al CoopIS 2011] T. Jin, J. Wang, and L. Wen: Efficient Retrieval of Similar Business Process Models Based on Structure. CoopIS 2011, pages 56-63
- [Berardi-Calvanese-DeGiacomo-Lenzerini-Mecella ICSOC 03] Daniela Berardi, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Massimo Mecella: Automatic Composition of E-services That Export Their Behavior. ICSOC 2003: 43-58
- [Gerede-Hull-Ibarra-S. ICSOC 04] C.E. Gerede, R. Hull, O.H. Ibarra, and J. Su: Automated composition of e-services: lookaheads. ICSOC, 2004, pages 252-262
- [Berardi et al ICSOC 04] D. Berardi, G. De Giacomo, M. Lenzerini, M. Mecella, and D. Calvanese: Synthesis of underspecified composite *e*-services based on automated reasoning. ICSOC, 2004, pages 105-114
- [Berardi-Calvanese-De Giacomo-Hull-Mecella VLDB 05] D. Berardi, D. Calvanese, G. De Giacomo, R. Hull, M. Mecella: Automatic Composition of Transition-based Semantic Web Services with Messaging. VLDB, 2005, pages 613-624
- [Gerede-Ibarra-Ravikumar-S. TCS 08] C. E. Gerede, O. H. Ibarra, B. Ravikumar, and J. Su: Minimum-cost delegation in service composition. Theor. Comput. Sci., 409(3):417-431, 2008
- [Hassen-Nourine-Toumani ICSOC 08] R. R. Hassen, L. Nourine, F. Toumani: Protocol-Based Web Service Composition. ICSOC, 2008, pages 38-53
- [Fu-Bultan-S. WWW 04] X. Fu, T. Bultan, and J. Su: Analysis of interacting BPEL web services. WWW, 2004, pages 621-630
- [Fu-Bultan-S. ISSTA 04] X. Fu, T. Bultan, and J. Su: Model checking XML manipulating software. ISSTA, 2004, pages 252-262
- [Gerede-S. ICSOC 07] C. E. Gerede and J. Su: Specification and Verification of Artifact Behaviors in Business Process Models. ICSOC, 2007, pages 181-192
- [Brand-Zafiropulo JACM 83] D. Brand and P. Zafiropulo. On communicating finite-state machines. Journal of the ACM, 30(2):323-342, 1983
- [Nigam-Caswell IBM Sys J 03] A. Nigam, N.S. Caswell: Business artifacts: An approach to operational specification. IBM Systems Journal, 42(3):428-445, 2003
- [Gerede-Bhattacharya-S. SOCA 07] C.E. Gerede, K. Bhattacharya, and J. Su: Static Analysis of Business Artifact-centric Operational Models. SOCA, 2007, pages 133-140
- [Gerede-S. ICSOC 07] C.E. Gerede and J. Su: Specification and Verification of Artifact Behaviors in Business Process Models. ICSOC, 2007, pages 181-192
- [Bhattacharya-Gerede-Hull-Liu-S. BPM 07] K. Bhattacharya, C. E. Gerede, R. Hull, R. Liu, and J. Su: Towards Formal Analysis of Artifact-Centric Business Process Models. BPM, 2007, pages 288-304
- [Hull et al WSFM 2010] R. Hull, E. Damaggio, F. Fournier, M. Gupta, F.T. Heath, S. Hobson, M.H. Linehan, S. Maradugu, A. Nigam, P. Sukaviriya, R. Vaculín: Introducing the Guard-Stage-Milestone Approach for Specifying Business Entity Lifecycles. WS-FM, 2010, pages 1-24
- [Sun-Xu-S.-Yang CoopIS 12] Y. Sun, W. Xu, J. Su, and J. Yang: SeGA: A Mediator for Artifact-Centric Business Processes, CoopIS, 2012
- [Sun-Xu-S. ICSOC 12] Y. Sun, W. Xu, and J. Su: Declarative Choreographies for Artifacts. ICSOC, 2012, pages 420-434
- [Xu-S.-Yan-Yang-Zhang CoopIS 2011] W. Xu, J. Su, Z. Yan, J. Yang, and L. Zhang: An Artifact-Centric Approach to Dynamic Modification of Workflow Execution. CoopIS, 2011, pages 256-273
- [Liu-S.-Yang CoopIS 11] X. Liu, J. Su, J. Yang: Preservation of Integrity Constraints by Workflow. CoopIS, 2011, pages 64-81
- [De Masellis-S. 2012] R. De Masellis and J. Su: Runtime Enforcement of First-order LTL Properties in Data-centric Workflows, in preparation, 2012