

# Separating Execution and Data Management: A key to Business-Process-as-a-Service (BPaaS)

---

Yutian Sun    Jianwen Su  
UC Santa Barbara

Jian Yang  
Macquarie U

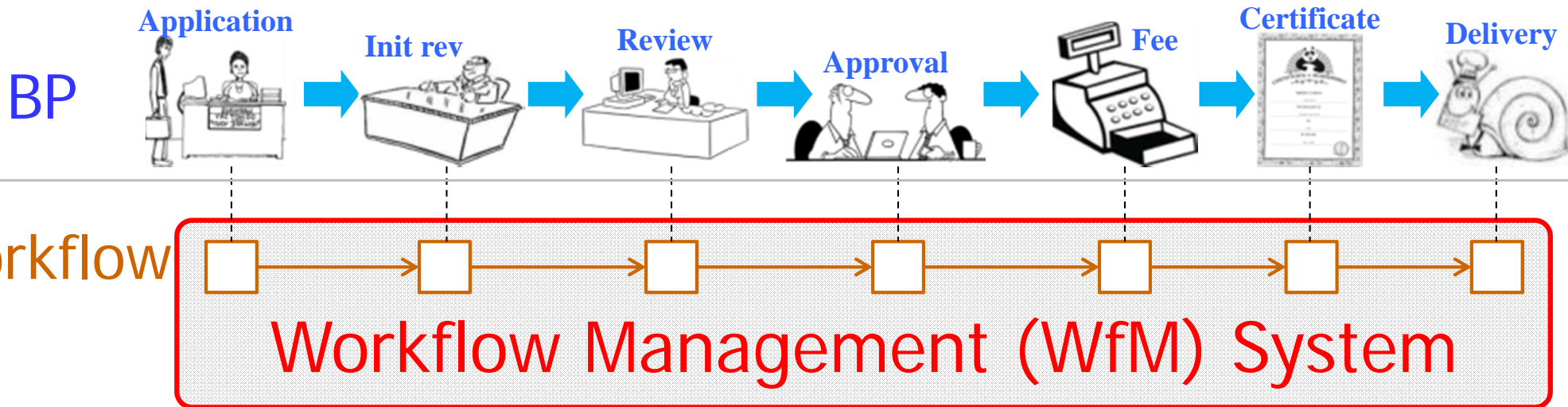
# Outline of the Talk

---

- The technical problem for supporting BPaaS
- Self-guided artifacts
- The SeGA Framework (support for BPaaS)
- Conclusions

# Business Processes & Workflow Management

- A BP is an assembly of tasks to accomplish an objective
  - ❖ Eg: Obtaining a Permit

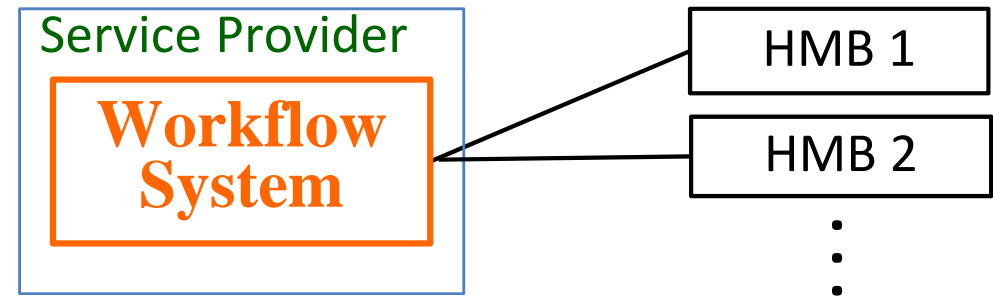


- Each **workflow** model matches a BP model
- Each workflow activity is a software program (□) that interfaces one task in BP
- **WfM system** manages executions, resources, documents, etc.

# Application Example: Property Management

- A Housing Management Bureau (HMB 房管局) manages *titles, licenses, permits, ...* for a region
- Each HMB runs its **own workflow system** for its business
- 30 provinces in China, each province has 10-50 HMBs
- Those hundreds HMB **workflow systems** are individually maintained—**very costly!**

- **Can one workflow system support multiple HMBs?**

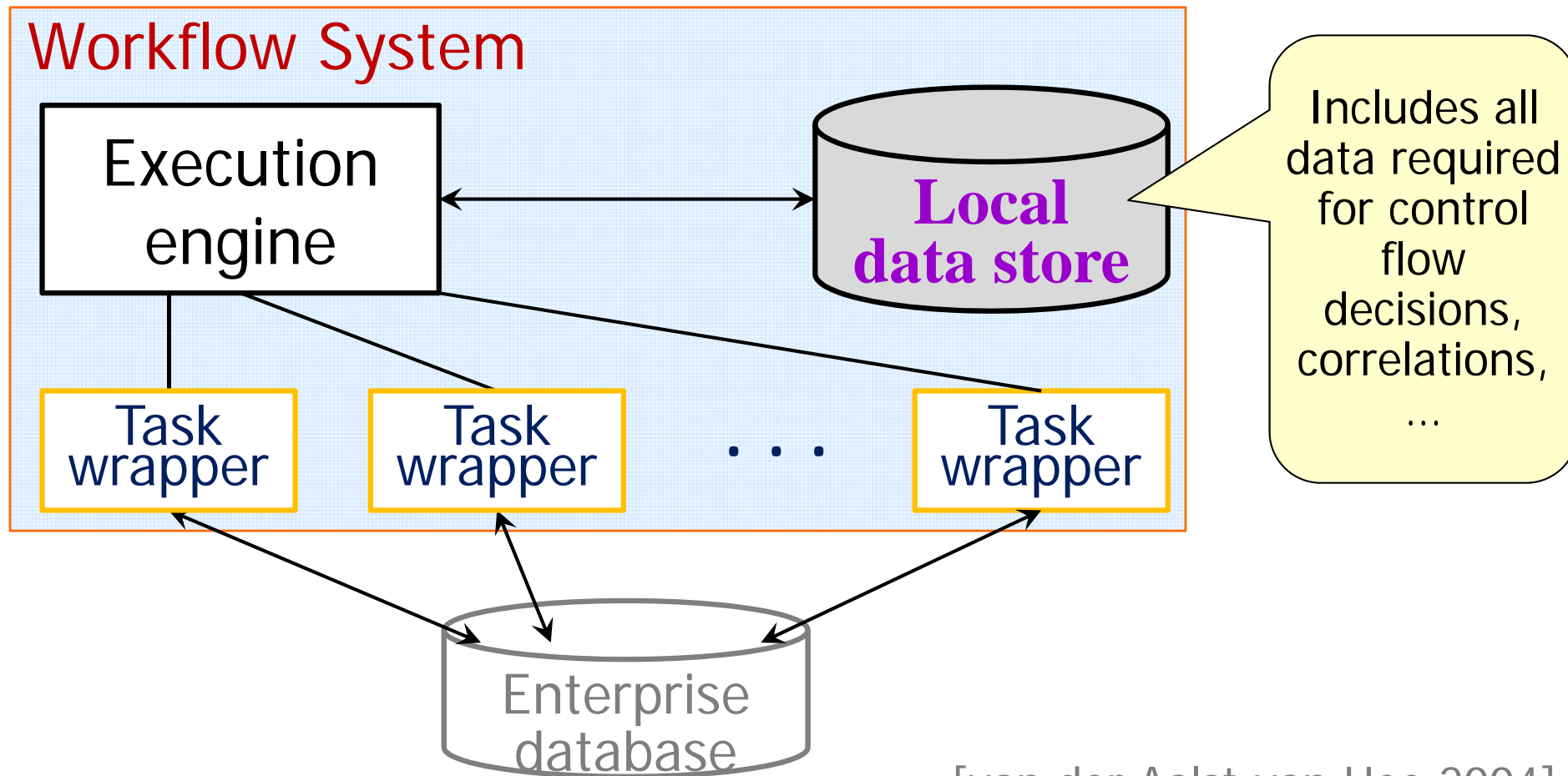


- Note:

- (1) BPs are similar but not identical (provincial/local policies); data sets are mostly disjoint (e.g., data for buildings)
- (2) BPs change often (e.g., low income housing added in 2013)

# Workflow Systems and Local Data

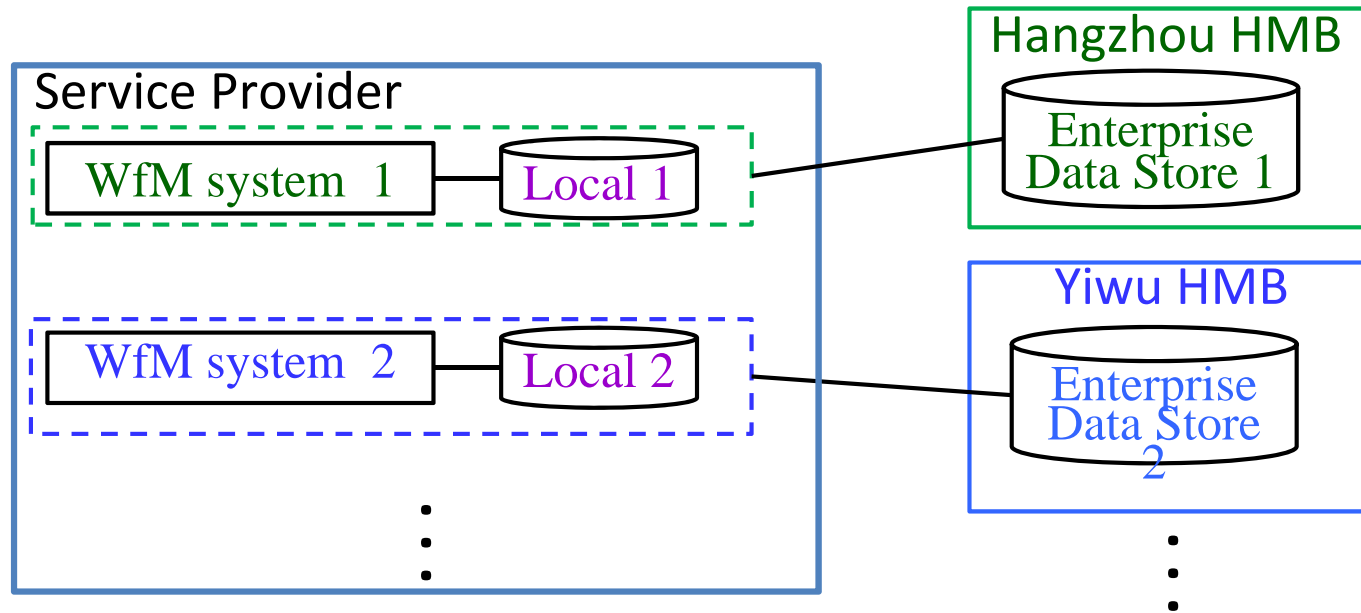
- A **workflow system** typically maintains *execution states, control flow related data, correlations, etc.* in a **local data store**



[van der Aalst-van Hee 2004]

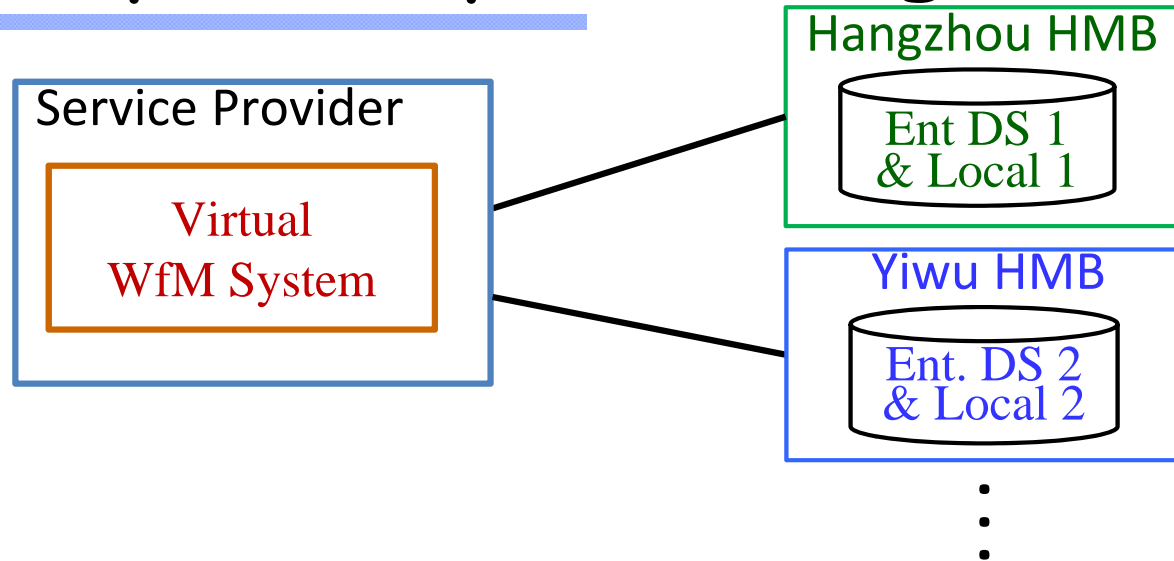
# Workflow Mgmt as Services? Not Effective

- With straight-forward approach, the service provider has to run one WfM system for each client HMB due to disparate **local data** for each HMB's workflow instances



- Arguments for keeping data at local sites:
  - ❖ Analytics with local enterprise data + WF control data
  - ❖ Heterogeneity of data across different local sites
  - ❖ HMB's want to hold their own data

# Our Goal: Separate processing from data



- Only one WfM system that maintains no data
- Keep each HMB's data separated
  - ❖ Independent decision where/how Hangzhou, Yiwu maintain their data
- A key enabling idea:
  - ❖ For each WF step, ship relevant data to WfM System
  - ❖ How to know which data is relevant?
  - ❖ Use BP model with holistic notion of data + process

# Self-Guided Artifacts (sga)

*A self-guided artifact contains everything an engine needs*

- A traditional business artifact (model):

*(Entity information model, Entity lifecycle model)*

- A self-guided artifact:

(BP specification, *Entity*, States, Dependencies, *L*)

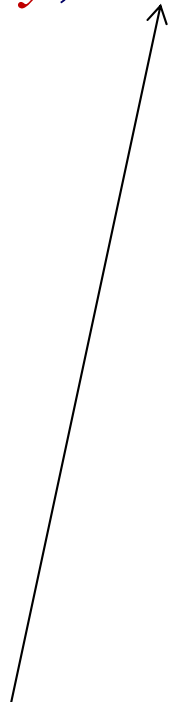
A document with the specification of the *entity lifecycle* model

ID of modeling language

actual business data

current states

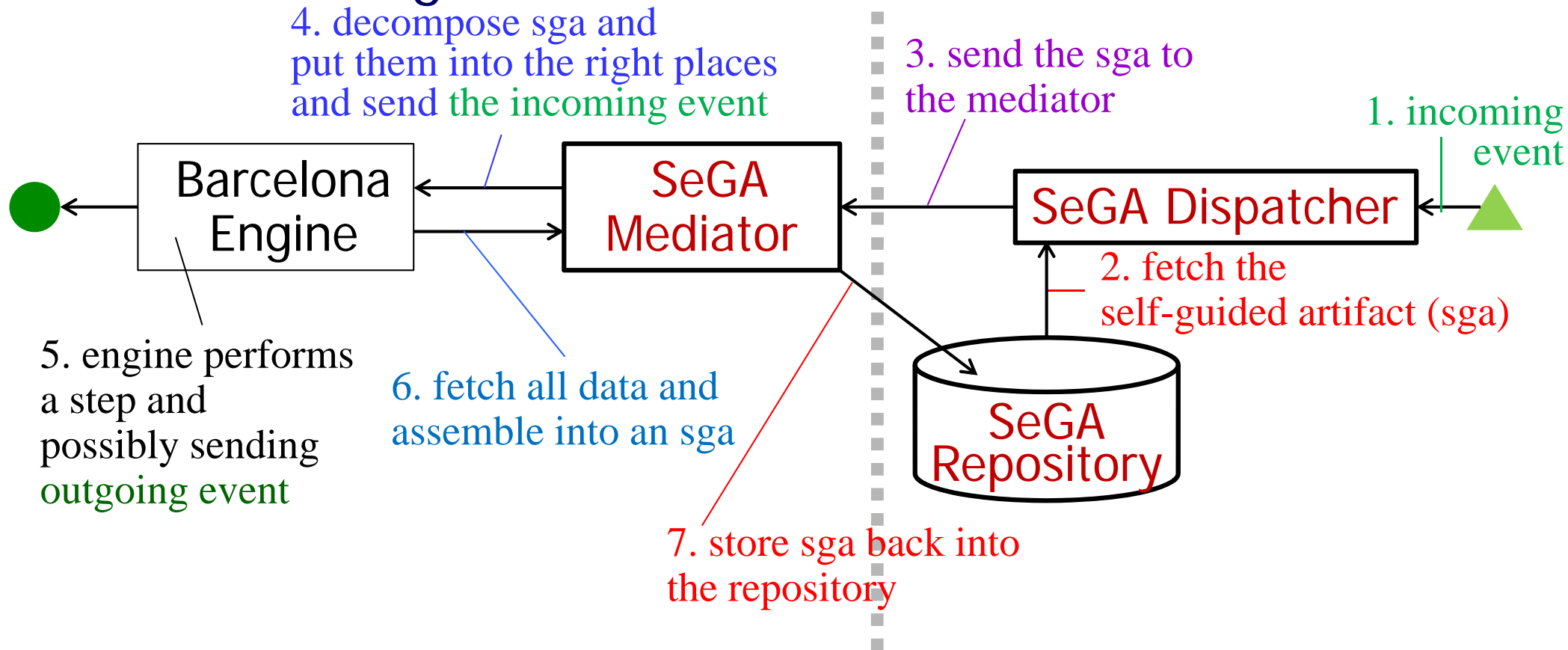
correlations resources





# The SeGA Framework

- Key idea: a process wrapper to supply all data (i.e., self-guided artifact) when the engine needs to run



- Both Barcelona and EZ-Flow are integrated with SeGA

# Supporting BPaaS

## WfM System (Service Provider)

4. decompose sga and put them into the right places and send the incoming event

Barcelona Engine

SeGA Mediator

3. send the sga to the mediator

1. incoming event

SeGA Dispatcher

2. fetch the sga

SeGA Repository

Hangzhou HMB

5. engine performs a step and possibly sending outgoing event

6. fetch all data and assemble into an sga

7. store sga back into the repository

SeGA Dispatcher

2. fetch the sga

SeGA Repository

YiWu HMB

- WfM system maintains no local data\*
- SeGA repository can be merged into enterprise db

[Sun-Su-Wu-Yang ICDE 2014]

\*in current prototype some data stored in Server due to ID mgmt issue 10

# SeGA Prototype and Current State

---

- A prototype developed that works with Barcelona and EZ-Flow
- Two operational test examples:  
Hangzhou HMB's (cooperative) BPs:
  - ❖ Early-sell permit approval (Barcelona via SeGA)
  - ❖ Maintenance space check (EZ-Flow via SeGA)
- A new project funded by Ministry of Sci. & Tech (China), Zhejiang Province, and Hangzhou City (>1.2m RMB):
  - ❖ To implement an operational prototype with a small number of HMBs

# Summary

---

- Main difficulty of providing BPaaS lies in **data management**
- Workflow systems need to have a holistic approach to managing **all of its data**
  - ❖ Self-guided artifact is one such approach
- SeGA elevates existing systems to support:  
**Independence of Data and Execution Management**
- With the principle, BPaaS is actually easy to support
  - ❖ Basis for moving workflow services to cloud while enabling local sites to maintain their data

# Some research challenges

---

- Conceptual data modeling for BPs beneficial, but SeGA raises many technical questions to be addressed:
  - ❖ Inefficient to ship large sega instance due to large data sets
    - Identify subset of artifact data needed for a given WF step?
  - ❖ Parallelism/txn consistency – what if 2 WfM steps require overlapping sets of artifact instances?
    - Use locking, and also subsets of artifact data
  - ❖ Auto-ID issue: current BP systems typically use DBMS to perform ID mgmt
    - Current SeGA prototype: retains some data in WfM server
    - For general solution the virtual WfMS server needs to explicitly manage artifact IDs
- Re-thinking WfMS arch. to include the above will enable robust support for the data independence principle