

Database and Application Interoperability in Industry:

Requirements, Technology, Research & Case Studies

Prof. Amit Sheth

Large Scale Distributed Information Systems Lab
Dept. of Comp. Science, The Univ. of Georgia
415 Graduate Studies Research Center
Athens GA 30602 USA

Tel. +1 706-542-2310, Fax: -2966
Email: amit@ajax.csrg.uga.edu

Tutorial notes, COMAD, December 1994, Bangalore, India.

Transactional Workflows: Research, Technologies, Markets and Applications

Amit Sheth

Large Scale Distributed Info. Sys. Lab
Dept. of Computer Science
415 GSRC, University of Georgia
Athens GA 30602-7404 USA
e-mail: amit@cs.uga.edu
Phone: (706) 542-3455, Fax: (706) 542-2966

*[Based on the Invited Talk at 10th Data Engineering, Feb. 1994, and
Tutorial at VLDB94, Chile, and COMAD94, India.]*

Collaborations/Acknowledgments: U. of Houston (M. Rusinkiewicz, ..),
Bellcore (N. Krishnakumar,..), GTE (D. Georgakopoulos)
MCC (M. Singh,..), ETH-Zurich, ..

Transactional Workflows: Research, Technologies, Markets and Applications

Amit Sheth

Large Scale Distributed Info. Sys. Lab
Dept. of Computer Science
415 GSRC, University of Georgia
Athens GA 30602-7404 USA
e-mail: amit@cs.uga.edu
Phone: (706) 542-3455, Fax: (706) 542-2966

*[Based on the Invited Talk at 10th Data Engineering, Feb. 1994, and
Tutorial at VLDB94, Chile, and COMAD94, India.]*

Collaborations/Acknowledgments: U. of Houston (M. Rusinkiewicz, ..),
Bellcore (N. Krishnakumar,..), GTE (D. Georgakopoulos)
MCC (M. Singh,..), ETH-Zurich, ..

6. Multisystem Applications and Transactional Workflows

- Basic concepts and specification of workflows
- Related research and technologies
- Components of a Workflow Management Systems
 - Prototypes and applications
 - Work in progress



Workflow Examples

- Business loan processing in Banks requires coordinating *user tasks* such as loan application entry and risk exception, and *application tasks* such as risk evaluation, risk update, and loan decision recording.

- Patient health care support in a health care group practice requires coordinating user tasks such as patient registration, doctor's record review/update, and lab work, and application tasks such as automated billing and statistics compilation.

- Service provisioning in Telcos require the coordinated execution of heterogeneous tasks on heterogeneous systems:

- application tasks such as contracts and terminal emulation on various mainframe based Operation Support Systems,
- transactions on workstation database management systems, and
- user tasks to resolve Requests for Manual Assists (RMAs) using Graphical User Interface (GUI) tools.

Other examples:

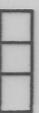
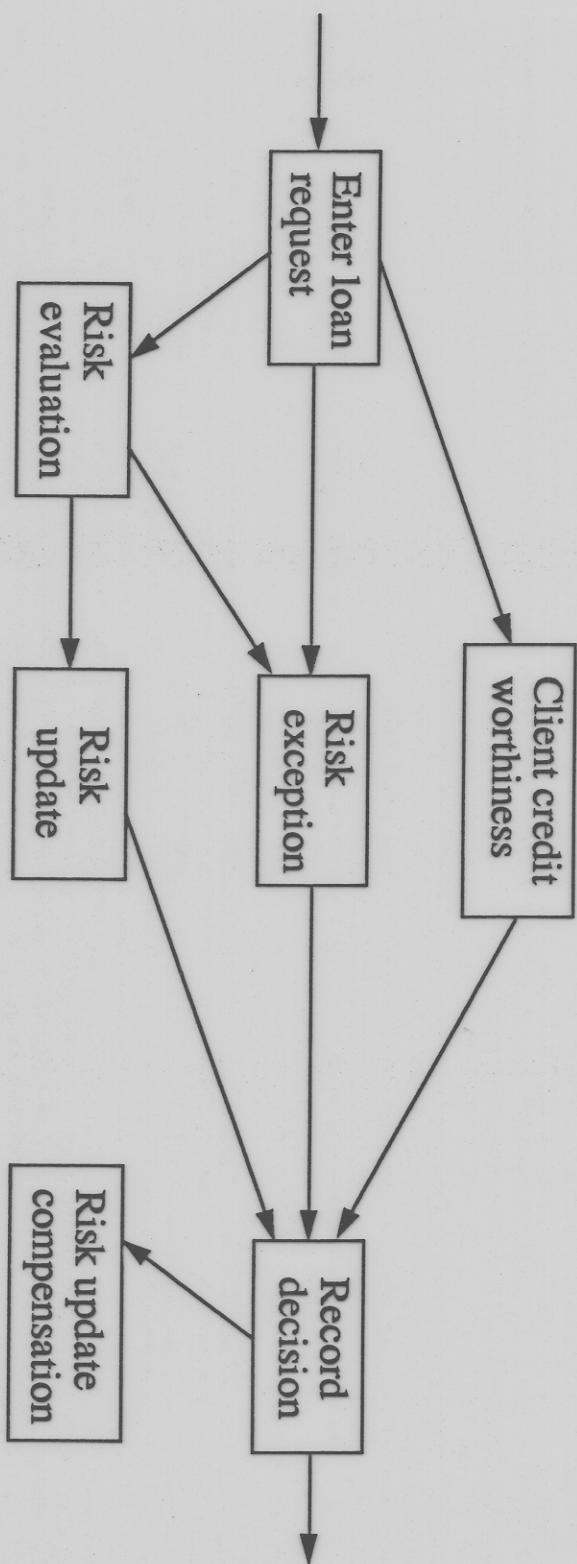
Office computing: electronic mail routing, meeting scheduling, course organizing

Data processing: purchase order processing, Expense voucher processing

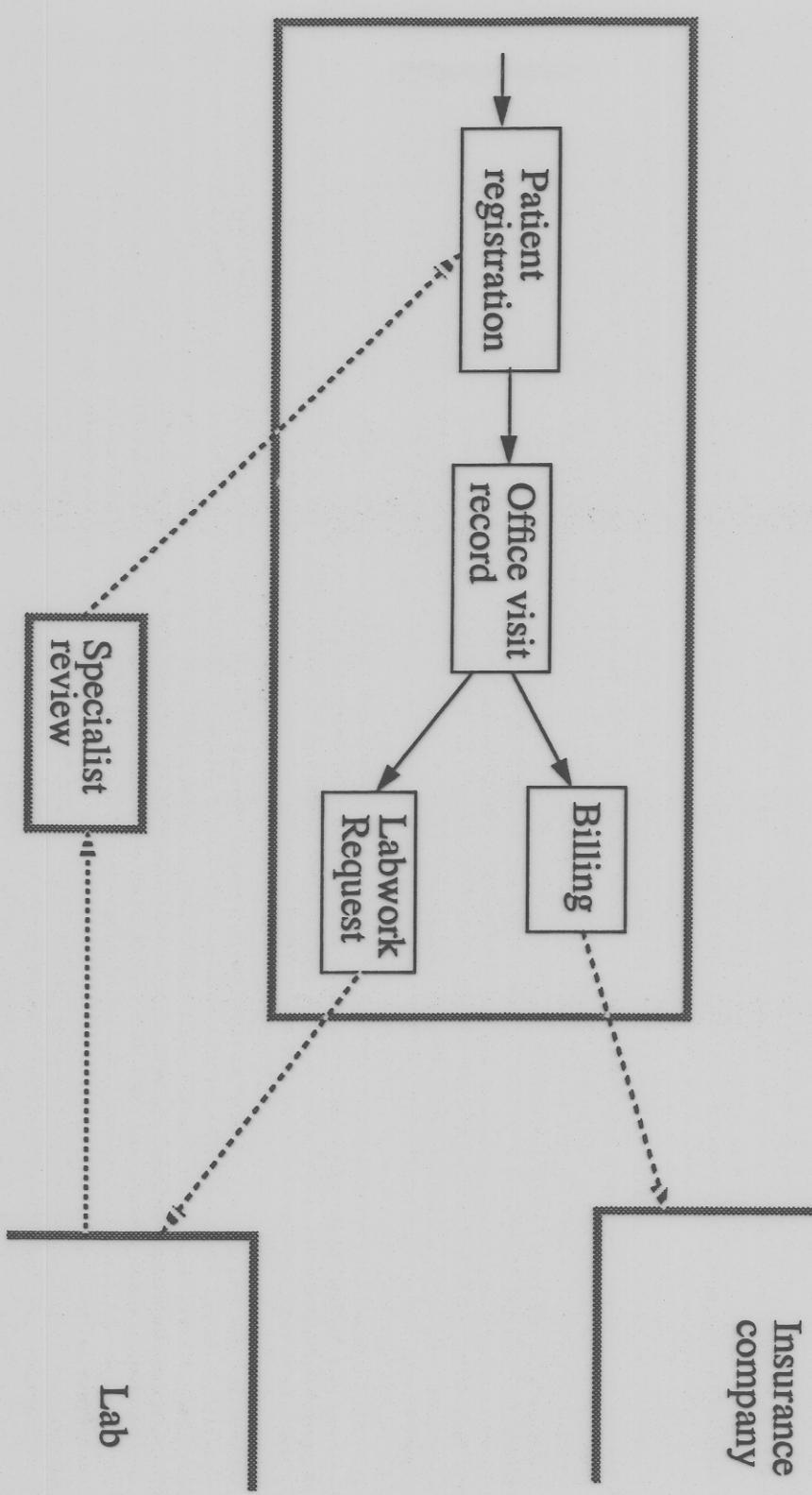
Manufacturing: product life-cycle



An Example Workflow in Business Data Processing: Business Loan Processing



An Example Workflow in a Healthcare Enterprise



Technology Perspective

for future commercial directions towards **TWF**
PLANNING & ENGINEERING

OPERATIONS

Trans.
Workflow

CTM/ETM

Workflow
Software

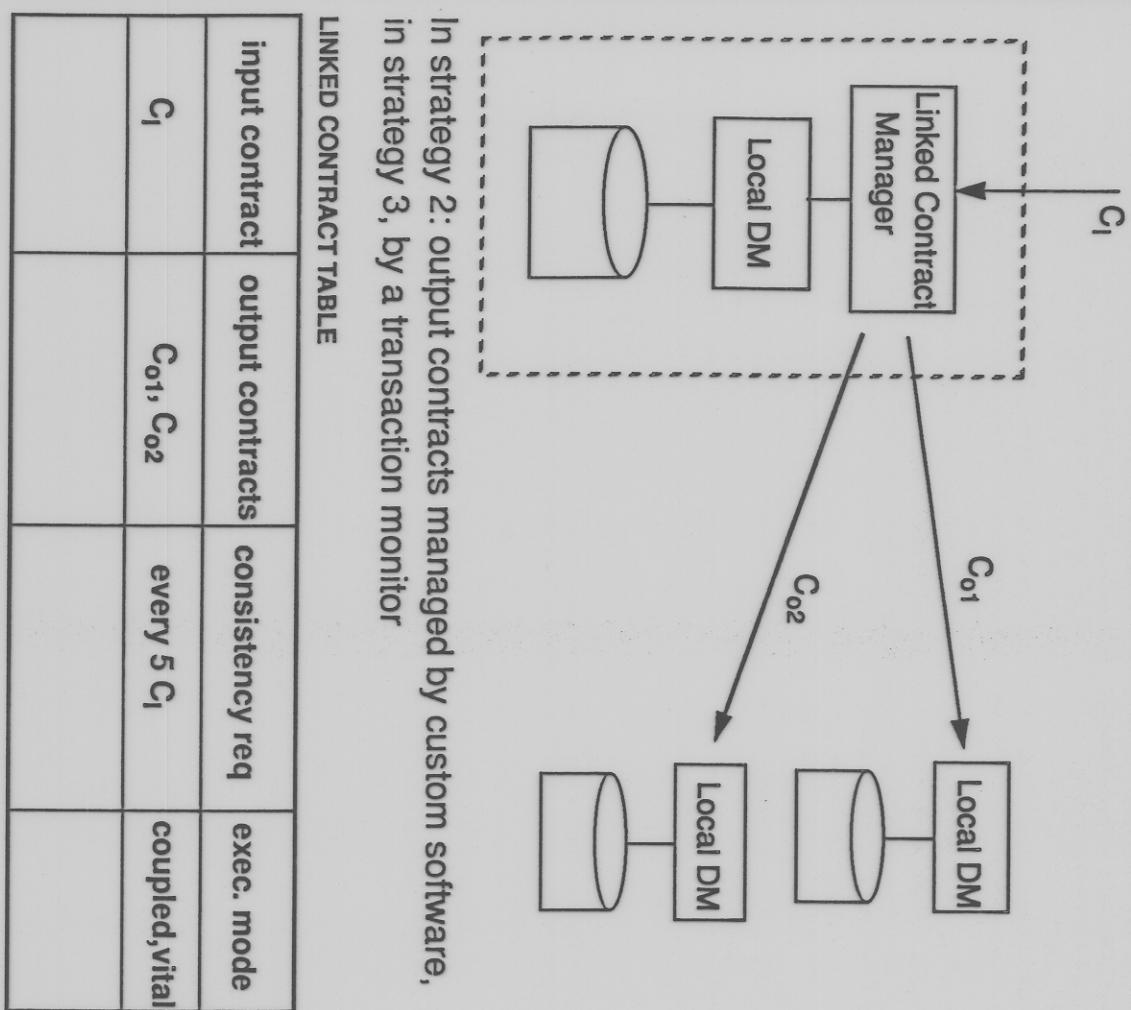
Service Negotiation

Distributed Proc.
Infrastructure
Enablers
- CORBA, DCE,
X.400, X.500

PROVISIONING



Enforcement Strategies 2 and 3



In strategy 2: output contracts managed by custom software, in strategy 3, by a transaction monitor

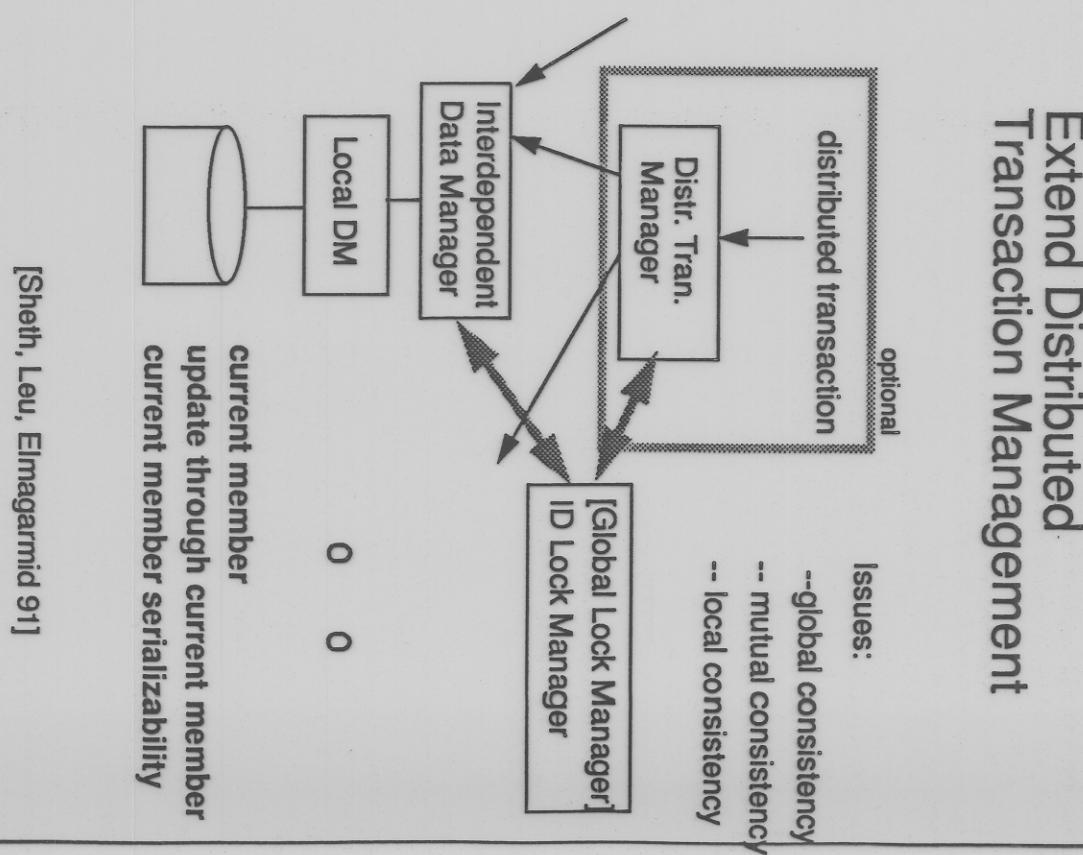
LINKED CONTRACT TABLE

input contract	output contracts	consistency req	exec. mode
C_l	C_{o1}, C_{o2}	every 5 C_l	coupled,vital

[Gomer Thomas 89]

[Gomer Thomas 89]

Enforcement Strategy 5



[Sheth, Leu, Elmagarmid 91]



Enforcement Strategy 4

Approach:

- Replication Server allows data inconsistencies
- Transaction, not data are replicated; copies are transactionally consistent
- Inconsistencies propagate back to a queue at the originator's site

• Choice of transaction models for tradeoff between level of consistency and efficiency/availability:

- primary first
- replication first
- version controlled
- 2PC

Technique:

- Client application updates DBMS.
- Log Transfer Manager sends transaction to Replication Server
- Replication Server forwards transaction to other Replication Servers
- Transactions are queued for unavailable systems

Sybase Replication Server [based on presentation at PDIS'93]



Capabilities of a Modern Data Cleanup System

- customizable, expressive, and flexible specification of simple to very complex data validation and clean-up rules,
- open data access to read, reference, and update data in relational databases using industry standard methods, or interfaces to legacy systems, and
- interactive operations through a graphical user interface, and
- automatic cleanup, using additional rules and programs, either in batch mode or scheduled periodically.



Current Products and Solutions for Data Cleaning

- A sample of current products and solutions:
 - Integrity Programming Environment, Validity Technologies Inc., USA
 - proprietary assembly level implementation for efficient string manipulation on files (not structured data)
 - MVS/TSO environments only
 - Database Synchronizer, Apertus Technologies Inc., USA
 - customized solution in mainframe environments



Current Approaches and Limitations:

- Manual approach: labor-intensive; introduces its own errors
- Semi-automatic approach: based on creating custom application with hard-coded data consistency constraints.
With these approaches:
 - significant effort is required to change the constraints; customers cannot customize them;
 - expressing and enforcing constraints for data present in different databases is difficult; and
 - simultaneously accessing databases from different vendors is very difficult, if not impossible.



An Example

Connects

CAC	FAC	From Co ID	From Type	To Co ID	To Type
...	...				
SC26YT4	00425055	SCZ6406850	EQ	SCZ6406851	CH

Equipment

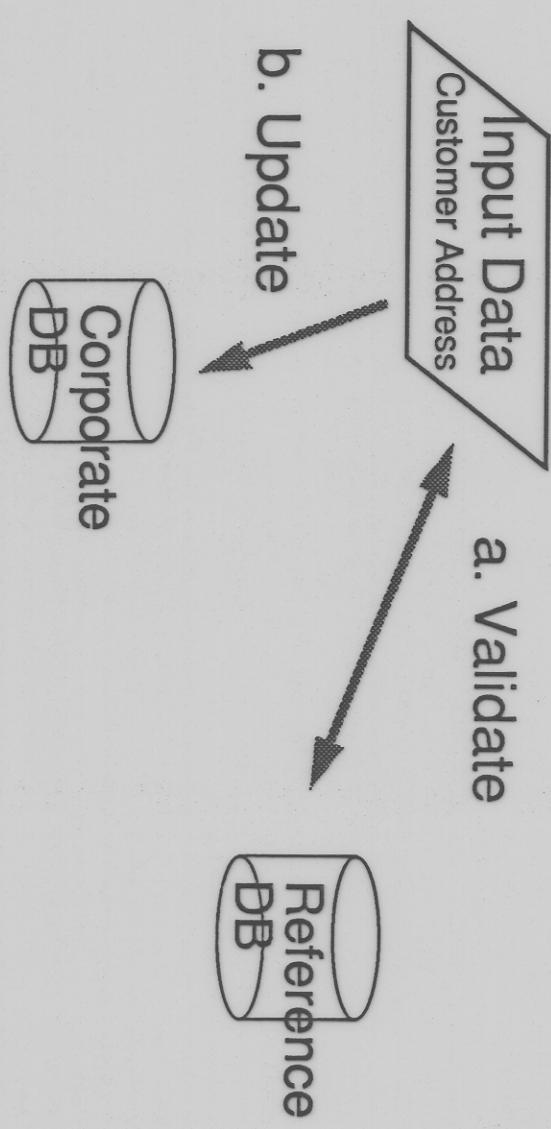
EQPT ID	Action	Loc ID	...
W			
WT		AUSNTXTL	
T			

Channel

Chan ID	Action	Loc A	Loc B	...
...	...			
SCZ6406851	W	FTWOTXH	DLLSTXTL	

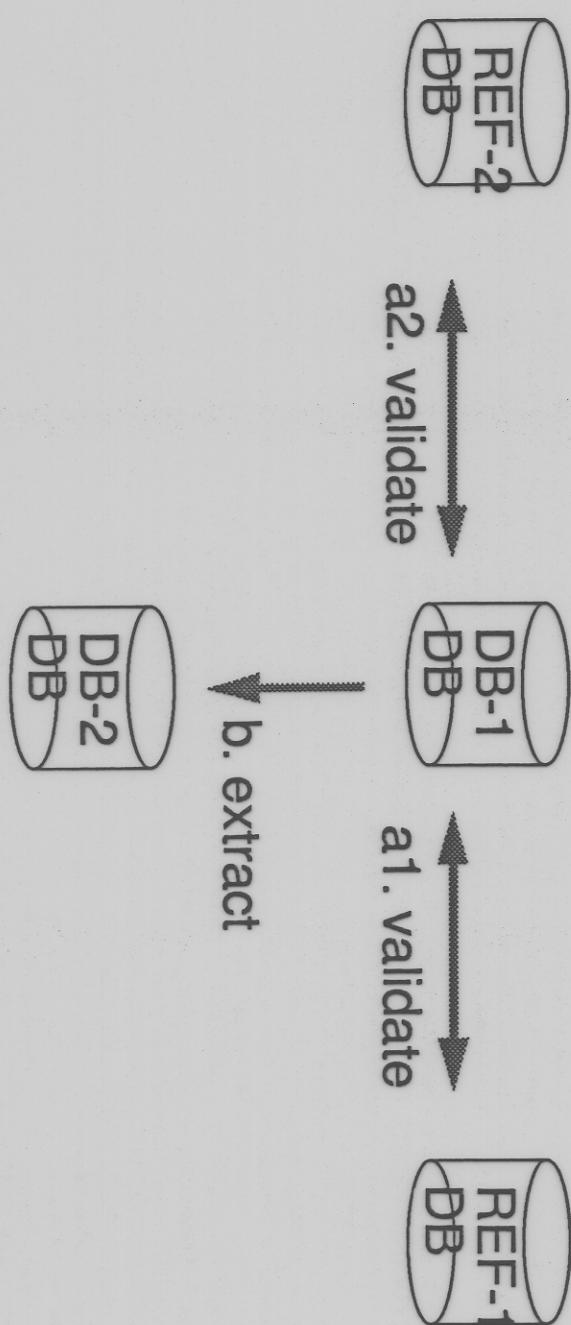
Scenario 1

- Validate input data against reference data before updating an application database.



Scenario 2

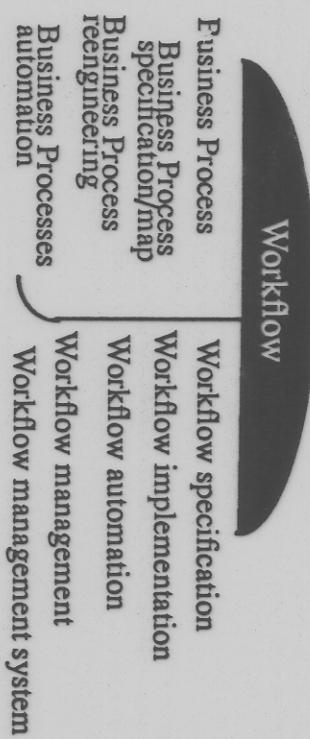
- Validate DB-1 data with reference databases before extracting/uploading data for DB-2.



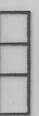
Workflow Management

WFM involves:

1. defining *workflows*, i.e., describing those aspects of a process that are relevant to controlling and coordinating the execution of its tasks (and possibly the skills of individuals or information systems required to perform each task), and
2. providing for automation and re-engineering (fast (re)design and (re)implementation) of the processes as business needs and information systems change



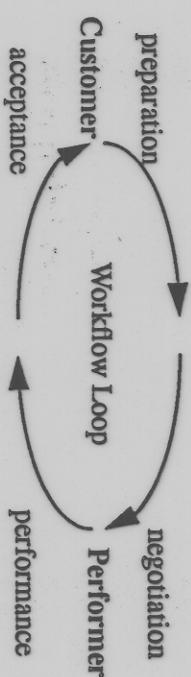
The “Workflow” Umbrella



Process Modeling Methodologies

Communication-based methodologies stem from Winograd/Flores “Conversation for Action Model” [WF87]

- 1.preparation - a customer requests an action to be performed or a performer offers to do some action
- 2.negotiation - both customer and performer agree on the action to be performed and define the terms of satisfaction
- 3.performance - the action is performed according to the terms established
- 4.acceptance - the customer reports satisfaction (or dissatisfaction) with the action

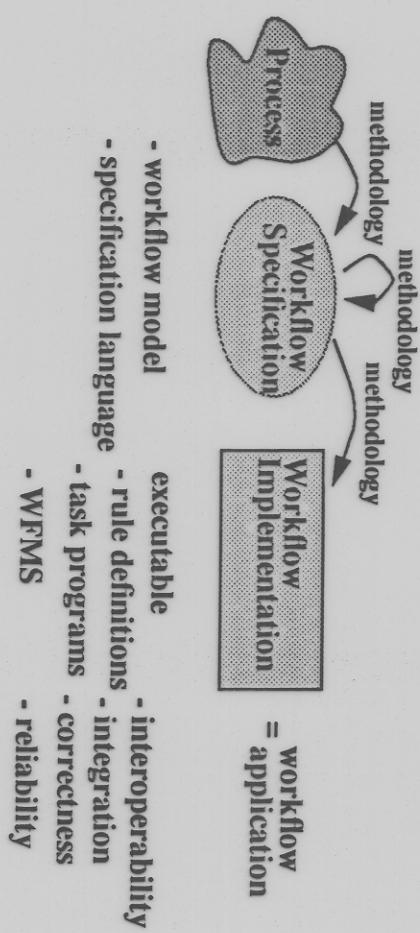


Activity-based methodologies focus on modeling the work instead of modelling the commitments among humans.

[Georgakopoulos, Hornick and Sheth 94]

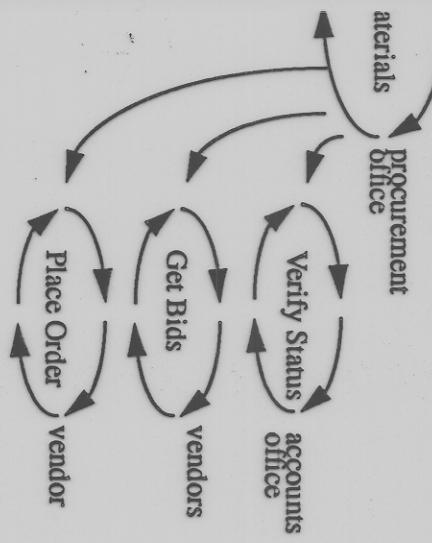


Workflow Management Issues

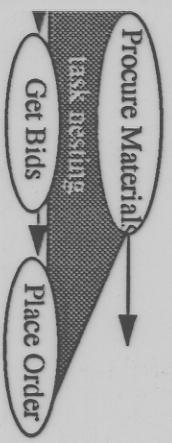


als Procurement Process

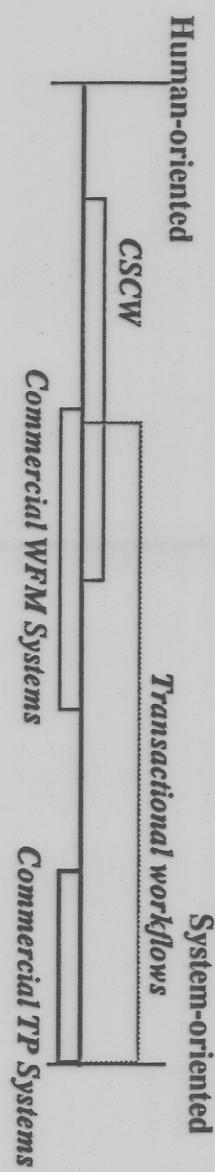
In-based Model:



process modeling:



Another Characterization



Transactional workflows are activities that involve coordinated execution of multiple related tasks on distributed/heterogeneous/autonomous information systems and support selective use of transactional properties at individual task and workflow levels.

In particular, they use transaction management concepts and technology for specifying and ensuring workflow correctness and reliability in distributed/heterogeneous/autonomous information system environments.



Work in Progress...

- Graphical instantiating and monitoring of workflows [MCC, Bellcore]
- Extended Multidatabase/Semantic Transactions (ETM), CTM [Georgakopoulos et al 94a], ATM vs. Transactional Workflow [Breitbart et al 93]
- Concurrency Control and Recovery that exploit application and system semantics [Jin et al 93a,b, Georgakopoulos et 94a]
- Partially distributed scheduler [one scheduler per workflow] (Jin et al 93b)
- Distributed Scheduler [Singh, MCC]
- Graphical specification and testing of workflows[Georgia]
- Workflow simulation environment [Georgia]
- ...

