Plan for Today's Lecture(s)

- Knowledge management
- Content management
- Enterprise Data Management
- Integration and Interoperability
- Supply Chains and Inter-Enterprise Information Exchange

Information Management and the DTS

Knowledge Management — **Content Management** — **Data Management**

"Unstructured Text" Processing

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

31 October 2013
Lecture 19.1 – Knowledge & Content Management

Knowledge Management

- Much collective knowledge is embodied in a firm's people, systems, management techniques, history of strategy and design decisions, customer relationships, and intellectual property like patents, copyrights, trademarks, brands, etc.
- Some of this knowledge is EXPLICIT, tangible, and traceable in the form of documents, databases, organization charts, policy and procedure manuals
- But much of it is TACIT: informal, not systematized in tangible form because it is held in the minds and experiences of people; a synonym is "KNOW-HOW"

Knowledge Management

- The goals of KM can be viewed as getting the tacit parts of this "intellectual capital" to be explicit
- "Knowing how to get things done" or "knowing how things work" includes what a person knows how to do, but also who they know
- So identifying and analyzing the "social network" or "knowledge network" of employees is a key part of knowledge management

Knowledge Management Goals

- Sharing solutions to customer problems
- Facilitating collaboration
- Locating people with relevant skills
- Managing unstructured content
- Providing greater access to existing information
- Improving traceability and justification for strategic (and controversial) decisions
- Recording the rationale for business process and information models

Knowledge Management Issues

- Many technologies have been used for KM -- Lotus Notes, Intranets, Wikis, Blogs...
- But at best, knowledge management techniques can only capture knowledge that is codifiable and transferable, and not all knowledge is
- And furthermore, employees have complex motivations for complying with or not complying with KM goals
- Enlightened firms and management try to align personal and corporate goals for knowledge management

Content Management

- "Content Management" narrowly defined involves the management of semi-structured content in a logical repository, usually in a multi-user collaborative context
- But "content management" necessarily involves authoring and delivery or there would be nothing to manage or no purpose in managing it

Content Authoring

- Authoring can be broadly defined as creating reusable "information assets" from different sources
- Reusable information sometimes means XML, but more generally means information resources with metadata
- Reusable information assets can be created by adding structure and metadata to existing information
- Non-text information assets can be described using XML text metadata

Content Management (narrow sense)

- Reliable storage and retrieval of components, documents, schemas, transforms, stylesheets...
- Componentizing a document by separating it into its constituent elements using user-defined names as boundaries
- Risk management functions like backup and archiving

Component Granularity

- Document level granularity
- Module level granularity
- Content unit level granularity
- Word level granularity

Content Delivery

- Content delivery usually begins when some set of components is retrieved from the repository and assembled to meet some specific requirement
- Assembly may involve both the assembly of a document type model and then the assembly of an instance that conforms to it
- The retrieved or assembled instance may need to be transformed to conform to different models for various contexts, users, or devices

"Flavors" of Content Management

- Document management
  - Often just searchable metadata and content with workflow support
- Web content management
  - More multimedia types than doc mgmt, more real-time workflow
- Digital asset management
  - Movies and audio

"Flavors" of Content Management

- E-mail management
- Records management
- Report management

A Single-Source Strategy

- Informal defintion:
  - Write once, reuse many times
  - Revise once, update everywhere
  - Transform many times for delivery
- Rigorous definition:
  - Enforce normalization techniques to prevent anomalies with duplicate content
  - Use transformations to convert content from one structure or context to another,

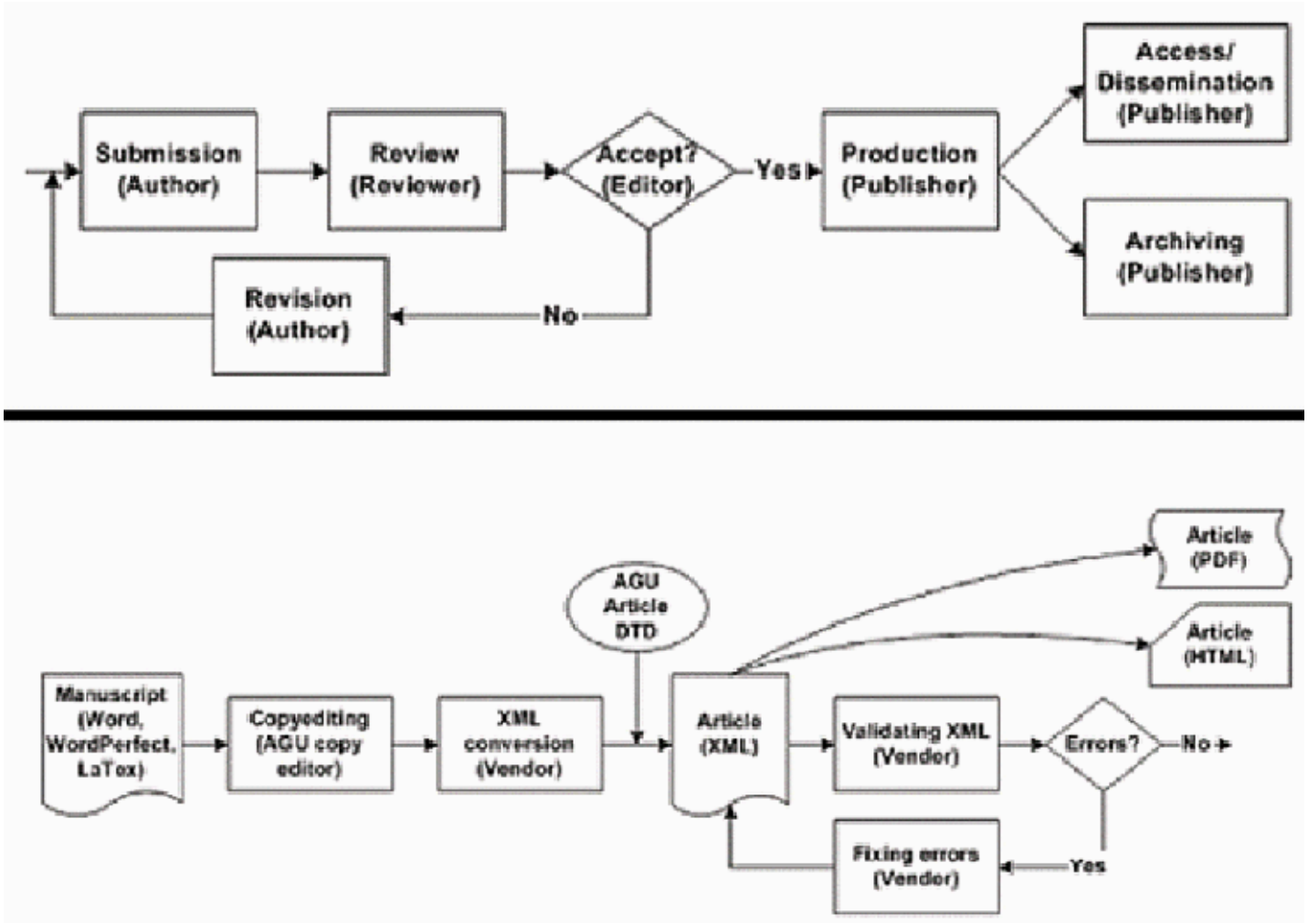
A Single-Source Strategy

- Single-sourcing is NOT a property of any content management system, only a goal to be achieved by one
- CASE STUDY: XML-Centric Workflow in Scholarly Publishing

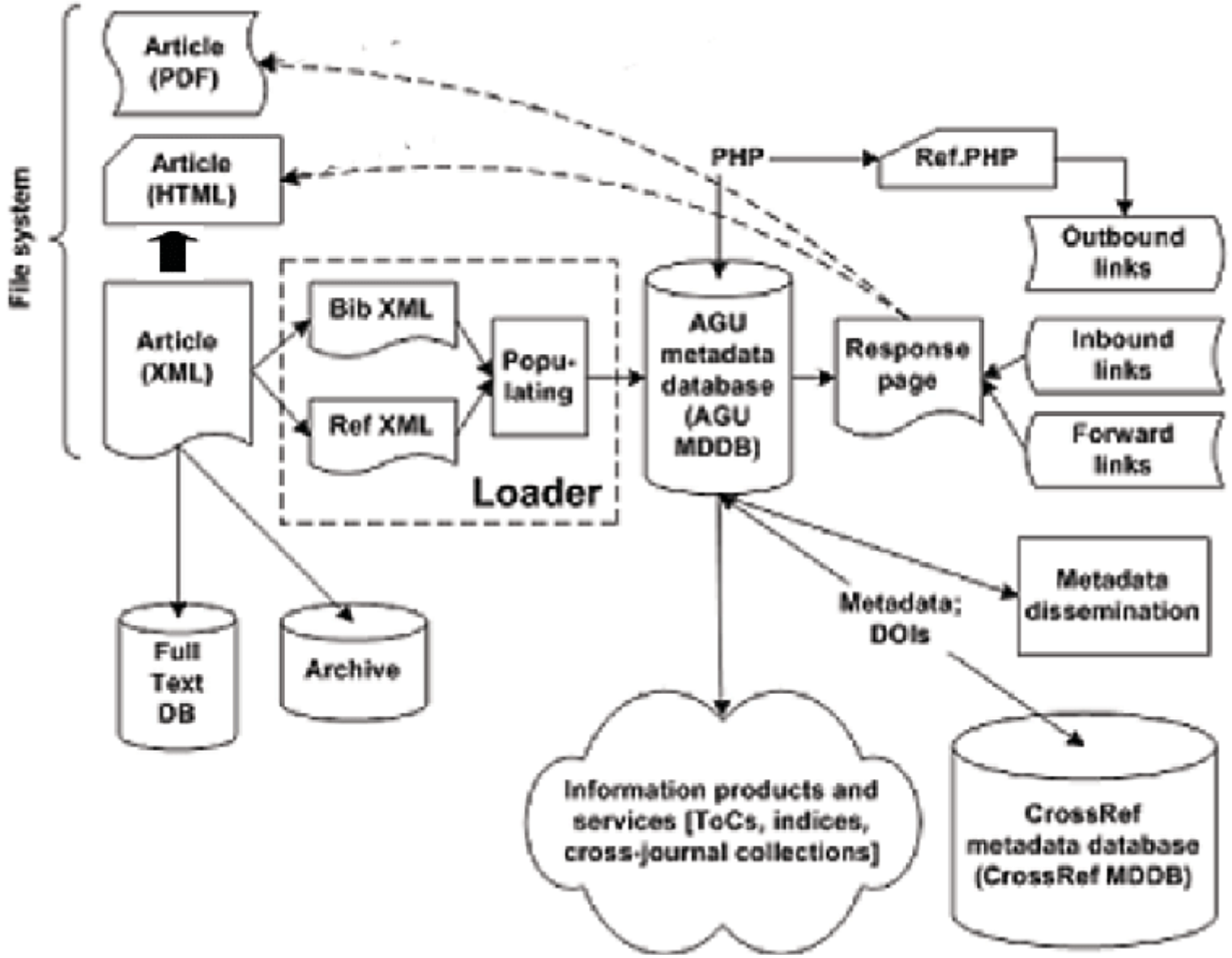
AGU Case Study

- How American Geophysical Union redesigned its publishing processes and technology
  - Substantially increased productivity in producing existing publications
  - Enabled many new kinds of publications
- Valid XML single source is the foundation for all the automated publishing and post-publication processes

AGU Authoring – Before and After



AGU Publishing System



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

31 October 2013
Lecture 19.2 – Enterprise Data Management

The Data that Enterprises Manage

- In addition to the "end-to-end" processes of authoring, management, and delivery for content many enterprises have end-to-end data or transactional processes
- Some of these data processes are separable from the content processes and others are intertwined (especially in e-commerce processes)
- These internal processes also extend to other enterprises in supply and demand chains or distribution channels

Enterprise Data Management Challenges

- Many business processes span multiple departments (or companies) or multiple business applications (run by separate departments)
- These "silos" or "stovepipes" are narrowly-focused, may have been created over time, and not have been designed to share information with each other
- Stovepipe applications naturally occur when engineering, manufacturing, sales, marketing, etc. are separated into different departments with limited interactions except when they hand off "finished work" to each other

Enterprise Data Management Challenges

- Each of these systems has a specific purpose and a data model customized for that purpose - so these models may be incomplete or incompatible with respect to each other
- In the worst case. manual re-entry of information between stovepipe systems is required
- Many of these problems are also occur between enterprises; the primary difference is whether they can be attacked unilaterally

Sarbanes-Oxley

- The Sarbanes-Oxley Act of 2002 was enacted to curb corrupt business activities and fraudulent accounting practices
- SOX (aka Sarbox) requires firms to implement adequate internal control structures and procedures and attest to their effectiveness.
- SOX requires sufficient auditing and traceability to relate the IT systems that carry out internal controls and the financial reporting process to the firm's financial statements

Ken Lay Does the “Perp Walk”

Sarbanes-Oxley

- SOX also requires that firms disclose "material" information about their operations and financial situation in a timely and predictable manner that trigger disclosure

XBRL

- Standardization of "Extensible Business Reporting Language (http://www.xbrl.org/WhatIsXBRL/) and standard models for the auditing document types and their interrelationships
- Excellent XBRL resource: http://xbrl.squarespace.com/
- Some people argue that effective internal controls should be viewed as a strategic investment, not just a defensive move

"Electronic Health Records: Just Around the Corner? Or over the Cliff?"

- A case study of the adoption of an electronic health record system by a small (4 physicians) medical office... discusses several interconnected organizing systems
- What are the resources being organized?
- What were the primary motivations for installing the system?

"Electronic Health Records: Just Around the Corner? Or over the Cliff?"

- Was the system they selected able to implement the organizing system(s) effectively?
- Were their expectations about installation, training, and operation reasonable?
- Of the problems they encountered, which were preventable, and which ones weren't?

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

31 October 2013
Lecture 19.3 – Integration and Interoperability

Integration

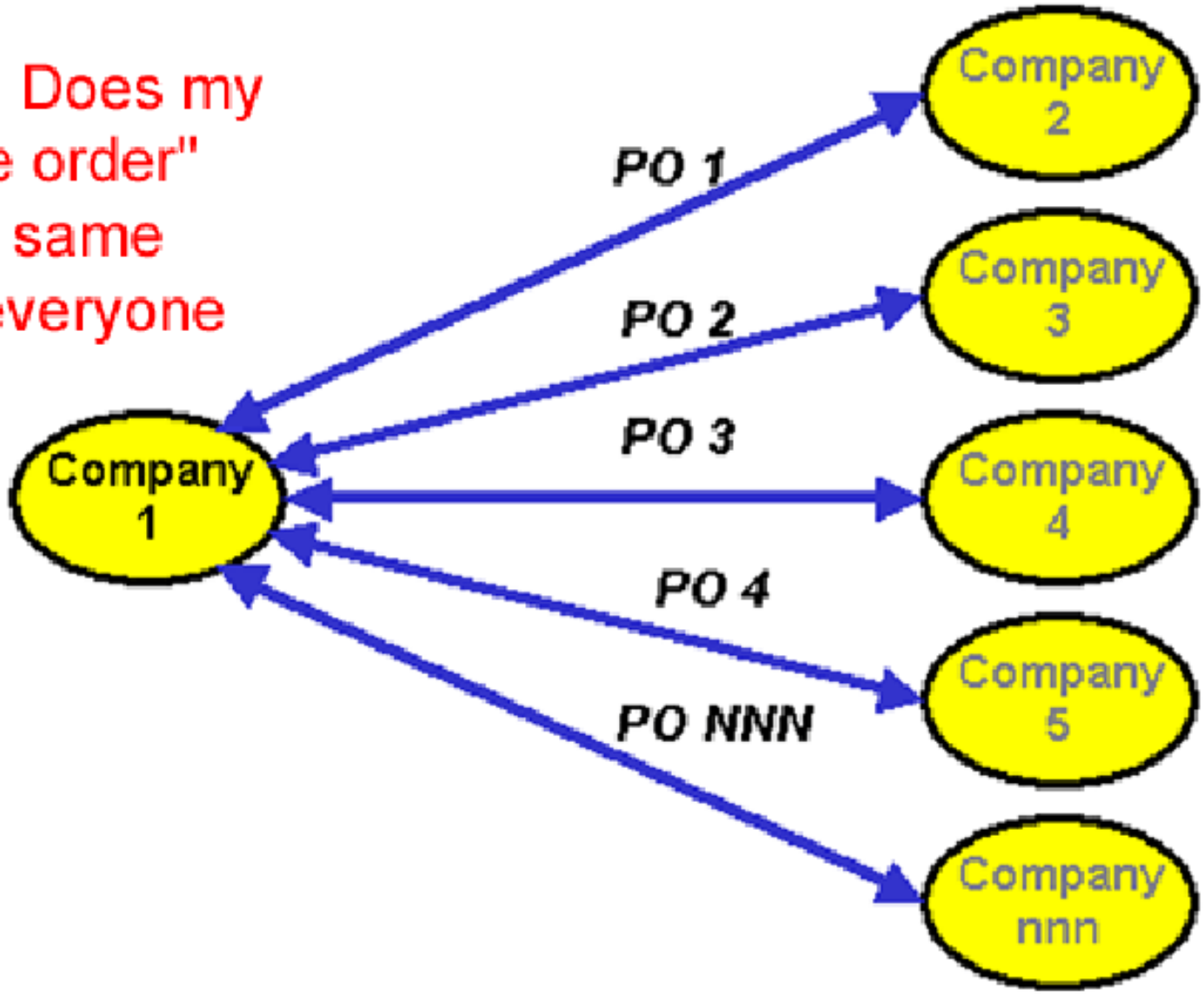
- Integration is the "controlled sharing of information" between two (or more) business systems, applications, or services within or between firms
- Integration means that one application can extract or obtain information from another one
- It doesn't mean that the information will work "as is" in the target application
- "Enterprise integration" - making different applications share information - has long been a substantial portion of the IT activities in many companies

Inter-Enterprise Data Integration

- Data integration between companies so they can do business with each other is also a huge IT challenge
- Suppose you publish your web service interface description; this says "my ordering service requires a purchase order that conforms to this schema"
  - This says "send me MY purchase order" not "send me YOUR purchase order"
- Will the purchase orders being used by other firms meet your interface requirement, either directly or after being transformed?

To Interoperate, or not to Interoperate?

Problem: Does my "purchase order" mean the same thing as everyone else's?



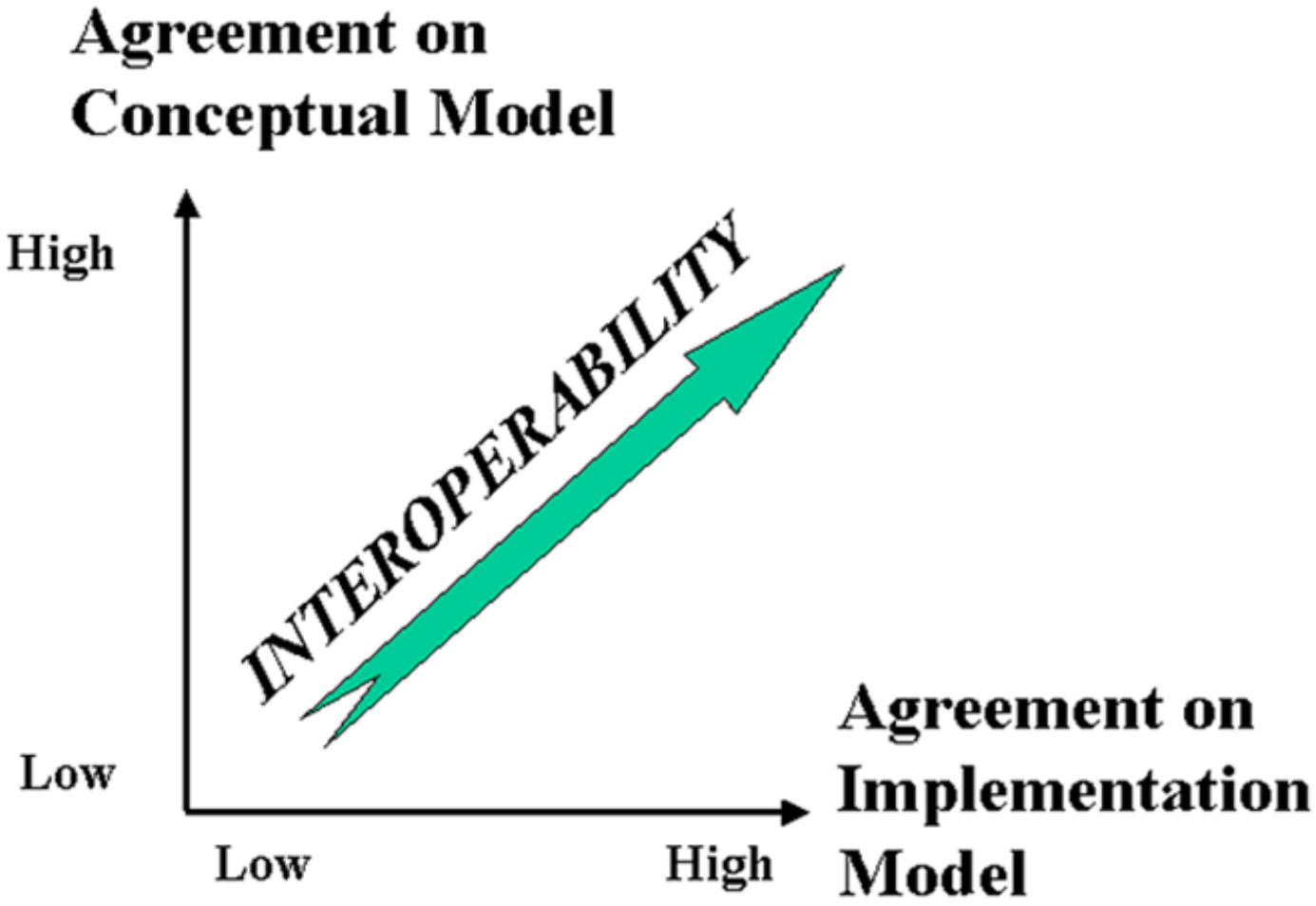
Syntactic, Structural, & Semantic Interoperability

- SYNTACTIC interoperability is just the ability to exchange information
- STRUCTURAL interoperability means that all of the expected information components are present with the same arrangement and granularity
- SEMANTIC interoperability requires that the content of the message be understood by the recipient application or process
- Semantic integration is the process by which this common semantic model is created

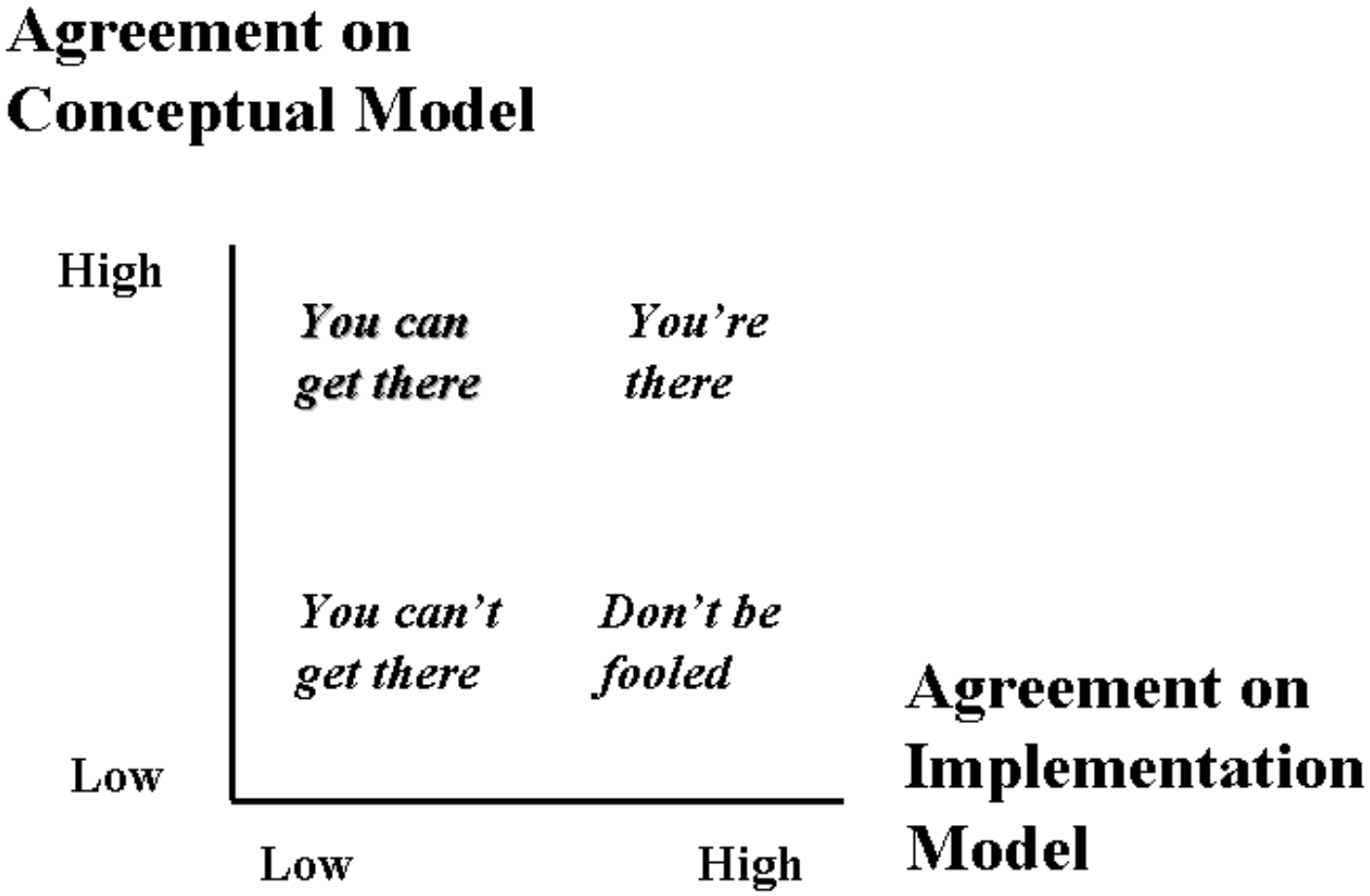
Interoperability isn't All or None

- Some interoperability problems can be detected and resolved by completely automated means
- Other problems can be detected and resolved with some human intervention
- Other problems can be detected but not resolved
- Some problems can go undetected

The Dimensions of Interoperability



The Dimensions of Interoperability



Ways NOT to Interoperate

- Elements with the same meaning can have different names ("IssueDate" vs "OrderIssueDate")
- Elements with the same meaning can have different names or different formats even when they have the same name (September 11, 2001, 9/11/2001 and 9-11-01; 11/09/01 in Europe)
- Same meaning, but encoded differently: in XML using attributes instead of elements

Ways NOT to Interoperate

- Differences in granularity ("Name" vs "LastName" "FirstName", "Address" vs "Street" "City" "PostalCode“)
- Overlapping coverage, different assemblies (Separate "Order" and "Customer" documents)
- Implied values embedded in forms or software

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

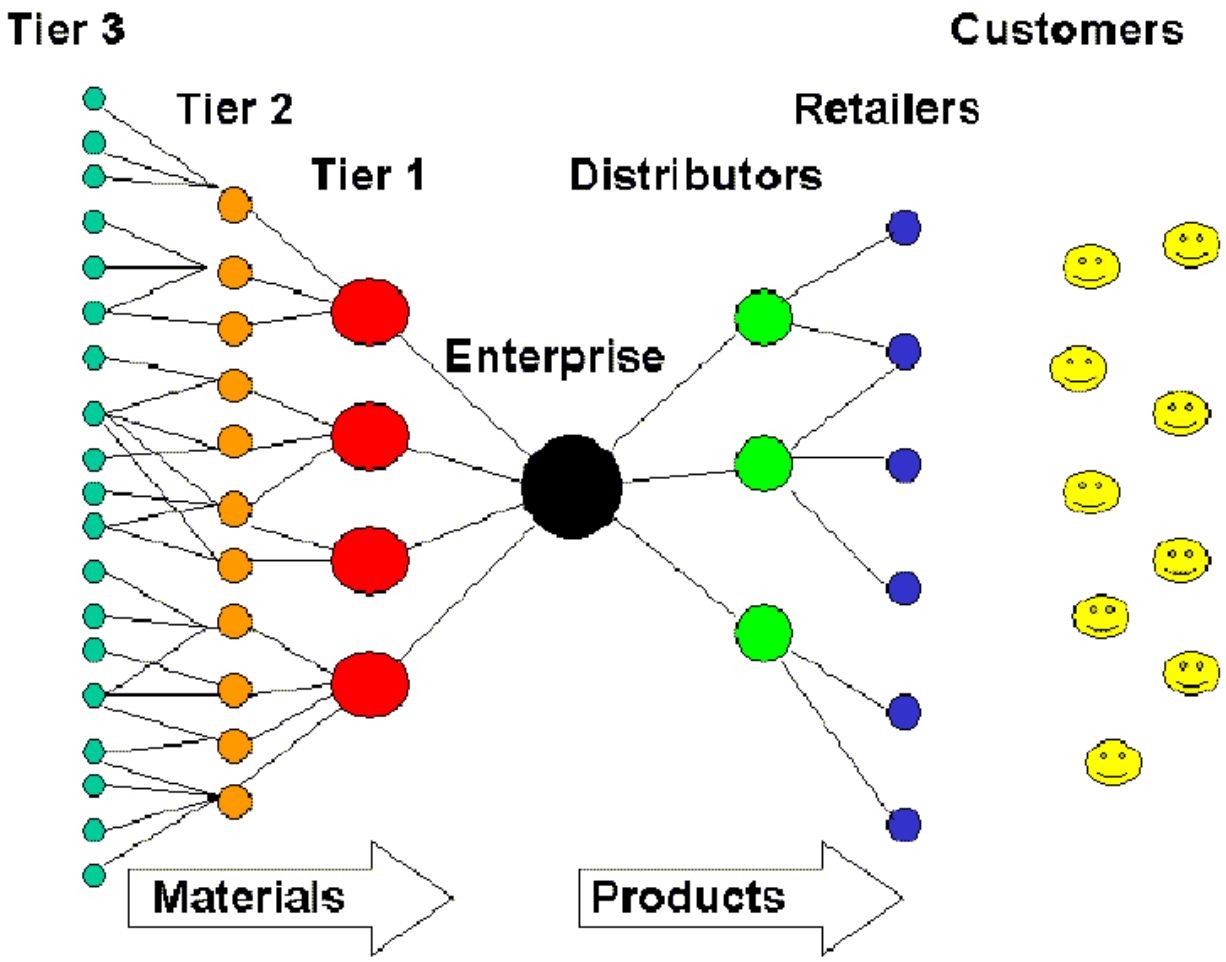
Robert J. Glushko
glushko@berkeley.edu
@rjglushko

31 October 2013
Lecture 19.4 – Supply Chains and Inter-Enterprise Information Exchange

The Supply Chain Pattern

- A supply chain is an organizing system that defines the end-to-end view of the buy-side and sell-side relationships of an enterprise
- A supply chain is the network of facilities and distribution capabilities an enterprise uses to:
  - "Source" (or "procure") raw materials (chemicals, ores, grains, ...) or components
  - Transform the materials or assemble the components into products
  - Deliver the products to customers (indirectly through distributors or stores or directly to the purchaser)

Conceptual Model of a Supply Chain



The Information Supply Chain

- The flow of materials and goods in a supply chain is accompanied by information about it
- But information about supply chain activities and processes is increasingly separated from the physical flow of materials and goods, and for information-based services there are no physical resources to move
  - (recall “Global Disaggregation of Information-Intensive Services” in previous lecture)

The Information Supply Chain

- Information also flows in the opposite direction from the customer, retailers, and distributors back into the supply chain – this is also called the DEMAND CHAIN
- The information supply chain has become especially important because increased global competition and better informed customers are forcing forms to shift from forecast to demand (i.e. customer) driven business models

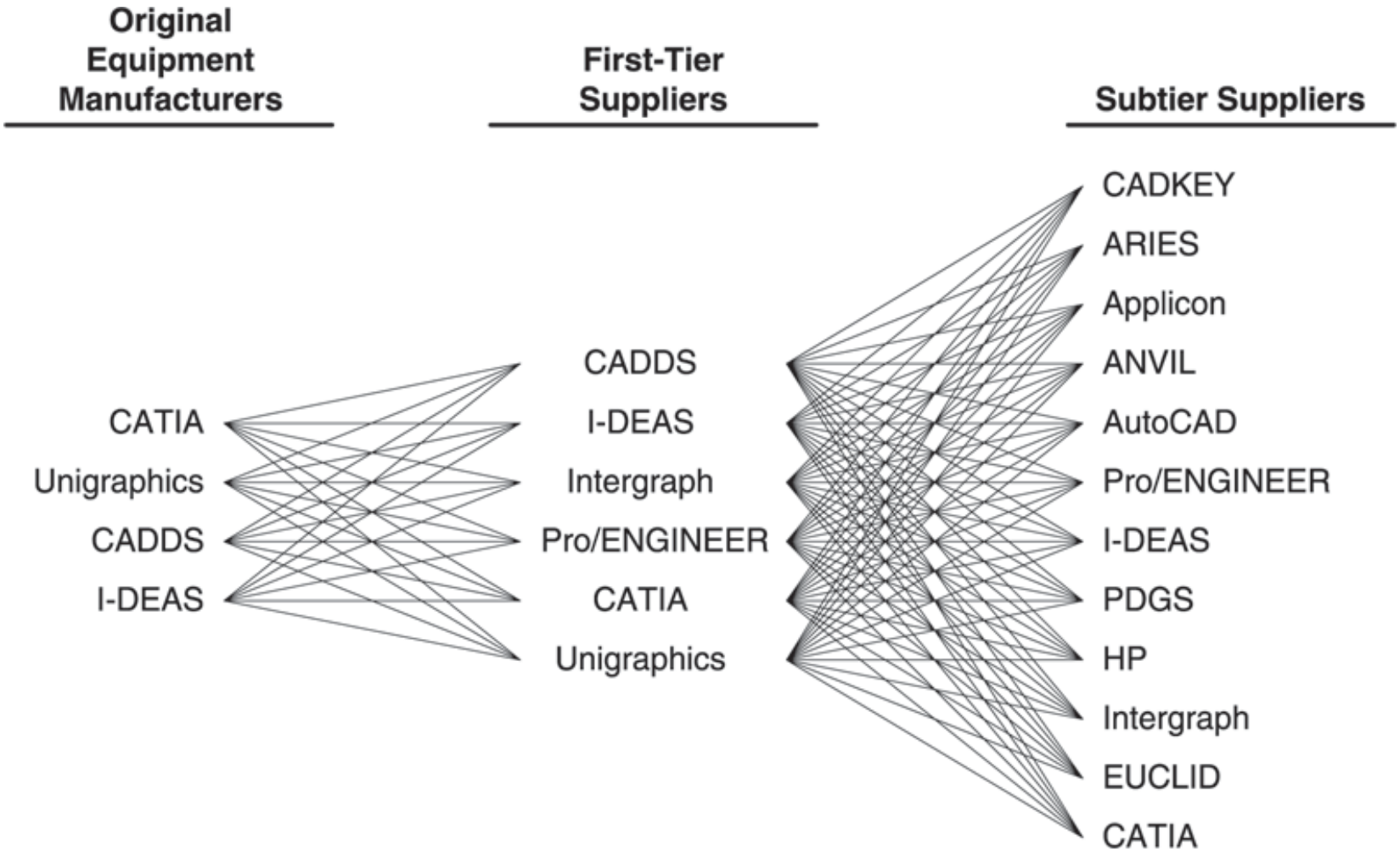
Design Issues for the Information Supply Chain

- What information is exchanged?
- Which entities in the supply chain are able to exchange information?
- What is the frequency of this information exchange?

"Interoperability Costs in the US Auto Supply Chain"

- Excellent case study about how a concurrent engineering business model escalates the information exchanges and interoperability problems in the organizing system centered around automobile manufacturers
- Analyzes various alternatives for data transfer, and finds that the choices made are not the optimal ones

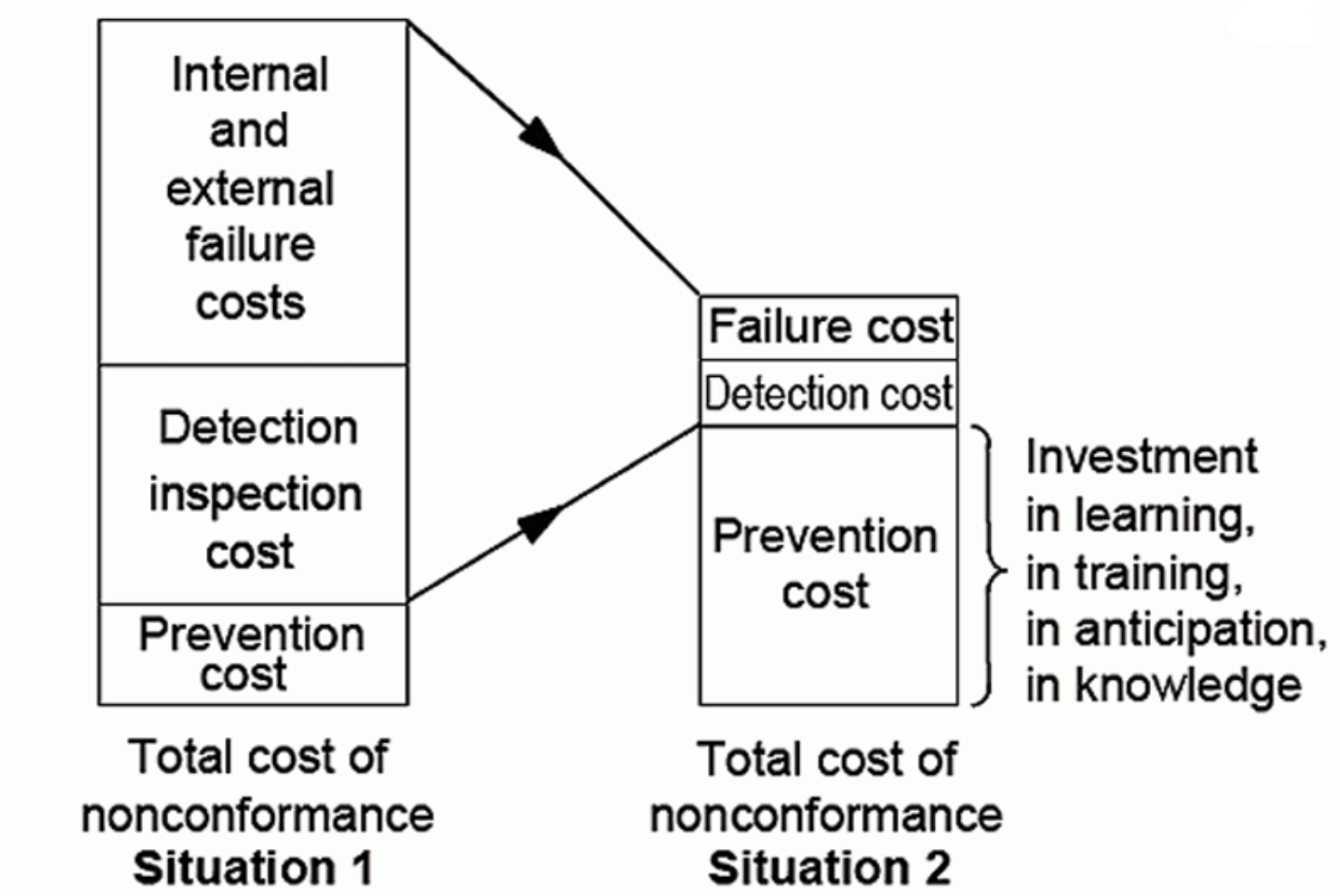
CAD/CAM Systems Proliferation



Juran's "Quality Costs" Framework

- Joseph Juran's "Quality Control Handbook" (1951) -- "cost of quality" framework determines how much to spend on quality at any point in the "quality system"
- The costs of preventing and finding quality problems (avoidance) ...
- ... must be balanced against the costs associated with those quality problems (mitigation)

The Case for Investing in Avoidance



Interoperability Avoidance Costs

Table I Sources of interoperability costs

Cost category	Source of cost	Components
Avoidance costs	Multiple CAD/CAM systems	CAD/CAM software licenses
		System maintenance
		System training
	Multiple translators	Translation software licenses
		Software training
	Outsourcing data translation	Third-party suppliers
	Investments in interoperability solutions	In-house interoperability research
		Activities in industry consortia

Interoperability Mitigation and Delay Costs

Table I Sources of interoperability costs

Cost category	Source of cost	Components
Mitigating costs	Poor quality CAD/CAM files	Scrapped models, designs, prototypes, parts, dies, etc. Manual data reentry
Delay costs	Delays	Car sales forfeited Delayed profits Delayed consumer benefits

Estimated Interoperability Costs

Table II Summary of annual interoperability costs: cost component approach

Source of cost	Costs by industry segment				
	OEMs	Suppliers	Tooling	Total	Percent of total
Avoidance costs	2,302	35,656	14,841	52,799	5
Mitigating costs	247,773	204,094	455,778	907,645	86
Subtotal	250,075	238,750	470,619	960,444	91
Percent segment revenue[a] (%)	0.075	0.083	11.914	0.513	
Delay costs				90,000	9
Total costs				1,050,444	100

Notes:
All figures are in thousands of US dollars unless otherwise stated
[a] See Brunnermeier and Martin (1999) for details of revenue estimates for the OEM (pp. 2-15), supplier (pp.2-18) and tooling segments (pp.2-20)
[b] We could not determine the distribution of costs for this category

Next Week

- Tuesday November 5 is midterm review
  - View video recording before Tuesday Nov 12
- Submit review questions to glushko@berkeley.edu by noon Monday Nov 4
- Thursday November 7 is midterm
  - 4 multi-part question; 1 is required, choose 3 of other 6
  - open book, open notes
  - Write on your computer, submit online and in print