# 12. Enterprise / Institutional Categorization and Standards

**INFO 202 - 8 October 2008**

**Bob Glushko**

## Plan for INFO Lecture #12

Introduction to standards and standards-making

Content standards; UBL case study

Format standards: Auto industry supply chain case study

# What is a Standard?

"A prescribed set of rules, conditions, or requirements concerning definitions of terms; classification of components; specification of materials, performance, or operations; delineation of procedures; or measurement of quantity and quality in describing materials, products, systems, services, or practices" (NIST)

"A freely implementable specification developed by consensus among the important stakeholders in some domain, working in a framework that encourages open participation provided by an organization chartered to create standards" Glushko CNET Interview

Standards define institutional categories:

- (legal / correct / compatible / safe / etc) VS (illegal / incorrect / incompatible / unsafe / etc.)

# What Is a Standard? [2]

Many standards define technical specifications and procedures to ensure a "common design" or "interoperability" for some product or process

"Common design" or "interoperability" aren't always clearly defined, but what they mean in practice is that two different companies or vendors could each implement the specifications or follow the procedures and produce equivalent results

Standards are especially important in industries or markets that have significant "network effects"

## Standards and Metcalfe's Law

The value of a standard often depends on how many people (or computers) conform to it

How do you encourage and enable others to conform to a standard?

- Standardization Approach 1: "Follow MY specifications or I won't do business with you"
- Standardization Approach 2: "Excuse me, here's my specifications, would you like to do business with me?"

## Who Sets Standards?

Some technical standards are set by governments in order to protect the public interest

Other standards are set by organizations created precisely for the purpose of setting standards

Other standards are set by groups of companies who may or may not establish an ongoing organization and governance procedures to maintain and extend a standards "family" around the initial standard

Other "standards" begin as proprietary vendor specifications that will become "de facto" standards if they are widely adopted and achieve market dominance

Often "de facto" standards are submitted to a standards organization with the goal of "rubber stamping" them as "de jure" ones

## Comparing "Standards" Organizations

"Standards" organizations differ in how they "lump and split" the technical problems and solutions in a domain (and whether their focus is "horizontal" or "vertical")

They typically portray themselves at the center of the universe

They have different membership rules, governance procedures, and standards-development methods

They have allegiances, interlocking standards or rivalries with other standards organizations

A small number of large companies participates in almost every e-business and information systems standard organization

## Using Standards in Your Information Organization Efforts

You may be required by the government or by your customers to conform to some standards for information organization

But to the extent that standards embody best practices, it is a good idea to follow them anyway

Nevertheless, this "motherhood advice" is made challenging to follow by the proliferation of standards, their sometimes competitive and sometimes complementary character, the variety of processes by which they are created, and their different likelihood of adoption (or survival)

Pay attention to the role of standards in the business models of the companies that advocate them

# MITRE Reference Model for Comparing Standards: Data Objects

What is the semantic granularity of the concepts being standardized?

Is there also a standard for how the concepts are encoding in some syntactic or physical representation?

Does the standard also specify "instance sets" or possible values for each data element concept?

# MITRE Reference Model for Comparing Standards: Structures

In addition to standards for data objects, are there standards for schemas or "document architectures" that structurally organize them?

Are there standards for the instances or interchange formats used by publishers/producers or expected by subscribers/consumers?

## MITRE Reference Model for Comparing Standards: Community Characteristics

Is there a primary stakeholder with decision making authority, or is authority distributed?

What are participants' obligations to support the standard?

Do the participants already share an understand of the domain to be standardized?

## Lessons From Standards Making: The "Person-Concept" Tradeoff

Semantic agreement comes at a cost driven by the number of people who require a shared understanding, and by the number of concepts they must all understand

So a small set of people can agree on a complex standard, or a large set of people can agree on a simple one

Especially when participants are just trying to agree on how to describe pre-existing shared concepts rather than having to define them first

## Lessons from Standards Making: Enterprise Data Standards

Standards-making can be successful when a "single authority exercises effective control over the system requirements, funding, the developers, and the users"

But a very large enterprise cannot hope to construct a single data model (or even a single-set of universally understood concept definitions) for all the data it requires

## Lessons from Standards Making: Incentives and Disincentives

Approaches that require perfect coordination and altruism are of no practical interest

Disincentives to agree on semantics arise if agreement means that someone has to change an implementation and pay the cost of doing so

Model-driven development tools that generate needed software artifacts (e.g., system and user interfaces) from specifications can encourage standards adoption

## Lessons from Standards Making: "Communities of Interest"

Standards making is best organized around naturally formed communities of interest rather than "org chart" organizations

Different types of communities might be needed to develop, deploy, and maintain a standard

## Vertical and Horizontal Content Vocabularies

Vertical:

- Particular industry or vertical market
- Detailed product semantics
- Specialized process semantics
- Sometimes called "domain-specific" languages

Horizontal

- Concepts that are common to all (or a large number of) vocabularies

# Vertical Vocabularies and EDI

Scores of economically important vertical vocabularies were developed in the 1970s and 1980s using the Electronic Data Interchange syntax and standards processes

These are still in wide use in supply chains in consumer packaged goods (e.g., WalMart's turf) and other industries

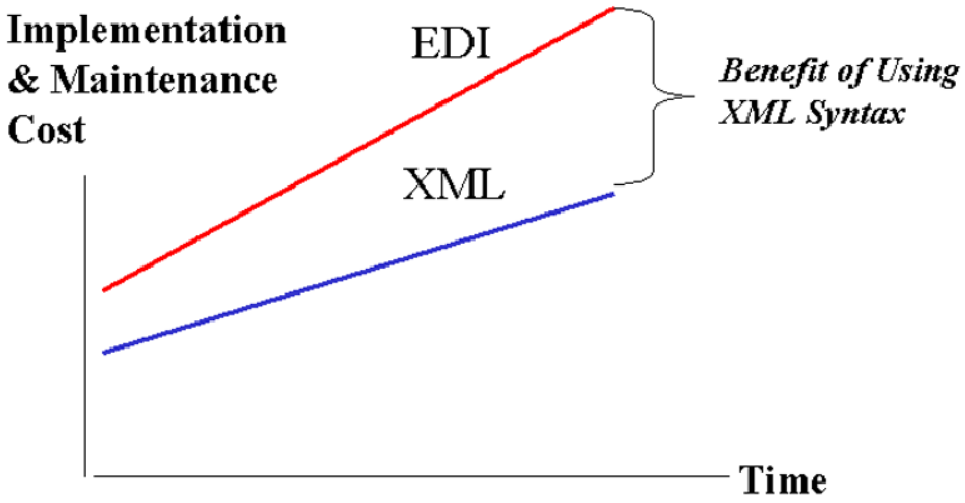
The semantic understanding encoded in EDI message standards is a large investment not easily discarded

But XML, web services, and other new concepts and technologies are generally used for new applications of electronic documents

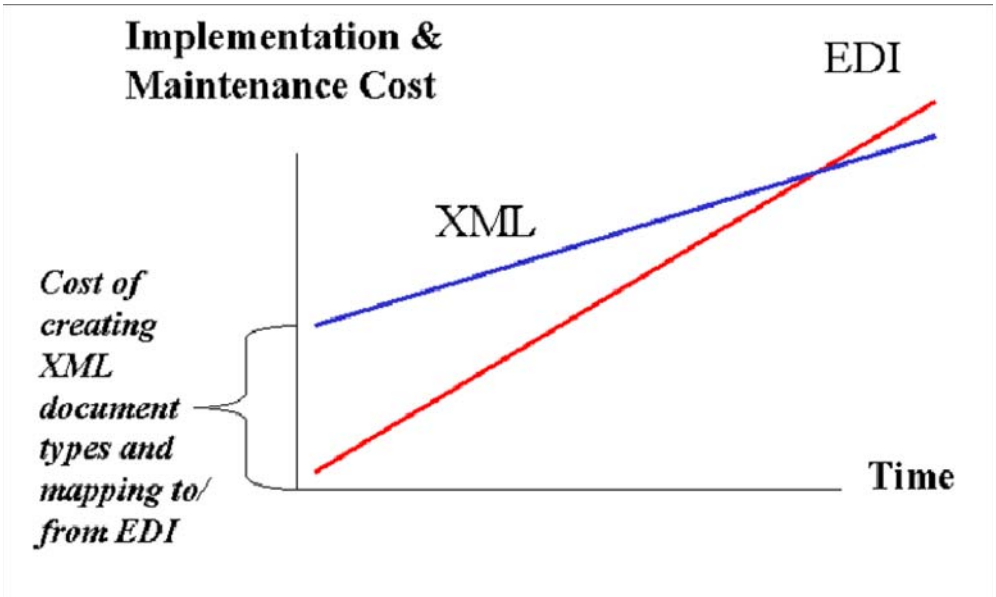
# EDIFACT Message Fragment

```
20000305:102'DTM+158:20000305:102'DTM+159:20000722:102'NAD+SU+9876543
NY'NAD+MI+88835::92'GIS+37'NAD+ST+72681::92'LIN+++93235494:IN'PIA+1+0
04'RFF+ON:XXX00004'QTY+79:6660:EA'DTM+51:19991225:102'DTM+52:20000304
91225:102'DTM+11:20000302:102'SCC+1++W:16'QTY+1:960:EA'DTM+158:200003
20000313:102'SCC+4++W:16'QTY+1:900:EA'DTM+158:20000320:102'QTY+1:900:
:1080:EA'DTM+158:20000403:102'QTY+1:1080:EA'DTM+158:20000410:102'QTY+
'QTY+1:630:EA'DTM+158:20000424:102'QTY+1:990:EA'DTM+158:20000501:102'
:102'QTY+1:810:EA'DTM+158:20000515:102'QTY+1:810:EA'DTM+158:20000522:
0529:102'QTY+1:810:EA'DTM+158:20000605:102'QTY+1:630:EA'DTM+158:20000
20000619:102'QTY+1:810:EA'DTM+158:20000626:102'QTY+1:810:EA'DTM+158:2
158:20000710:102'QTY+1:766:EA'DTM+158:20000717:102'SCC+2'QTY+3:12610:
:20000416:102'SCC+3'QTY+3:17485:EA'DTM+51:19991225:102'DTM+52:2000052
0:EA'UNT+73+770001'UNZ+1+77'UNB+UNOA:2+BFT:ZZ+CAI:ZZ+000305:2338+78++
```

## Why You'd Encode A New Vocabulary in XML and Not EDI




## Why EDI Hasn't Gone Away



# Why Semantic Interoperability Problems Are Often Inevitable with Vertical Standards

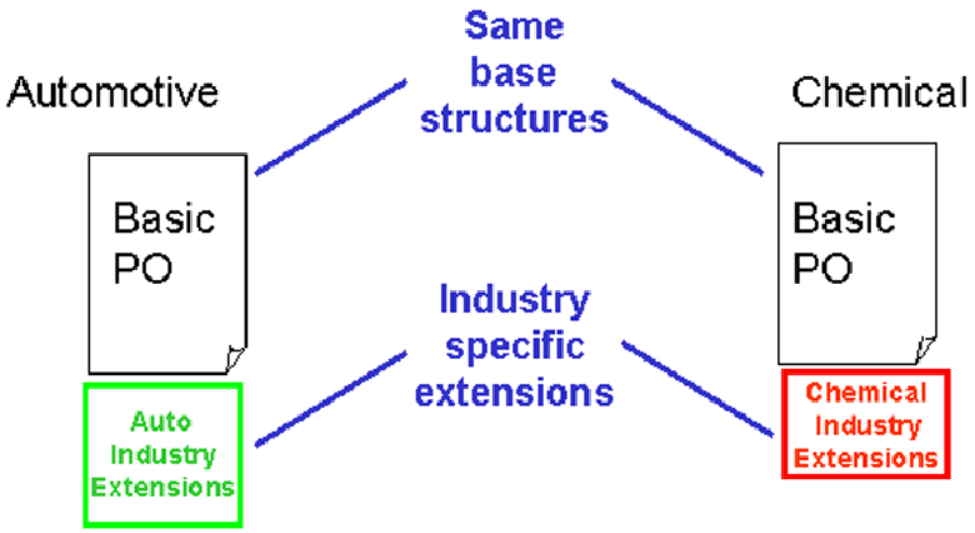
Each new vocabulary for a particular industry is a step forward for that community, but proliferates definitions of information models that are common to many of them

Since the distinctive or specialized parts of each vocabulary are the industry-specific "vertical" parts, a lot of attention gets paid to them

In contrast, relatively less effort is given to the "horizontal" parts that seem more familiar or understandable

Nevertheless, any large company – even highly verticalized ones – engages in diverse business activities that require it to understand multiple vocabularies at different times

# Vertical & Horizontal Vocabularies Coexist



# The Quest for "Core Components"

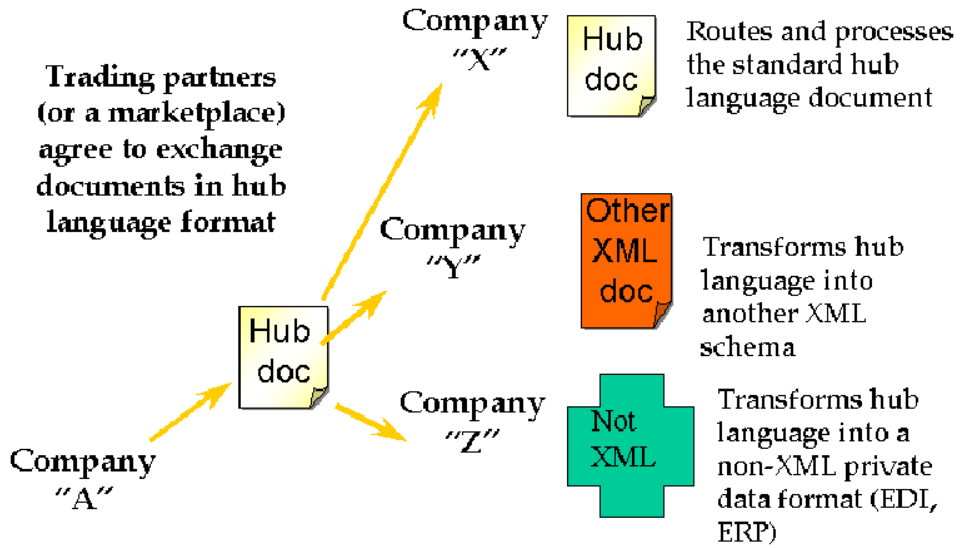
Horizontal vocabularies would be extremely useful as building blocks for vertical ones because their interoperable semantics would be the "glue" in "hub languages"

But these "core components" are hard to define because they never occur by themselves; they are always "wrapped" in some vertical context

Furthermore, because vertical vocabularies that need to interoperate exist in different syntaxes (XML, X12 EDI, EDIFACT), many people argue that core components must be maintained in a conceptual format

This philosophy imposes additional requirements to standardize "how context is applied" and the "naming and design rules" for encoding in the target syntax

# An Interchange or Hub Language



# The Need for a "Universal Business Language"

Too many redundant and incompatible document models of common business documents like catalogs, orders and invoices

Too many redundant adapters and gateways to enable trading relationships across domain boundaries

# UBL

31 document types built with a common architecture/metamodel needed for supply chains (European) and International Trade (Asia and US)

These document types use a library of XML schemas for reusable aggregate data components such as "Address," "Item," and "Payment"

The library is a standard implementation of the ebXML/CEFACT "core components" types ("Amount," "Code," DateTime," etc.

An OASIS standard

# The Road to UBL

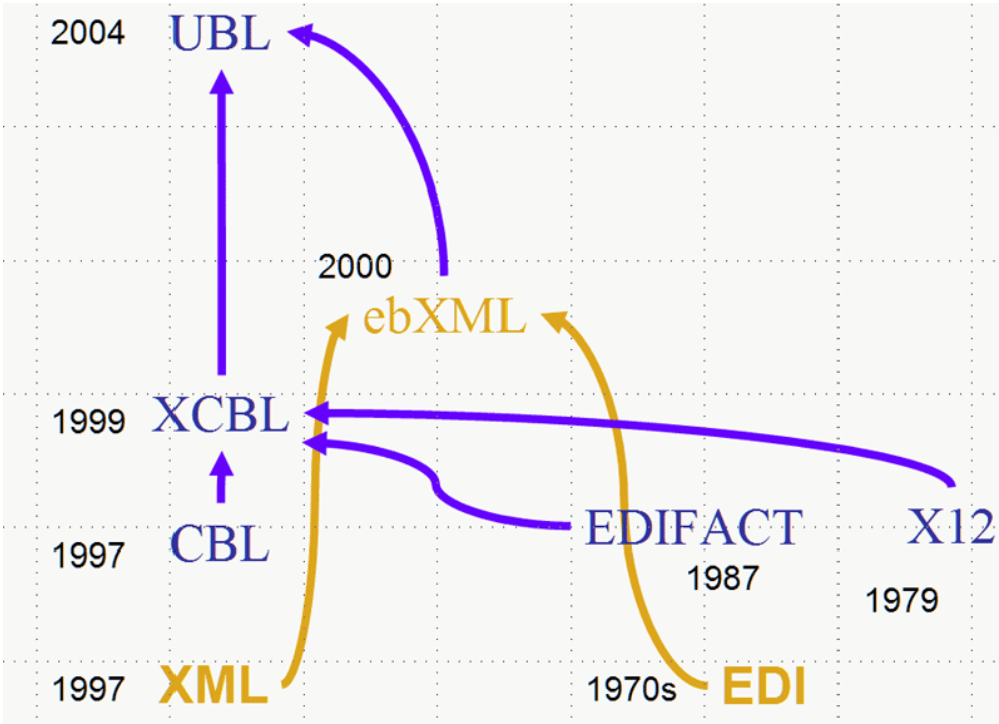
(early 1990s) - Ad hoc efforts in EDIFACT to "harmonize" core components across verticals

1997- XML Common Business Library
is 1st XML horizontal vocabulary, incorporated EDIFACT semantics and code lists

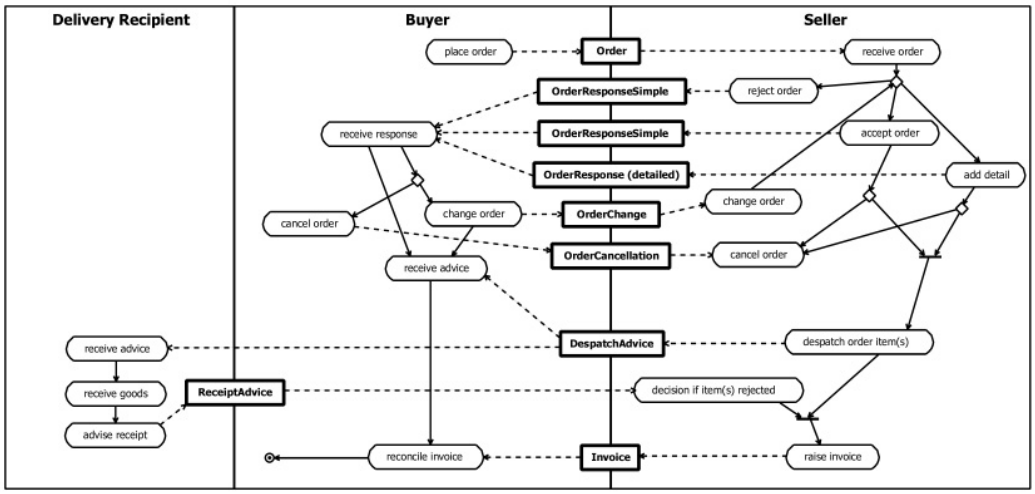
1999 - ebxml
initiative of EDIFACT and OASIS to develop syntax-neutral "core components"

2001 - Universal Business Language
effort begins, building on xCBL and ebXML Core Components

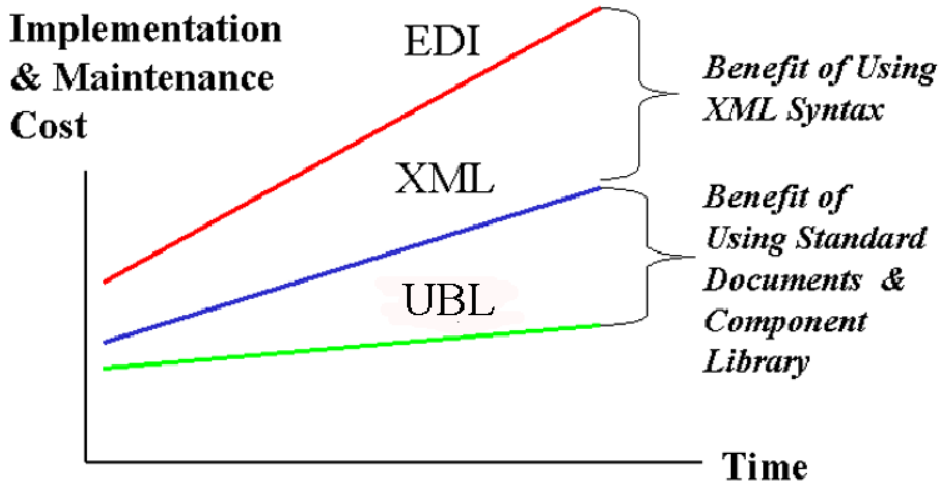
# UBL's Complex Ancestry



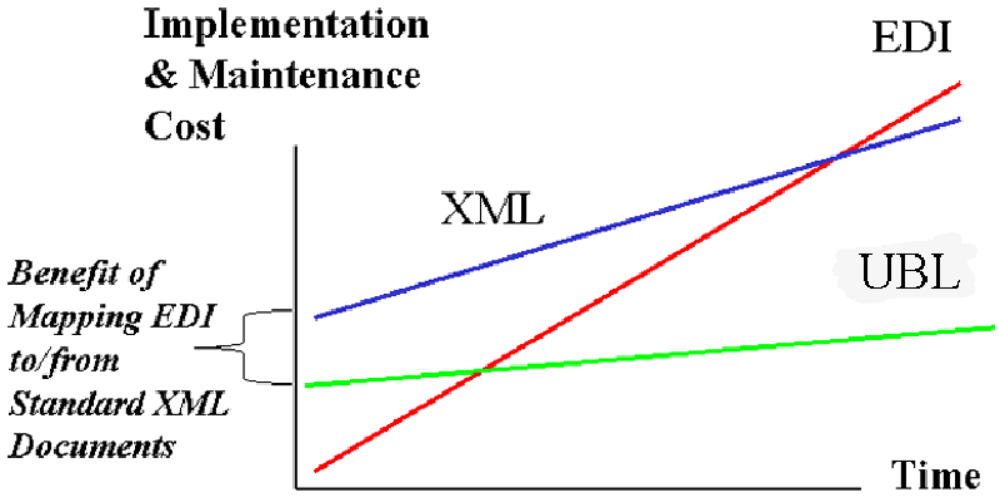
# UBL 1.0 Document / Process Scope




# XML Hub Language vs. EDI



# XML Hub Language Payoff




# Mapping in and out of Hub Language

If all parties/applications/services rely on a hub language for their external interfaces, an exponential interoperability challenge becomes a linear one

*Mapping*
tools for transforming instances from an internal information model to another one are ubiquitous as standalone tools and as parts of application servers

EXAMPLE: Altova MapForce

## *"Adoption of UBL in Denmark"*

What is the overall goal that motivates this project?

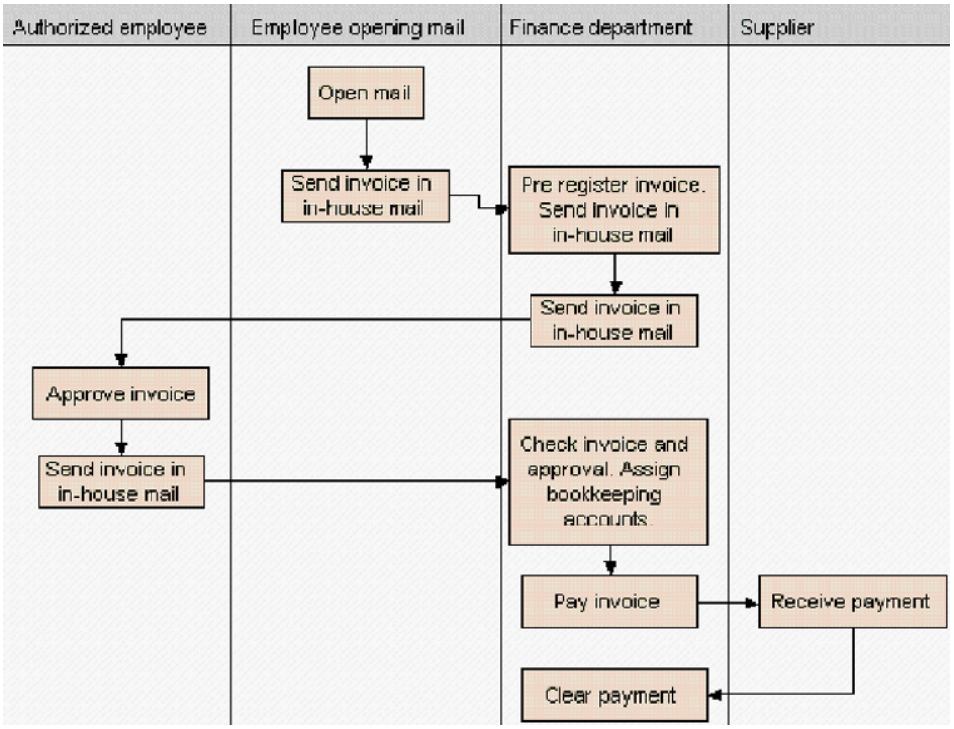
What are the key technical pre-requisites for this project?

How much does it cost to process an invoice?

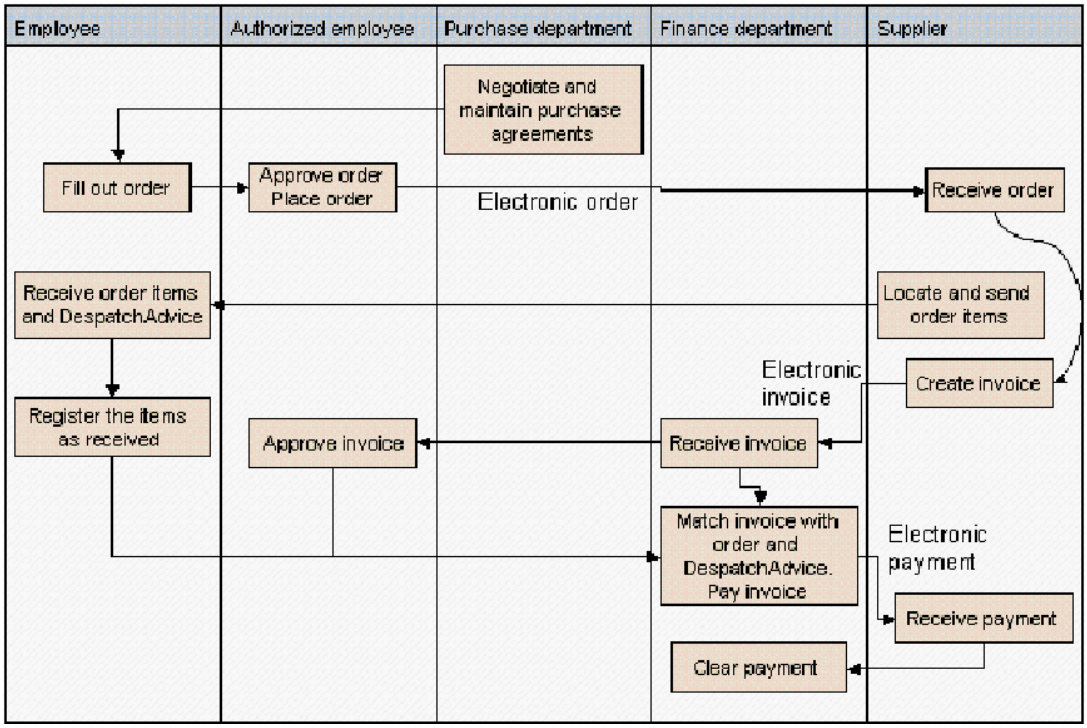
What will happen next with this project?

Would this kind of project be possible in the US?

## Manual Invoice Process



# Optimal Req-to-Pay Process




# 3 Ways to Send an Electronic Invoice

Send directly (e.g., from ERP system) to the government (using EDI value-added network as transport)

Enter invoice information into Web form

Send paper invoice to scanning agency that will create the UBL invoice and forward it

## *"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 "ecosystem"

Analyzes various alternatives for data transfer, and finds that the choices made are not the optimal ones

Concepts and lessons apply to other industries with "data exchange-intensive" supply chains

## Alternatives for Data Transfer Between Two Systems

Manual re-entry

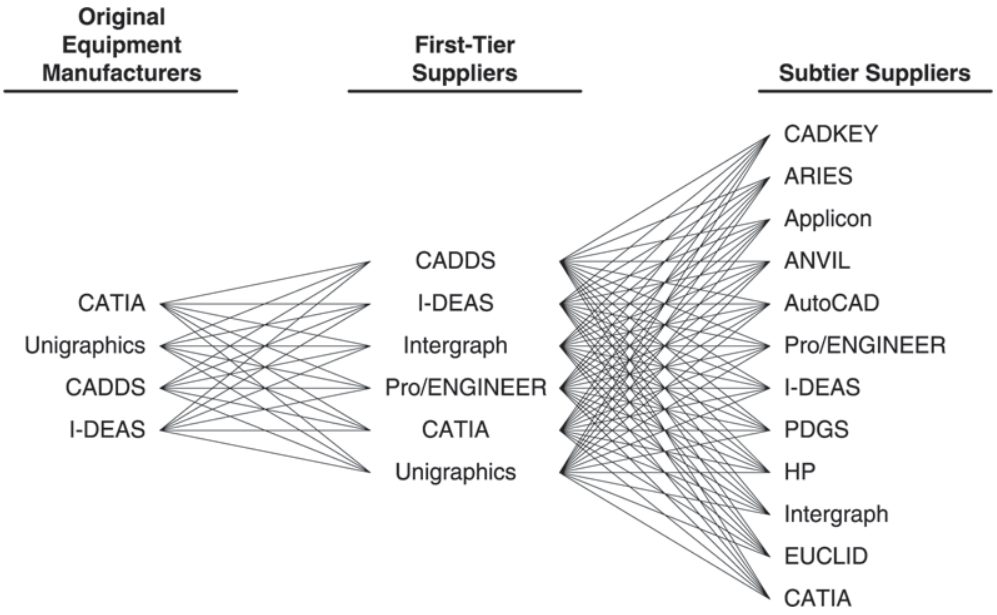
Everyone has to learn to "speak" all the languages

- Native formal transfer
- Point-to-point translation

Everyone has to learn just one new language but it has to be the same one

- Dominant players impose their language on their ecosystem
- Multiple vocabularies exist, but there is at least one "interchange" or "hub" language designed to facilitate translations between "native" vocabularies

# 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"

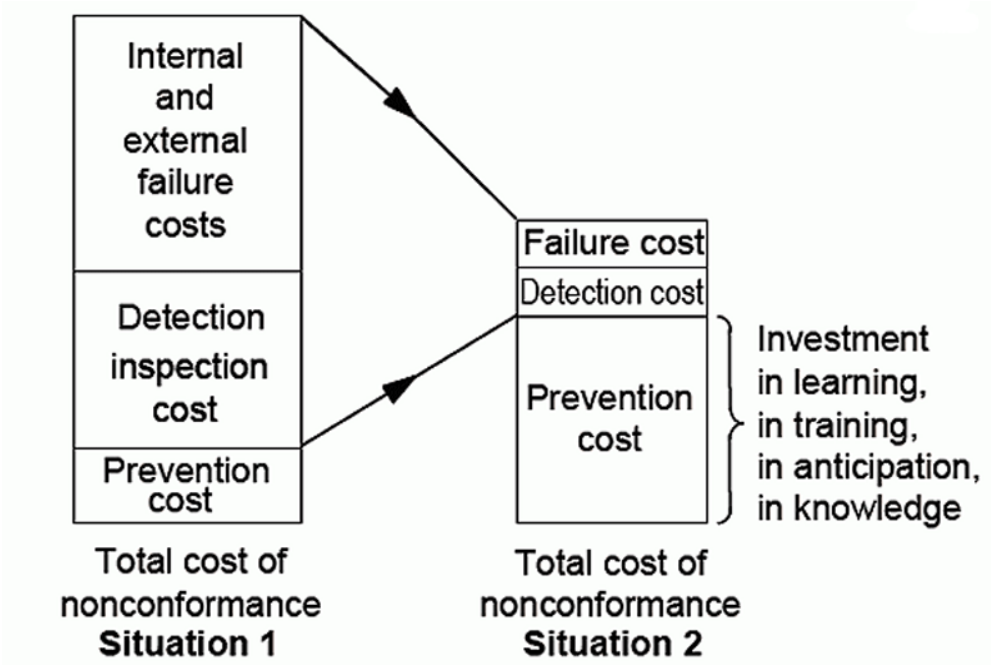
The costs of preventing and finding quality problems (avoidance) ...

- Prevention costs (design reviews, training, guidelines, knowledge...)
- Appraisal costs (tests, process control measurements, reports, evaluations,...)

... must be balanced against the costs associated with those quality problems (mitigation):

- Internal failure costs (costs incurred before the product or service is delivered: scrap, rework, lost time, unused capacity, ...)
- External failure costs (cost incurred when quality problems reach customers: returns, recalls, complaints, field services, warranty repairs, liability lawsuits,...)

# 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.<br>Manual data reentry |
| **Delay costs** | Delays | Car sales forfeited<br>Delayed profits<br>Delayed consumer benefits |

# Estimated Interoperability Costs

**Table II** Summary of annual interoperability costs: cost component approach

| | Costs by industry segment | | | | |
|---|---|---|---|---|---|
| **Source of cost** | **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

# Readings for INFO Lecture #13

Catherine Marshall and Frank Shipman, "Which Semantic Web?"

Nigel Shadbolt, Wendy Hall, and Tim Berners-Lee. "The Semantic Web Revisited"

Anupriya Ankolekar, Markus Krotzsch, Thanh Tran, and Denny Vrandecic. "The two cultures: Mashing up Web 2.0 and the semantic web."