Document Engineering

Strategic Computing and Communications Technology

12 March 2007

Bob Glushko

glushko@ischool.berkeley.edu

Plan for Today's Lecture

Document Engineering in the News

The Data vs Document Divide

Documents and Processes -- Yin and Yang

Design Patterns in Document Engineering

Synthesizing Four Modeling Approaches

What Is Document Engineering?

DOCUMENT ENGINEERING is a methodology for specifying, designing, and deploying the information models and repositories that enable document-centric applications

DOCUMENT ENGINEERING is a synthesis of information and systems analysis, business process modeling, electronic publishing, and distributed computing

DOCUMENT ENGINEERING has much in common with the field of Information Architecture, but extends its scope beyond web site and web application design

Intel, Wal-Mart and others Push Electronic Health Records

Tsunami Aid Delayed by Incomplete Shipping & Customs Documents

Borneo Bulletin Online
Your Gateway to Borneo

Tsunami aid stuck in port for months

JAKARTA, Indonesia (AP) - More than 70 shipping containers packed with mosquito nets, school supplies and timber for tsunami survivors have been stuck at an Indonesian port for months because of inadequate documentation, an official said Wednesday.

Salesforce.com Connects its Front End to Back Ends

Global Shippers Give Customers Real-Time Cargo Info

FedEx Kinko's Announces Web-based Printing, Tracking Services

FedEx Kinko's Adds Direct Mail, Print Online to Portfolio

Tuesday January 9, 9:30 am ET

Customers Benefit from Flexibility of New Services and Tool

DALLAS--(BUSINESS WIRE)--FedEx Kinko's, an operating company of FedEx Corp. (NYSE: FDX - News), today announced the introduction of two new offerings -- FedEx Kinko's(SM) Direct Mail Services and Print Online -- further establishing the company as a convenient resource for office, printing and shipping services.

Law Firm Business Models Disrupted by Document Assembly Software

Disrupting conventional law firm business models using document assembly

DARRYL R. MOUNTAIN[1]

Abstract

Document assembly software is a technology that is fundamental to disrupting law firms. This article uses the framework set out by Clayton Christensen in *The Innovator's Dilemma* and subsequent books to examine the range of business models that use document assembly software, from those that are sustaining in relation to law firms to those that are disrup-

The Common Themes in These News Items

Enormous amounts of existing (paper) documents and legacy processes would benefit from automation, process re-engineering, transformation to service-oriented architectures

New business processes are created / coordinated / choreographed via the management and exchange of electronic documents

Standards / patterns for documents and business processes are essential

Information technology and business processes are co-evolving with many ways to create business value

Document Engineering (1)

file:///C:/Documents%20and%20Settings/glushko/My%20Documents/Lectures/Varian-200703...

Is THIS a Typical Document?

CALL me Ishmael. Some years ago—never mind how long precisely—having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world. It is a way I have of driving off the spleen, and regulating the circulation. Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypos get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people's hats off—then, I account it high time to get to sea as soon as I can. This is my substitute for pistol and ball. With a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship. There is nothing surprising in this. If they but knew it, almost all men in their degree, some time or other, cherish very nearly the same feelings towards the ocean with me.

Or is THIS a Typical Document?

Or Maybe THIS is a Typical Document?

Amount   12   If line 10 is larger than line 9, subtract line 9 from line 10. This is
               the **amount you owe**. For details on how to pay, see page 20   ► 12

```
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++
```

Contrasting Methodologies for Documents and Data

Documents and data have had two different disciplines or methods of analysis that have had little intersection

Document-centric analysis

Data-centric analysis

Document Engineering (1)

file:///C:/Documents%20and%20Settings/glushko/My%20Documents/Lectures/Varian-200703...

Document Analysis

Documents are *Artifacts* or *Renditions* that combine content, structure and appearance

The goal of document analysis is a model of a document's content and structure that is separate from its presentational characteristics

This model of the document and those in its equivalence class is called a markup language or document schema or document type

Data-Centric Analysis

Goal is to understand and describe the properties and relationships between information components or objects.

This understanding is represented in conceptual models that organize the components efficiently to support a broad range of contexts or applications.

The conceptual model is also typically called a schema, but this is generally meant to be a "database schema" rather than a "document schema"

The Data vs. Documents Divide

Document Engineering (1)

file:///C:/Documents%20and%20Settings/glushko/My%20Documents/Lectures/Varian-200703...

But A Catalog Is Data (Document)

And a Reference Book is Document (Data)

Document Engineering (1)

file:///C:/Documents%20and%20Settings/glushko/My%20Documents/Lectures/Varian-200703...

So it's a Continuum: The Document Type Spectrum
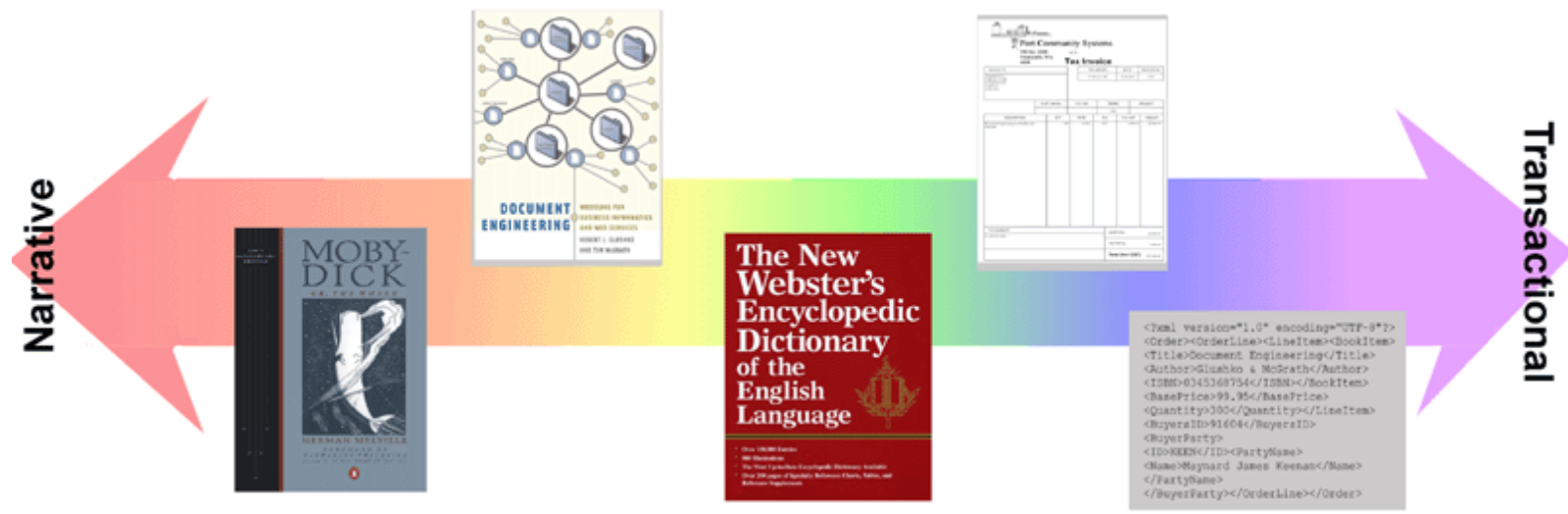
And Business Processes Involve The Entire Spectrum of Documents
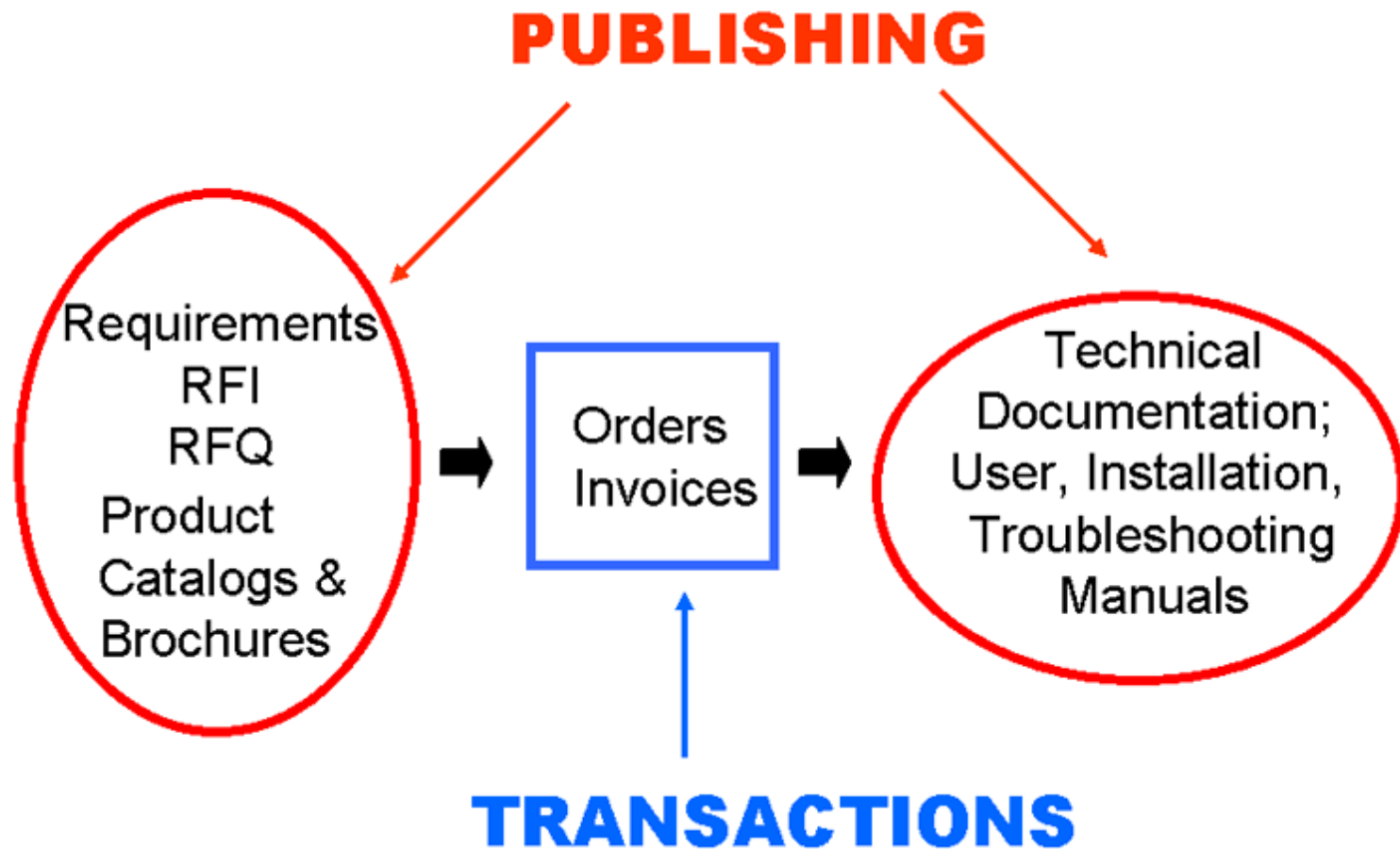
How Should We Understand the Relationships Between Documents and Processes?

Business activities involve documents and the processes that produce and consume them

By understanding the information in the documents, we learn what kinds of processes are possible

By understanding the processes, we learn what kinds of information are needed

A Process-Centric Depiction

Product Catalog

Order

Invoice
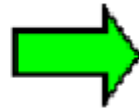
Buyer Creates Order

Seller Processes Order & Creates Invoice

A Document-Centric Depiction
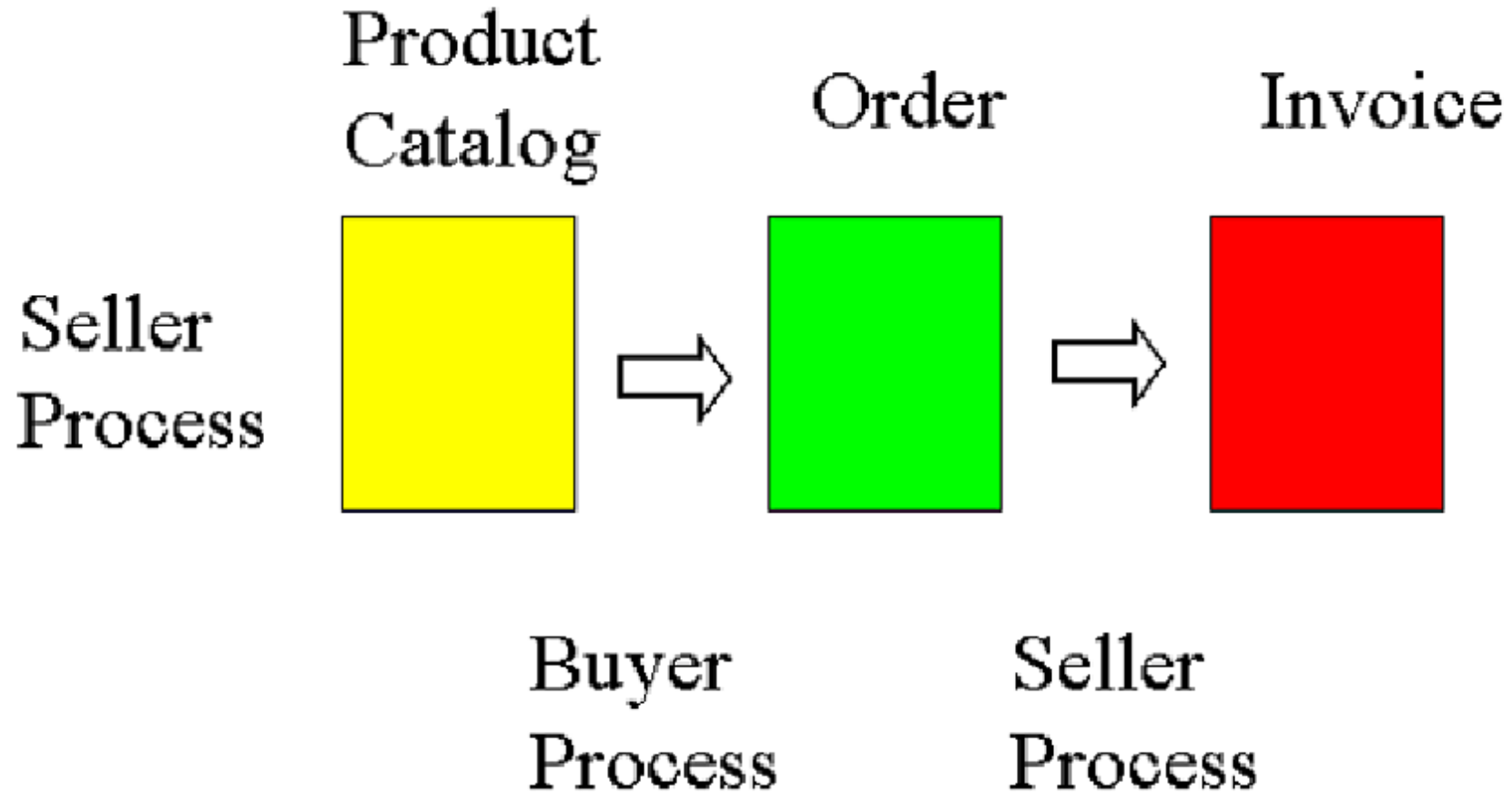
So How Can We Understand Documents/Data/Processes In a Systematic Way?

A focus on processes progressively refines a broadly scoped description of business activities

- "Making it all work" from a business perspective

- Inherently a "top down" approach

A focus on documents and data emphasizes the information objects and flows in a domain and the requirements for implementing a system in it:

- "Making it all work" from a technical perspective

- Inherently a "bottom up" approach

Documents and Processes -- Yin and Yang

Design Patterns in Document Engineering

The essence of Document Engineering is its systematic approach for discovering and exploiting the relationships between patterns of different types

- *Business model or organizational* patterns: marketplace, auction, supply chain, build to order, drop shipment, vendor managed inventory, etc.

- *Business process* patterns: procurement, payment, shipment, reconciliation, etc.

- *Business information* patterns: catalog, purchase order, invoice, etc. and the components they contain for party, time, location, measurement, etc.

Document Exchange Patterns

Businesses have long dealt with each other by exchanging documents

We use concepts like "supply chains" and "distribution channels" as metaphors for the coordinated or choreographed flow of information and materials/products between businesses

These are complex patterns composed from the document exchange pattern

The processes that comprise these patterns are "glued together" by overlapping information components in the documents

Document Exchange (Physical Model)

Document Exchange (Conceptual Model)
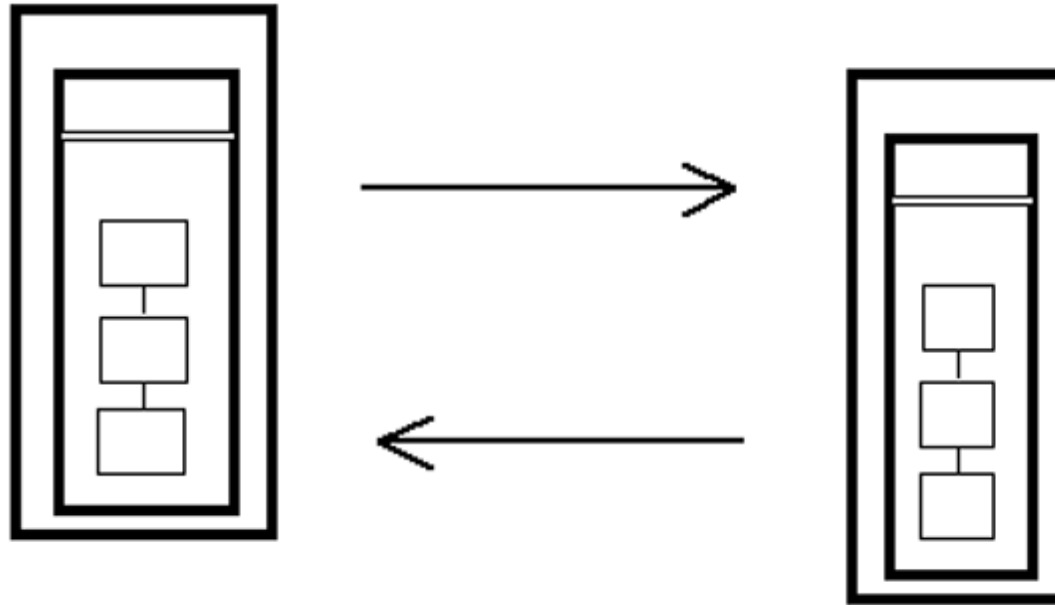
If you send me a **catalog request**, I will send you a **catalog**

If you send me a **purchase order** and I can fulfil it, I will send you a **purchase order response**

Document Engineering (1)

file:///C:/Documents%20and%20Settings/glushko/My%20Documents/Lectures/Varian-200703...

The Drop Shipment Pattern

Customer selects a book from an online bookstore

Customer pays with credit card

Book arrives via express shipper two days later

The Virtual Store as Choreographed Document Exchanges
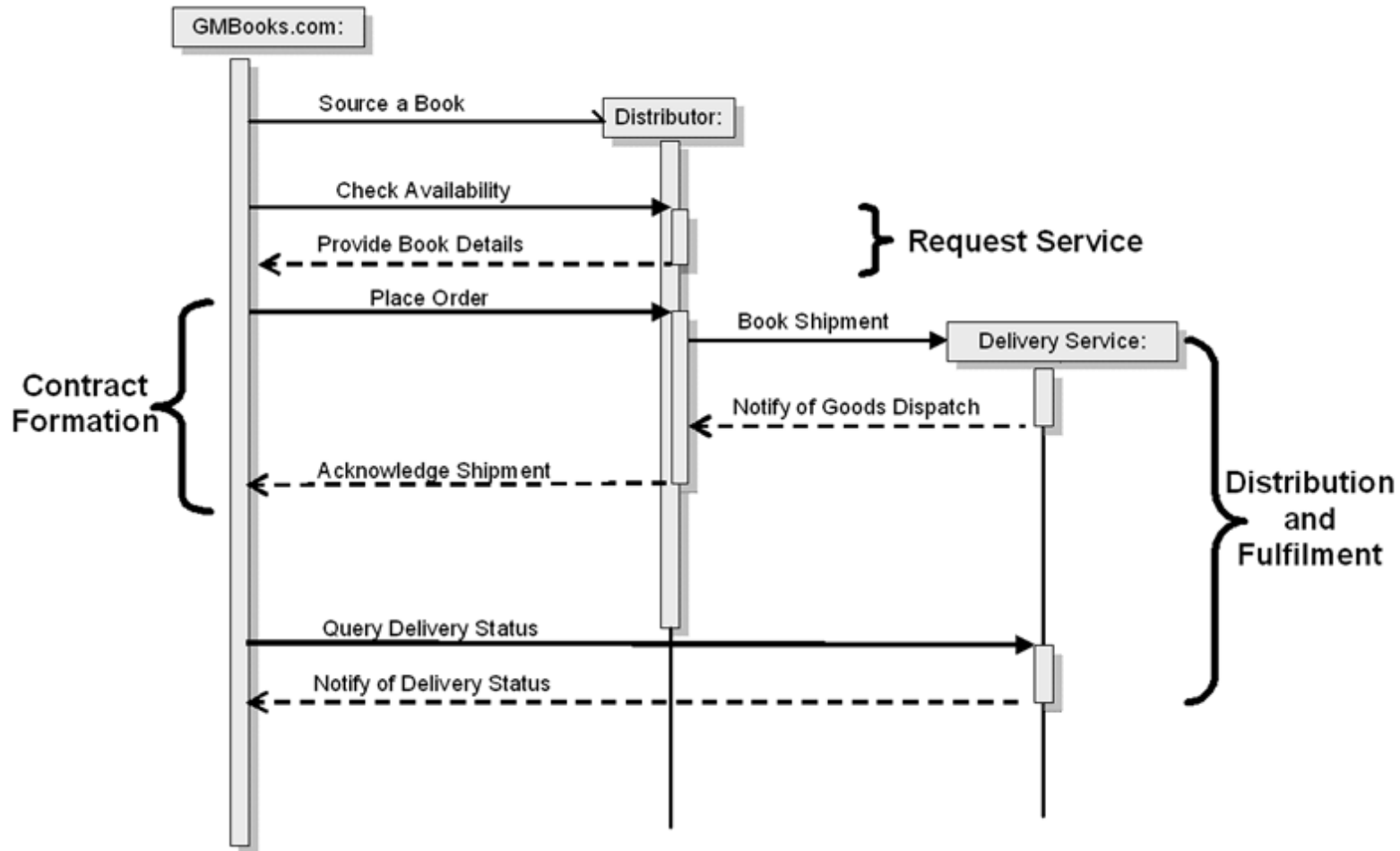
Overlapping Information Models in the Virtual Store
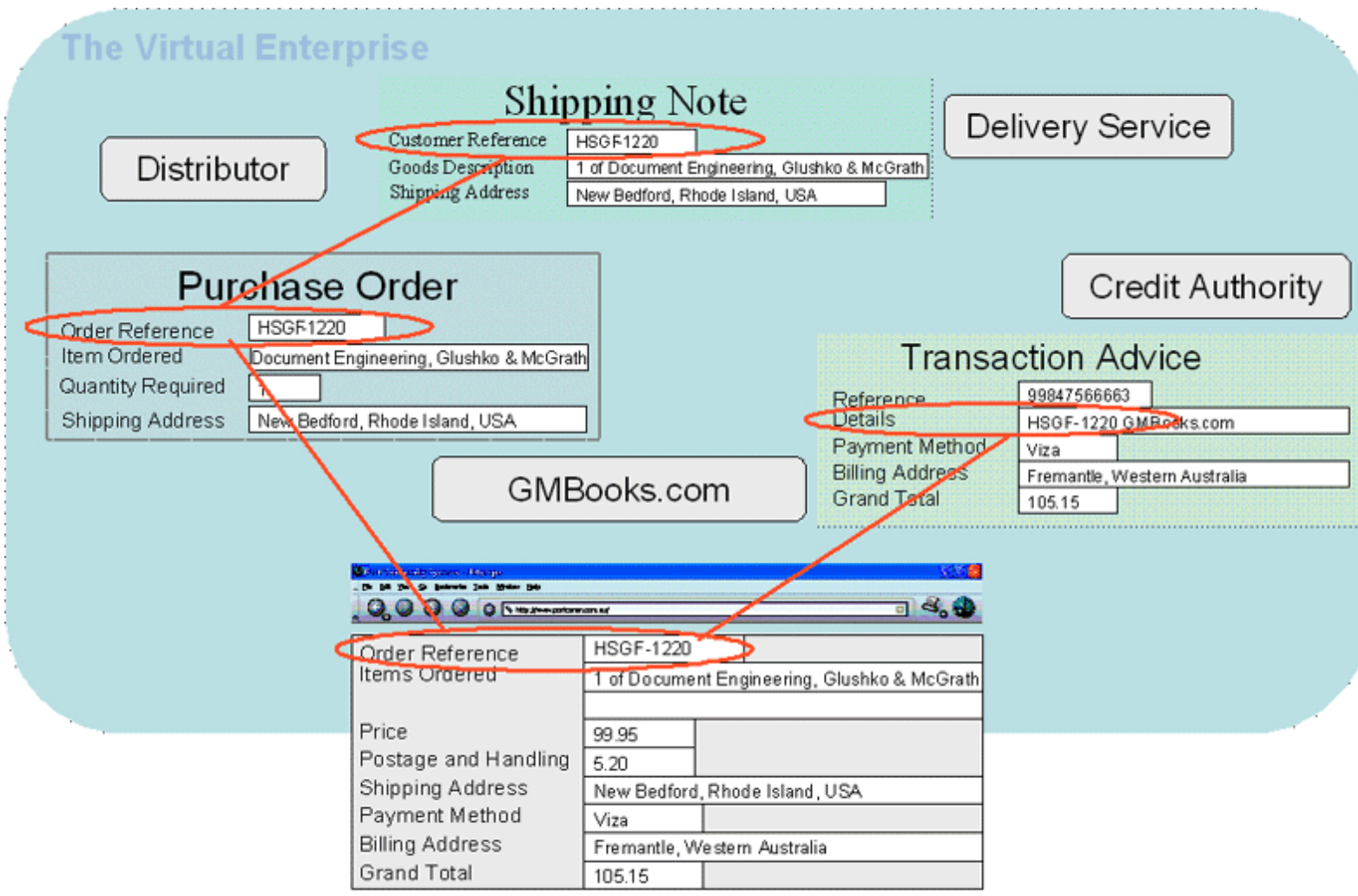
Patterns in the "Model Matrix"
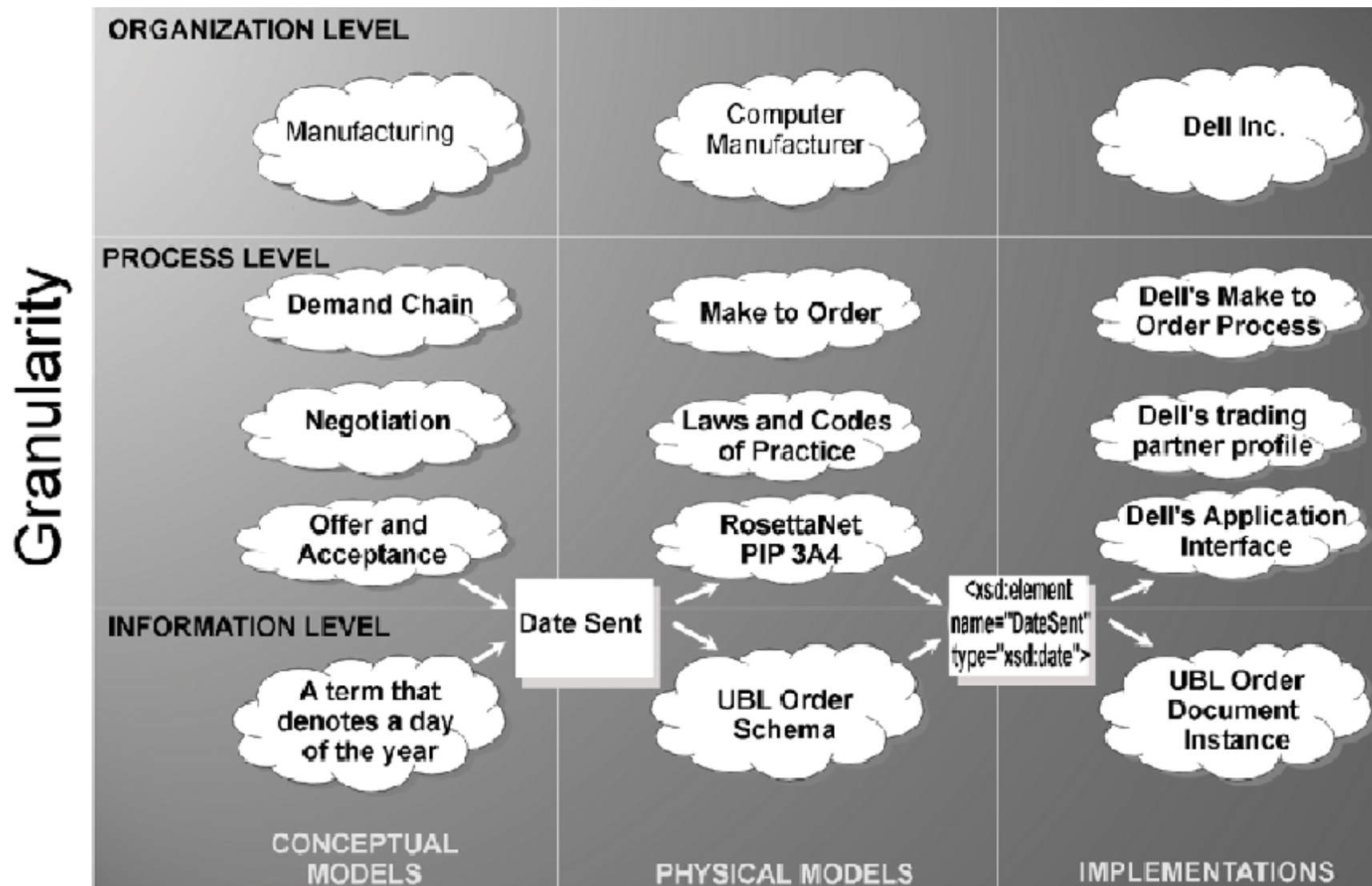
"Meeting in the Middle"

To "bridge the gap between strategy and implementation" we need models of the desired business processes and the documents that they will produce and consume at the same level of detail and implementability

Reaching the middle from the top down ensures that a business model is feasible

Reaching the middle from the bottom up ensures that we are designing and optimizing the activities that add the most value

"Meeting in the Middle" -- 4 Modeling Approaches
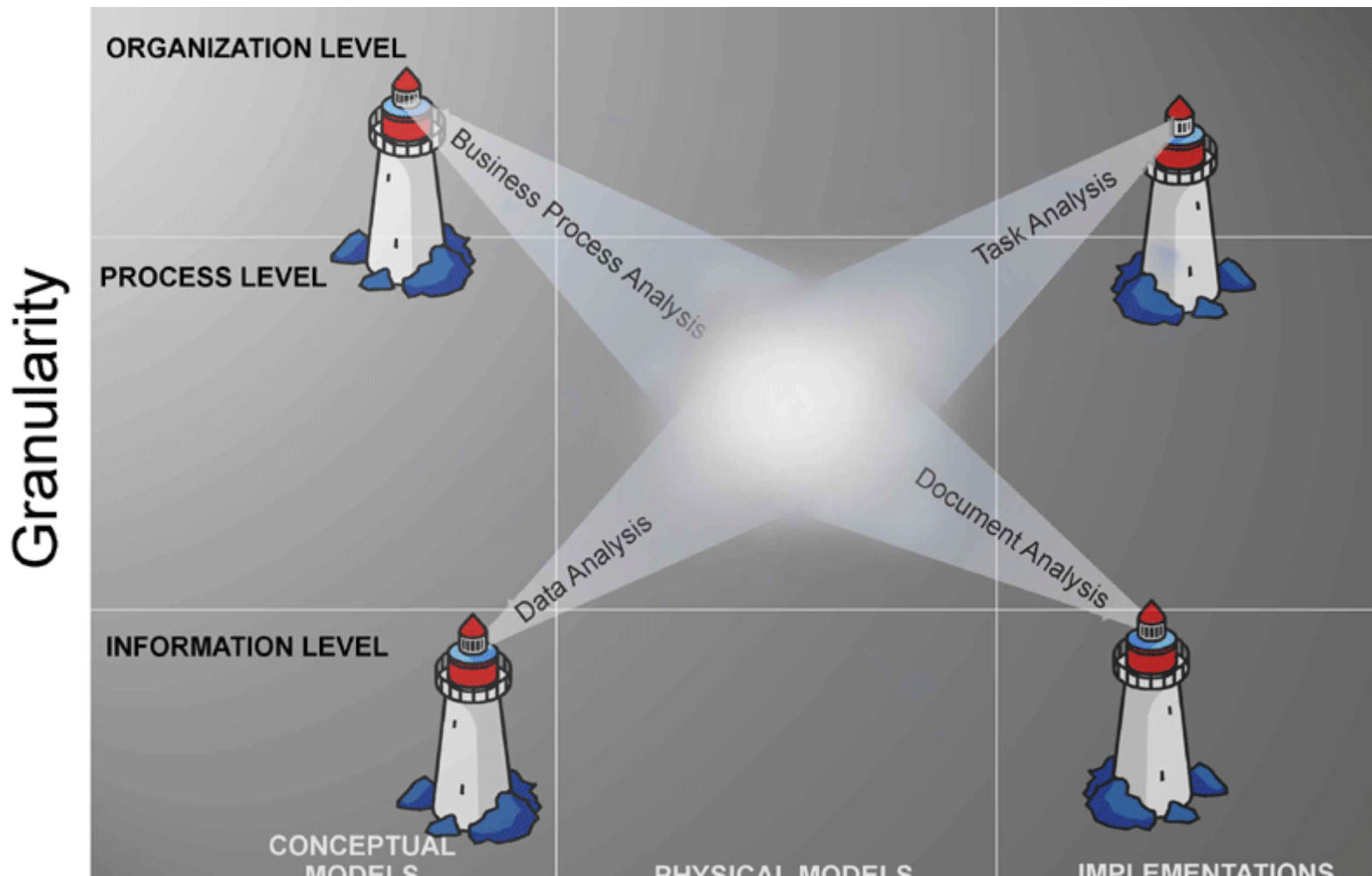
The Document Engineering Approach
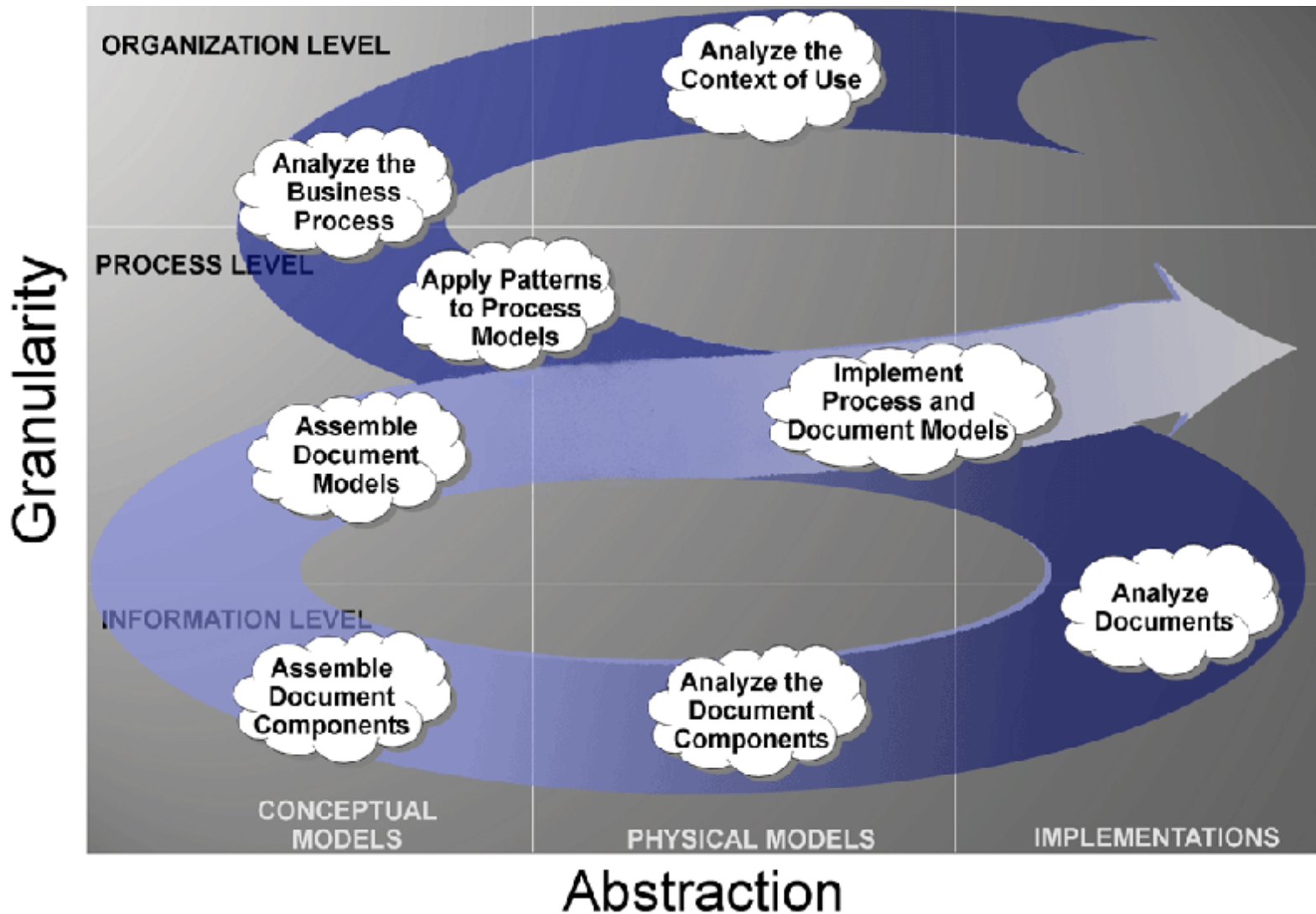
So Document Engineering Isn't About XML

XML is a useful technology for Document Engineering, but using XML doesn't make you a document engineer

The *best* thing about XML is the ease with which you can create a new vocabulary for a particular type of document

XML is just the syntax in which we encode document models... what really matters is how we modeled the documents

Creating Models is Easy, But Creating GOOD Models is Hard

The *worst* thing about XML is the same as the best thing – the ease with which you can create a new vocabulary

No way around the classical problems of classification and naming we know from philosophy, linguistics, cognitive psychology, and information science

XML is NOT "self-describing"

There are often multiple vocabularies for the same or related domains and especially for the common information models that are used in more than one domain

A Checklist for Describing Projects and Case Studies

D -- data types and document types

O -- organizational processes

C -- context (types of products or services, industry, geography, regulatory considerations)
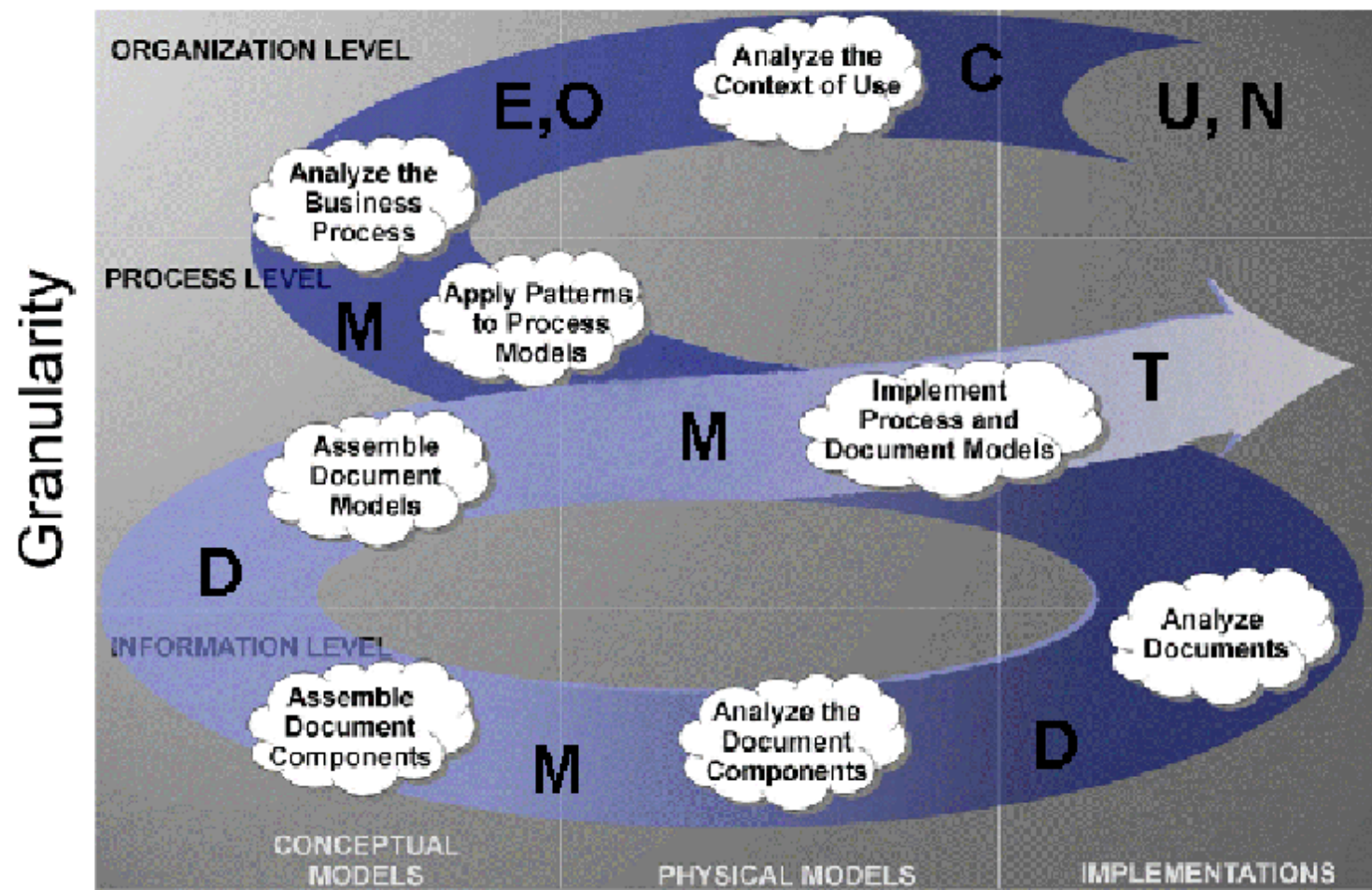
U -- user types and special user requirements

M -- models, patterns, or standards that apply

E -- enterprises and eco systems (e.g., trading communities, standards bodies)

N -- the needs (business case) driving the enterprise(s)

T -- technology constraints and opportunities

D-O-C-U-M-E-N-T in the Document Engineering Approach

Summary: Document Engineering's Big Ideas

"Document Engineering" is evolving as a new discipline for specifying, designing, and implementing electronic documents and information-intensive services

Best practices in Document Engineering require and reinforce the identification and reuse of patterns of information exchange of different levels of granularity and abstraction

Business activity always involves both "narrative" and "transactional" documents – so we need analysis and design methods that work for both ends of this "Document Type Spectrum"

Document Engineering can emphasize what these analysis and design approaches have in common rather than highlighting their differences

The methodology we've systematized for Document Engineering

Document Engineering - The Book

Acknowledgments

Much of this material comes from a book called *Document Engineering: Modeling for Business Informatics and Web Services* by Robert J. Glushko and Tim McGrath. MIT Press (2005)

Three years of students at the University of California, Berkeley have contributed to its development through courses and research projects with the first author

The methodology has been significantly refined through its use by the library content team of the Universal Business Language initiative, led by the second author