

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

The screenshot shows the top of the Wall Street Journal website. The logo "WSJ.com THE WALL STREET JOURNAL ONLINE" is prominent. To the right, there is an "Article Search" box with a "Play" button and a link to "Advanced Search". Further right, a "Quotes" section is partially visible. Below the logo, a navigation bar includes "Free Dow Jones Sites" and the date "As of Wednesday, November 29, 2006". A vertical menu on the left lists various sections: Home, News, Technology, Markets, Personal Journal, Opinion, and Weekend & Leisure. Below this menu are links for "TODAY'S NEWSPAPER", "MY ONLINE JOURNAL", "MULTIMEDIA & EXTRAS", "MARKET DATA & TOOLS", "FIND A JOB", and "FIND A HOME". A "Special Offer" section at the bottom left encourages users to "Subscribe to the print".

Big Employers Plan Electronic Health Records

By GARY MCWILLIAMS
November 29, 2006; Page B1

Several big employers are about to deliver an electronic jolt to the U.S. health-care system.

Next week, [Intel Corp.](#), [Wal-Mart Stores Inc.](#), [British Petroleum](#) and others will disclose a plan to provide digital health records to their employees and to store them in a multimillion-dollar-data warehouse linking hospitals, doctors and pharmacies. Their goal: to cut costs

Tsunami Aid Delayed by Incomplete Shipping & Customs Documents

Borneo Bulletin

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



InformationWeek
BUSINESS INNOVATION POWERED BY TECHNOLOGY

CMP
United Business Media

IW on the GO
Get News On Your Phone
Submit

NEWS | WINDOWS | SECURITY | OUTSOURCING | INTERNET | SOFTWARE | HARDWARE | MANAG

Industries Tech Center: • News By Vertical Industry • Government • Blog • Healthcare • Ba
• InformationWeek Download

Salesforce.com Unveils Integration Tools For Back-Office Systems

Salesforce.com introduced tools to integrate its customer relationship management service with enterprise resource planning systems, Oracle 11i, and technology from 25 partners.

By [Antone Gonsalves](#)

- » E-Mail
- » Print
- » Discuss
- » Write To Editor
- » Digg
- » Slashdot

Global Shippers Give Customers Real-Time Cargo Info

The screenshot shows the top of a web browser displaying the Wall Street Journal Online homepage. At the top left is the WSJ.com logo. To its right is the text "THE WALL STREET JOURNAL ONLINE". Further right are two search boxes: "Article Search" with an "Advanced Search" link below it, and "Quotes & Research" with radio buttons for "Symbol(s)" and "Name". Below the search boxes is a navigation bar with "Free Dow Jones Sites" and the date "As of Thursday, January 4, 2007". On the left side is a vertical menu with links: Home, News, Technology, Markets, Personal Journal, Opinion, Weekend & Leisure, TODAY'S NEWSPAPER, MY ONLINE JOURNAL, MULTIMEDIA & EXTRAS, MARKET DATA & TOOLS, FIND A JOB, FIND A HOME, and Special Offer. The main content area features the article title "Global Shippers Play Catch-Up In Information Age" by Leila Abboud, dated January 4, 2007, on page B3. The article text begins with "MARSEILLE, France -- New technologies have allowed factories and retailers to tighten the global supply chain in recent years, but one link has struggled to keep up: shipping." and continues with "The world's major ocean-shipping lines -- the pack mules of global trade -- have been slow to introduce new software to match the information revolution taking place in other industries. Now, shipping lines are coming under increasing pressure from customers to feed a steady stream of information about the location of the shippers' goods and other data not usually available in the past, such as the temperature at which the cargo is being..."

WSJ.com THE WALL STREET JOURNAL ONLINE

Article Search Advanced Search

Quotes & Research Symbol(s) Name

Free Dow Jones Sites As of Thursday, January 4, 2007

Home
News
Technology
Markets
Personal Journal
Opinion
Weekend & Leisure
TODAY'S NEWSPAPER
MY ONLINE JOURNAL
MULTIMEDIA & EXTRAS
MARKET DATA & TOOLS
FIND A JOB
FIND A HOME
Special Offer
Subscribe to the print Journal today and receive 8 weeks FREE! Click Here!

Global Shippers Play Catch-Up In Information Age

By LEILA ABBOUD
January 4, 2007; Page B3

MARSEILLE, France -- New technologies have allowed factories and retailers to tighten the global supply chain in recent years, but one link has struggled to keep up: shipping.

The world's major ocean-shipping lines -- the pack mules of global trade -- have been slow to introduce new software to match the information revolution taking place in other industries. Now, shipping lines are coming under increasing pressure from customers to feed a steady stream of information about the location of the shippers' goods and other data not usually available in the past, such as the temperature at which the cargo is being

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

International Journal of Law and Information Technology © Oxford University Press 2006; all rights reserved
doi:10.1093/ijlit/ean019

Disrupting conventional law firm business models using document assembly

DARRYL R. MOUNTAIN¹

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 disrupting the business of law firms. It looks at those business models that are...

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

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?

Department of the Treasury—Internal Revenue Service

Form **1040EZ** **Income Tax Return for Single and Joint Filers With No Dependents** (99) **2003** OMB No. 1545-0675

Label
(See page 12.)
Use the IRS label.
Otherwise, please print or type.

Label HERE

Presidential Election Campaign (page 12)

Your first name and initial	Last name	
If a joint return, spouse's first name and initial	Last name	
Home address (number and street). If you have a P.O. box, see page 12.		Apt. no.
City, town or post office, state, and ZIP code. If you have a foreign address, see page 12.		

Note. Checking "Yes" will not change your tax or reduce your refund.
Do you, or your spouse if a joint return, want \$3 to go to this fund?

Your social security number

Spouse's social security number

▲ Important! ▲
You **must** enter your SSN(s) above.

You Spouse

Yes No Yes No

Income	<p>1 Wages, salaries, and tips. This should be shown in box 1 of your Form(s) W-2. Attach your Form(s) W-2. 1</p> <p>2 Taxable interest. If the total is over \$1,500, you cannot use Form 1040EZ. 2</p> <p>3 Unemployment compensation and Alaska Permanent Fund dividends (see page 14). 3</p> <p>4 Add lines 1, 2, and 3. This is your adjusted gross income. 4</p> <p>5 Can your parents (or someone else) claim you on their return? Yes. Enter amount from <input type="checkbox"/> worksheet on back. No. If single, enter \$7,800. If married filing jointly, enter \$15,600. See back for explanation. 5</p> <p>6 Subtract line 5 from line 4. If line 5 is larger than line 4, enter -0-. This is your taxable income. 6</p>	
Payments and tax	<p>7 Federal income tax withheld from box 2 of your Form(s) W-2. 7</p> <p>8 Earned income credit (EIC). 8</p> <p>9 Add lines 7 and 8. These are your total payments. 9</p>	

Note. You **must** check Yes or No. }

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

```
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 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



But A Catalog Is Data (Document)

Industrial Or Light Weight Bags On A Roll

Bags are perforated allowing easy tear off for in-store or assembly line use. Choose industrial 2-mil or extra heavy 4-mil for parts fittings and hardware. Lightweight bags are .5 mil and ideal for produce and lighter weight items. Table or wall mount dispenser available below.



Industrial Bags On A Roll

Size (In.)	Bags Per Roll	Part No.		Price	
		2 Mil	2 Mil	4 Mil	4 Mil
4 x 6	1000	88400LU	25.55	88430LU	45.55
6 x 9	1000	88403LU	39.15	88433LU	59.90
8 x 10	1000	88406LU	47.15	88436LU	85.65
10 x 12	1000	88409LU	68.25	88439LU	125.30


Discount per part no.: Less 5% 12-23 rolls; 15% 24 rolls or more.

Lightweight Bags On A Roll

Size (In.)	Bags Per Roll	Part No.	Price Per Carton of 2 Rolls		
			.5 Mil	1-11	12-23
10 x 15	2000	88080LU	52.80	50.16	44.88
10 x 20	1500	88085LU	52.80	50.16	44.88
Dispenser		88090LU	17.80 each		

Lay-Flat Poly Tubing Rolls

Simply cut tubing to your exact length and seal with the Consolidated Impulse Heat Sealer found on page 80. Choose 2 mil or 4 mil tubing stock. Ideal for a variety of different size parts. FDA approved.



2 Mil

Part No.	W x L (In. x Ft.)	Price/Roll	Part No.	W x L (In. x Ft.)	Price/Roll
89965LU	2 x 2100	33.66	89972LU	12 x 2100	101.18
89967LU	3 x 2100	39.40	89973LU	14 x 2100	119.86
89968LU	4 x 2100	55.90	89974LU	16 x 2100	138.60
89969LU	5 x 2100	63.85	89975LU	18 x 2100	147.72
89971LU	6 x 2100	62.66	89976LU	20 x 2100	160.05
89979LU	8 x 2100	78.64	89940LU	24 x 1700	168.43
89981LU	10 x 2100	94.40	89941LU	36 x 1100	163.49

4 Mil

Part No.	W x L (In. x Ft.)	Price/Roll	Part No.	W x L (In. x Ft.)	Price/Roll
89980LU	2 x 1050	33.66	89993LU	12 x 1050	101.18
89982LU	3 x 1050	39.40	89994LU	14 x 1050	119.86
89983LU	4 x 1050	55.90	89996LU	16 x 1050	138.60
89984LU	5 x 1050	63.85	89997LU	18 x 1050	147.72
89986LU	6 x 1050	62.66	89998LU	20 x 1050	160.05
89989LU	8 x 1050	78.64	89942LU	24 x 850	168.43
89992LU	10 x 1050	94.40	89943LU	36 x 550	163.49

Discount per part no.: Less 5% 5-11 rolls; 10% 12-23 rolls; 15% 24 rolls or more.

And a Reference Book is Document (Data)

1.1 Measurement of Light

1.103 Range of Light Intensities Confronting the Eye

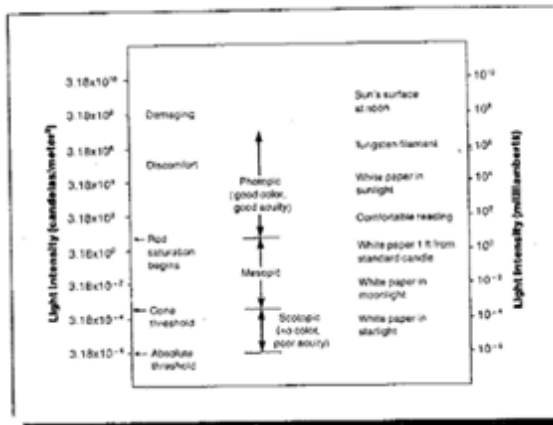


Figure 1. The range of light intensities that confront the human eye. (Adapted from C. H. Graham [Ed.], *Vision and visual perception*. Copyright © 1965 by John Wiley & Sons, Inc. Reprinted with permission.)

Key Terms

Illumination level; luminance; mesopic vision; photopic vision; scotopic vision

General Description

The human eye is sensitive to a wide range of light intensities, from a minimum visible level of $\sim 0.00003 \text{ cd/m}^2$ to an upper tolerance limit of over $300,000 \text{ cd/m}^2$. Vision at very low levels of illumination (e.g., starlight) is termed **scotopic** vision and is mediated by the **rods**; visual acuity is poor with scotopic vision and no sensation of color (hue) occurs. Vision at high intensity levels (e.g., daylight) is known as **photopic** vision and is mediated by the **cones**; photopic vision is characterized by high visual acuity and

the perception of color. **Mesopic** (mixed) vision (mediated by both rods and cones) occurs with intermediate light intensities (e.g., moonlight).

Figure 2 shows how outdoor brightness decreases during twilight. **Dark adaptation** of the eye with declining illumination is at least as rapid as this normal decline in ambient illumination at evening. Figure 3 shows how the luminance of a test patch changes with the angular elevation of the sun above the horizon.

Constraints

- Sensitivity to light depends on the eye's state of **adaptation**. Maximum scotopic sensitivity requires $\sim 1 \text{ hr}$ of dark adaptation even after as little as a few minutes' exposure to photopic light levels. The time course of **light adaptation** is similar for rods and cones and is much faster than dark adaptation, requiring only a few minutes' exposure at a high luminance level.

Key References

1. Bochs-Coxer, J. G., Czerniacki, L., & Mead, L. C. (1944). Daylight training of pilots for night flying. *Aeronautical Engineering Review*, 3, 9-26.
2. Graham, C. H. (Ed.) (1965). *Vision and visual perception*. New York: Wiley.

Cross References

- 1.101 Range of visible energy in the electromagnetic radiation spectrum

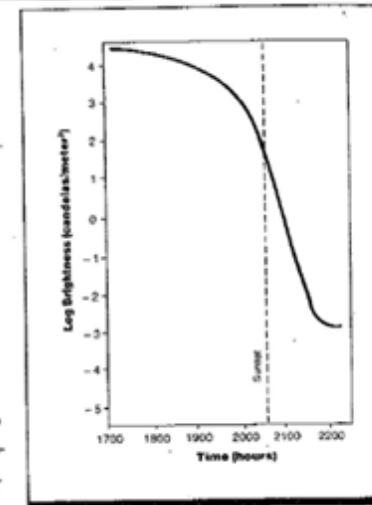


Figure 2. The decrease in brightness from daylight (5:00 pm) to darkness (sunset $\sim 9:30 \text{ pm}$) on July 14, 1942 (adapted from Ref. 2). Angle of test patch in relation to sun not given.

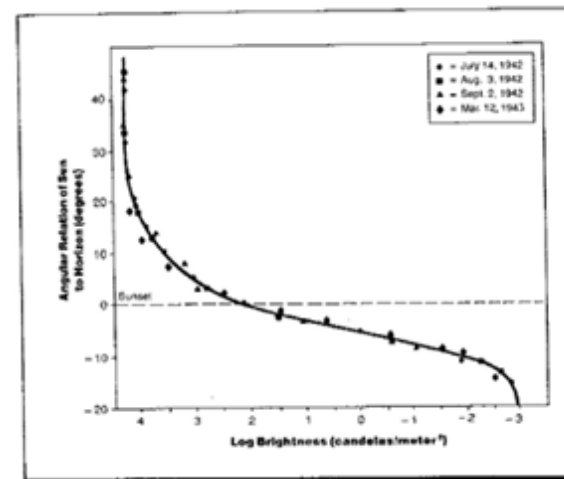
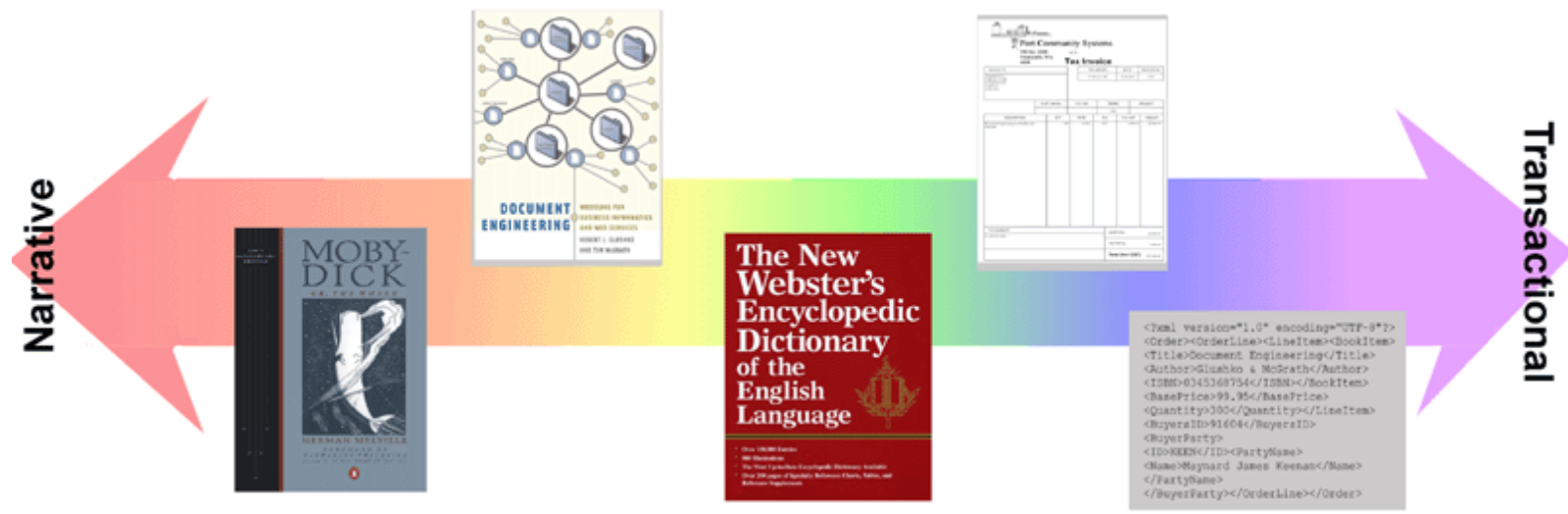
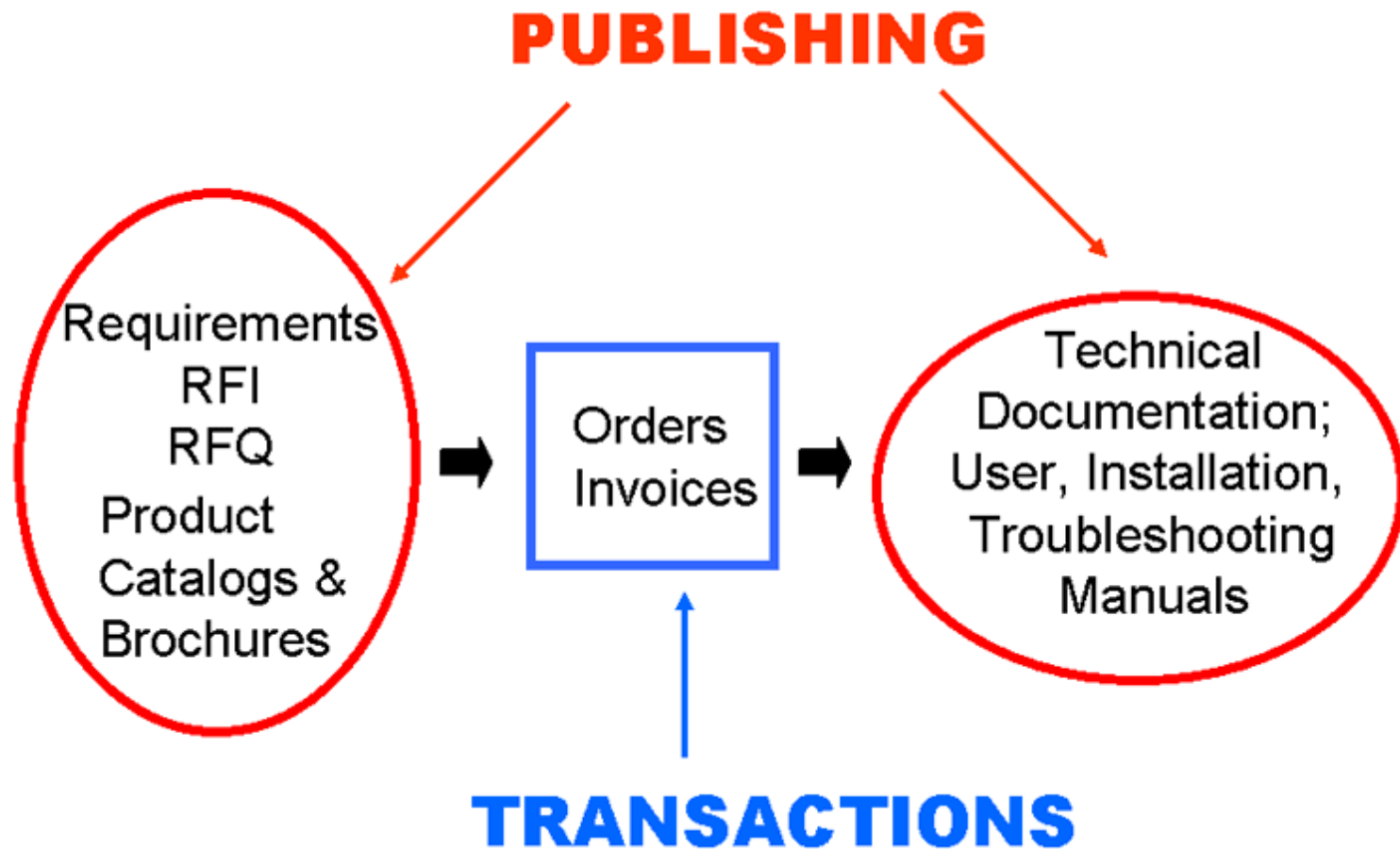


Figure 3. The relationship between the brightness of a stimulus patch in a constant position and the angular elevation of the sun relative to the horizon on four days during different seasons of the year. (From Ref. 1)

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

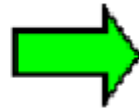
By understanding the processes, we learn what kinds of information are needed

A Process-Centric Depiction

Product
Catalog

Order

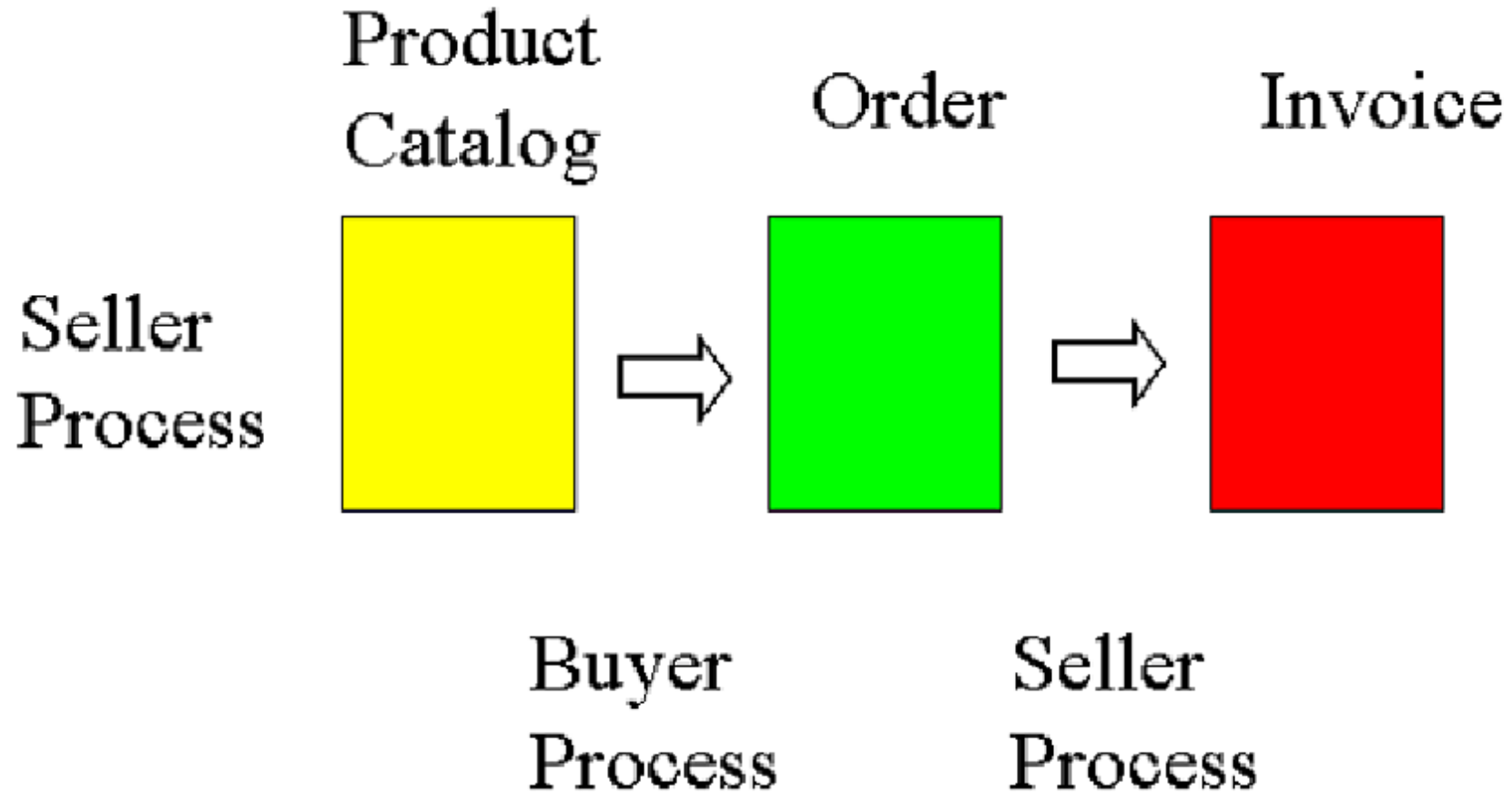
Invoice



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

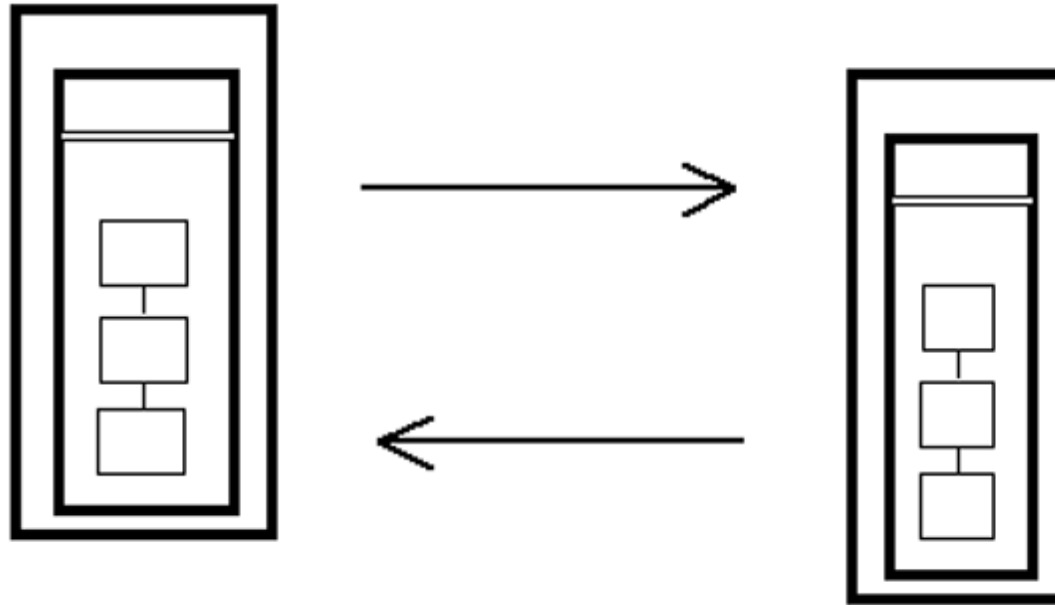
These are complex patterns composed from the document exchange pattern

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**

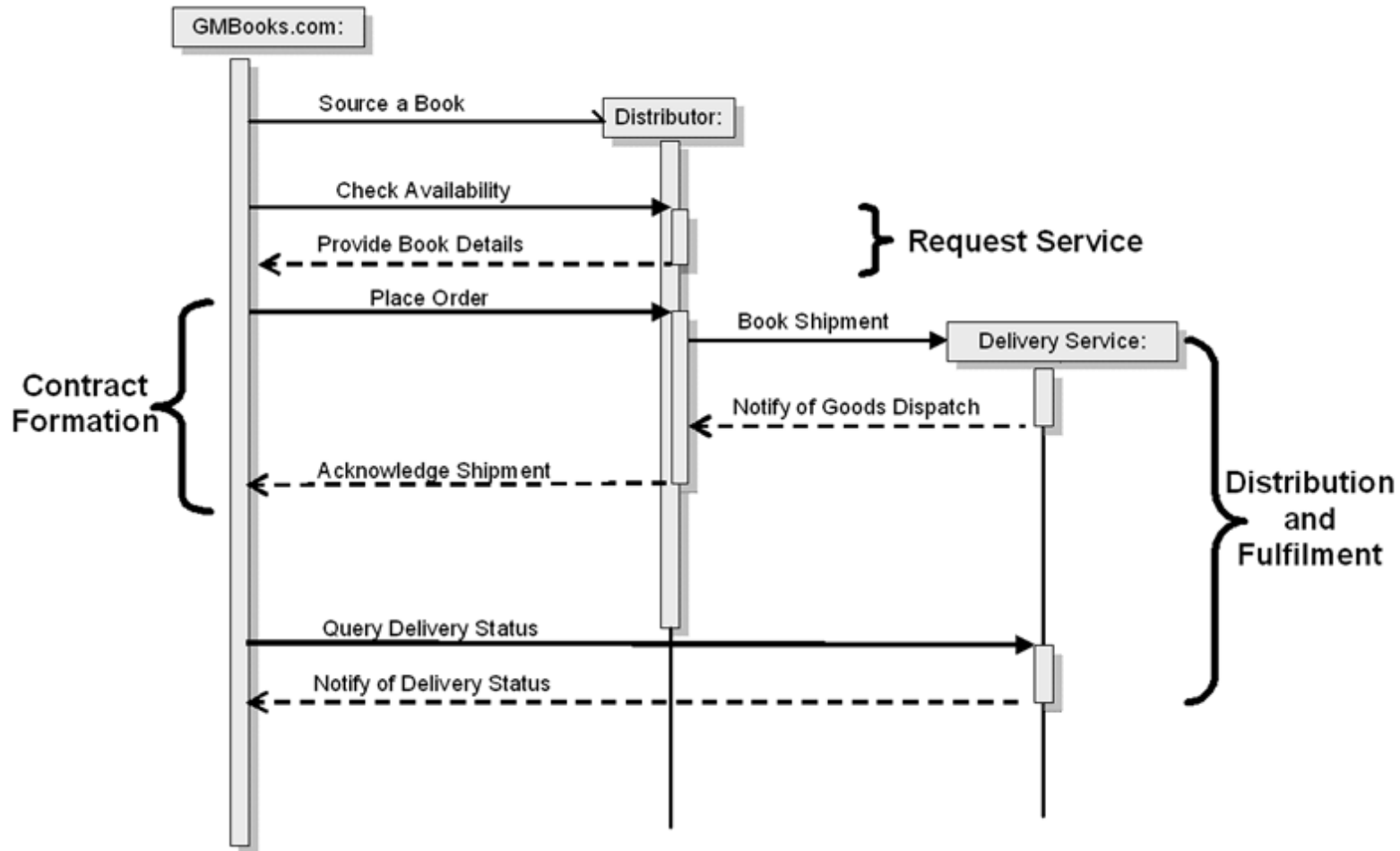
The Drop Shipment Pattern

Customer selects a book from an online bookstore

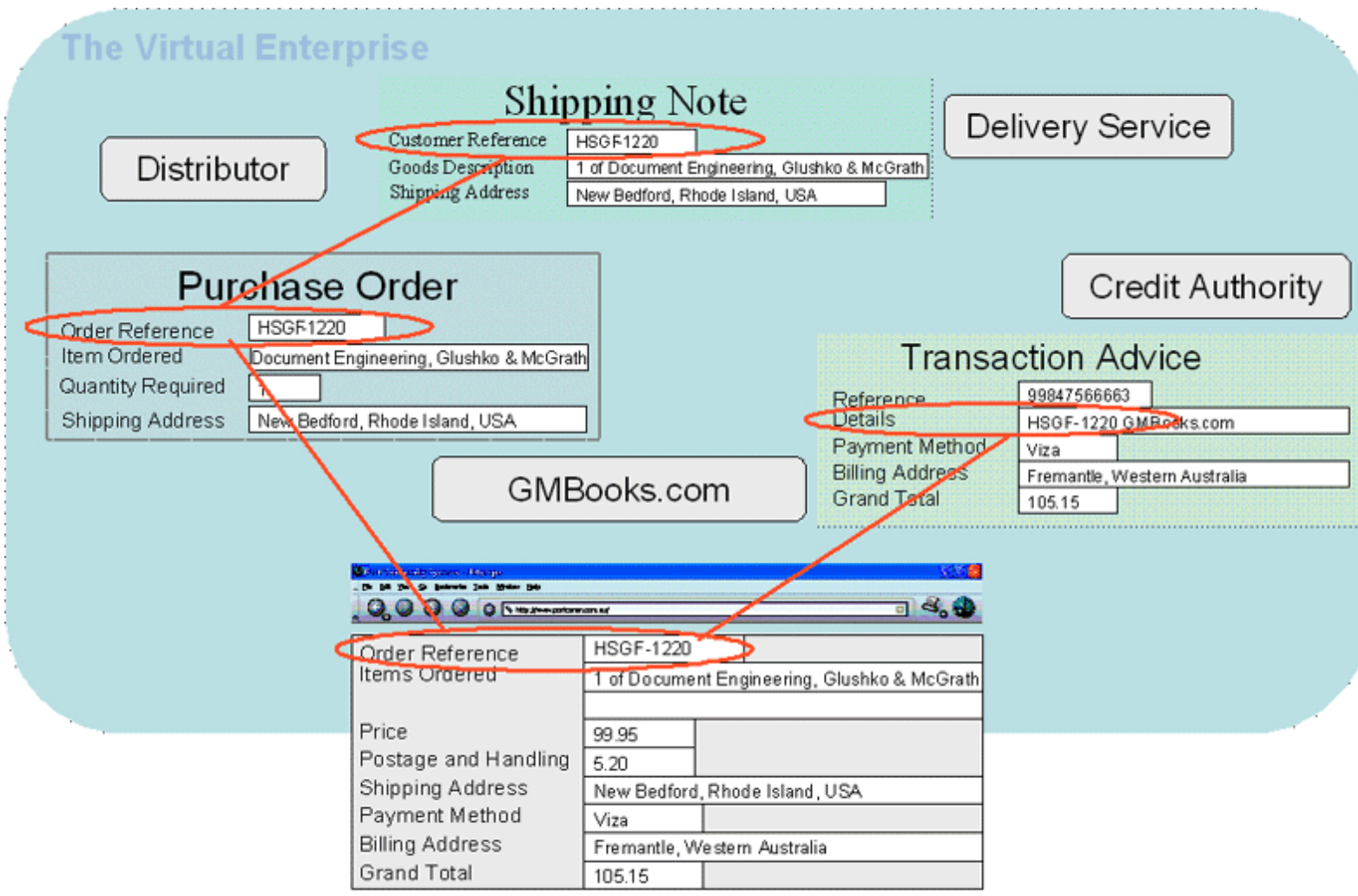
Customer pays with credit card

Book arrives via express shipper two days later

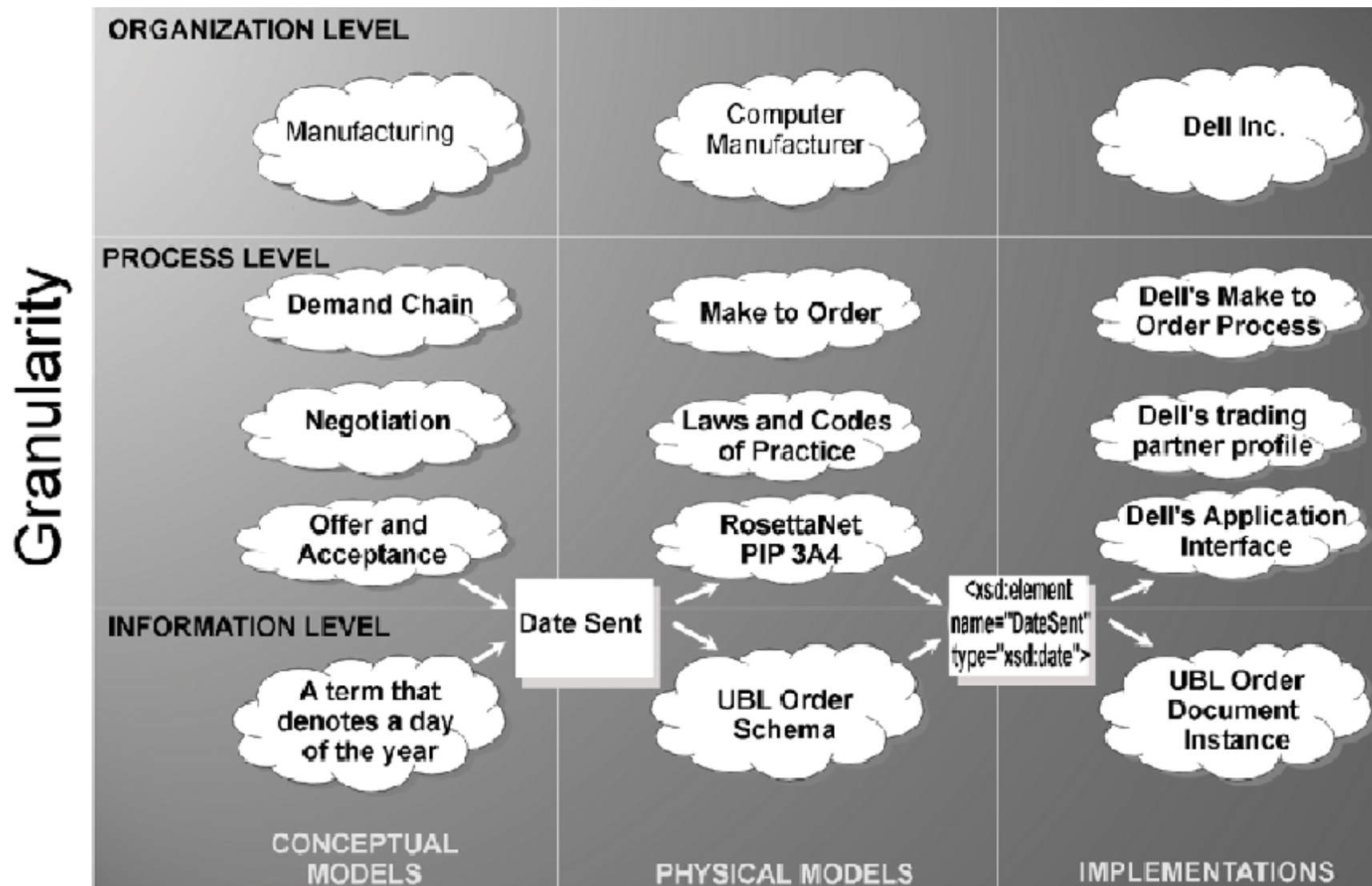
The Virtual Store as Choreographed Document Exchanges



Overlapping Information Models in the Virtual Store



Patterns in the "Model Matrix"



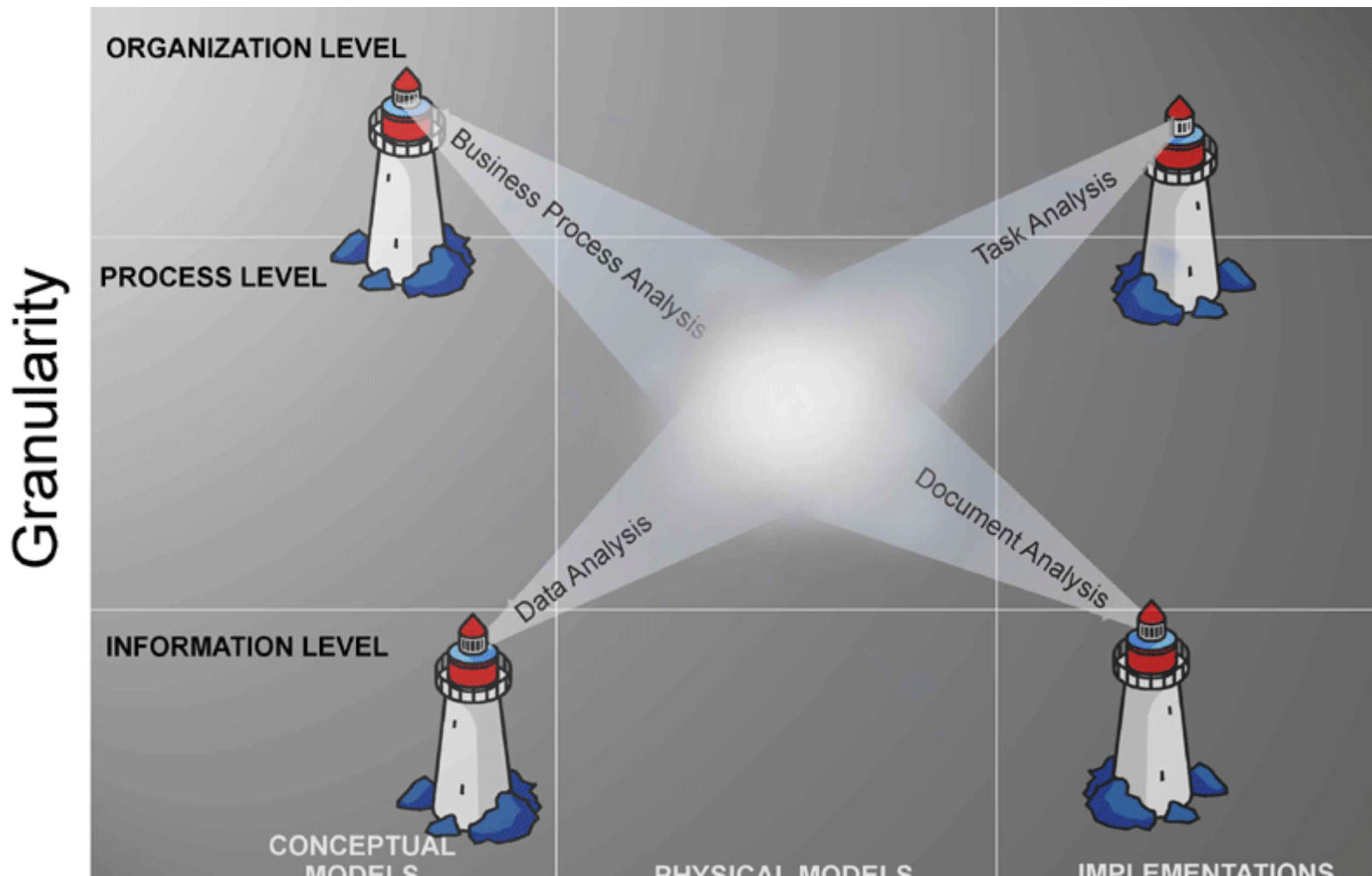
"Meeting in the Middle"

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

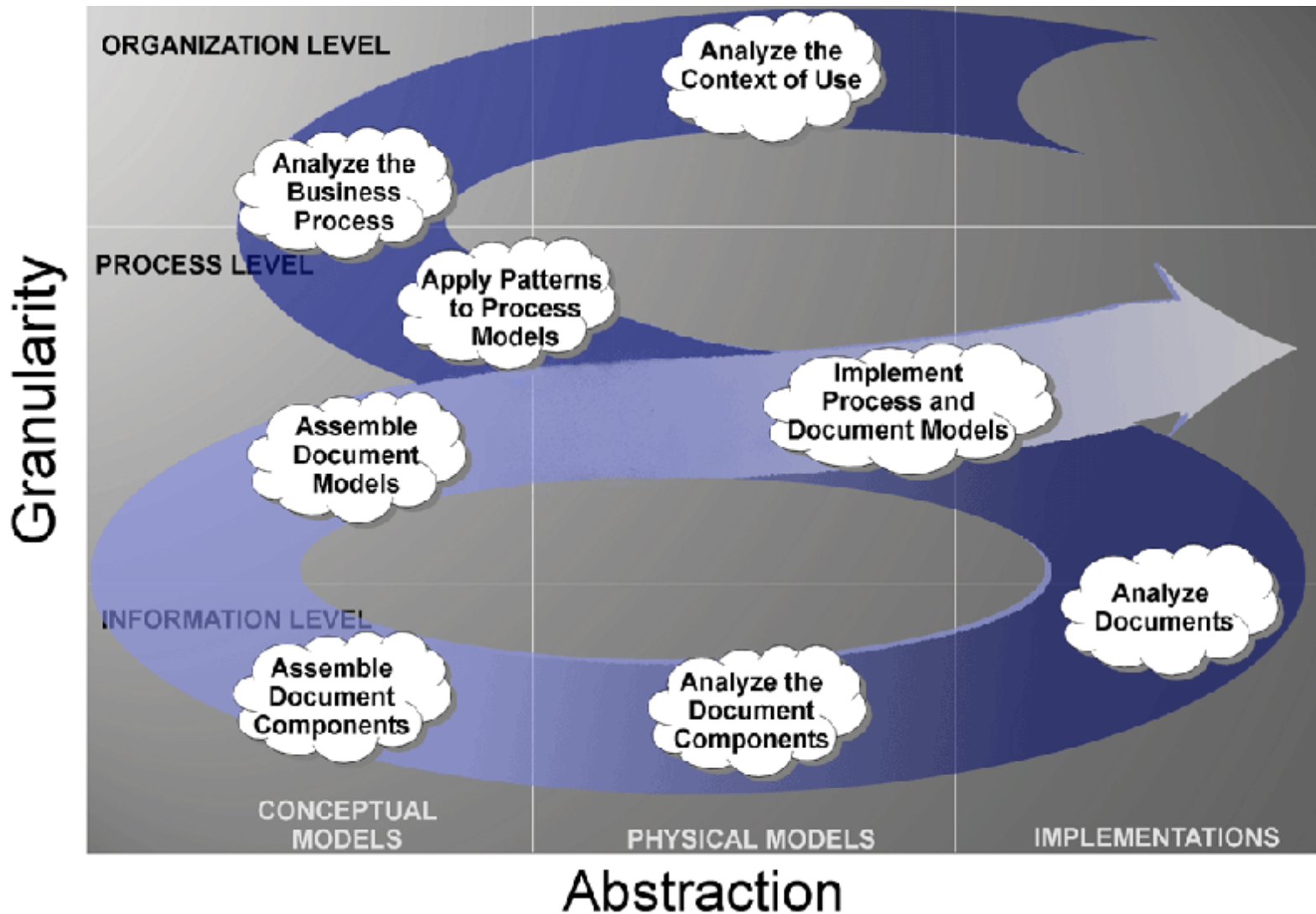
Reaching the middle from the top down ensures that a business model is feasible

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

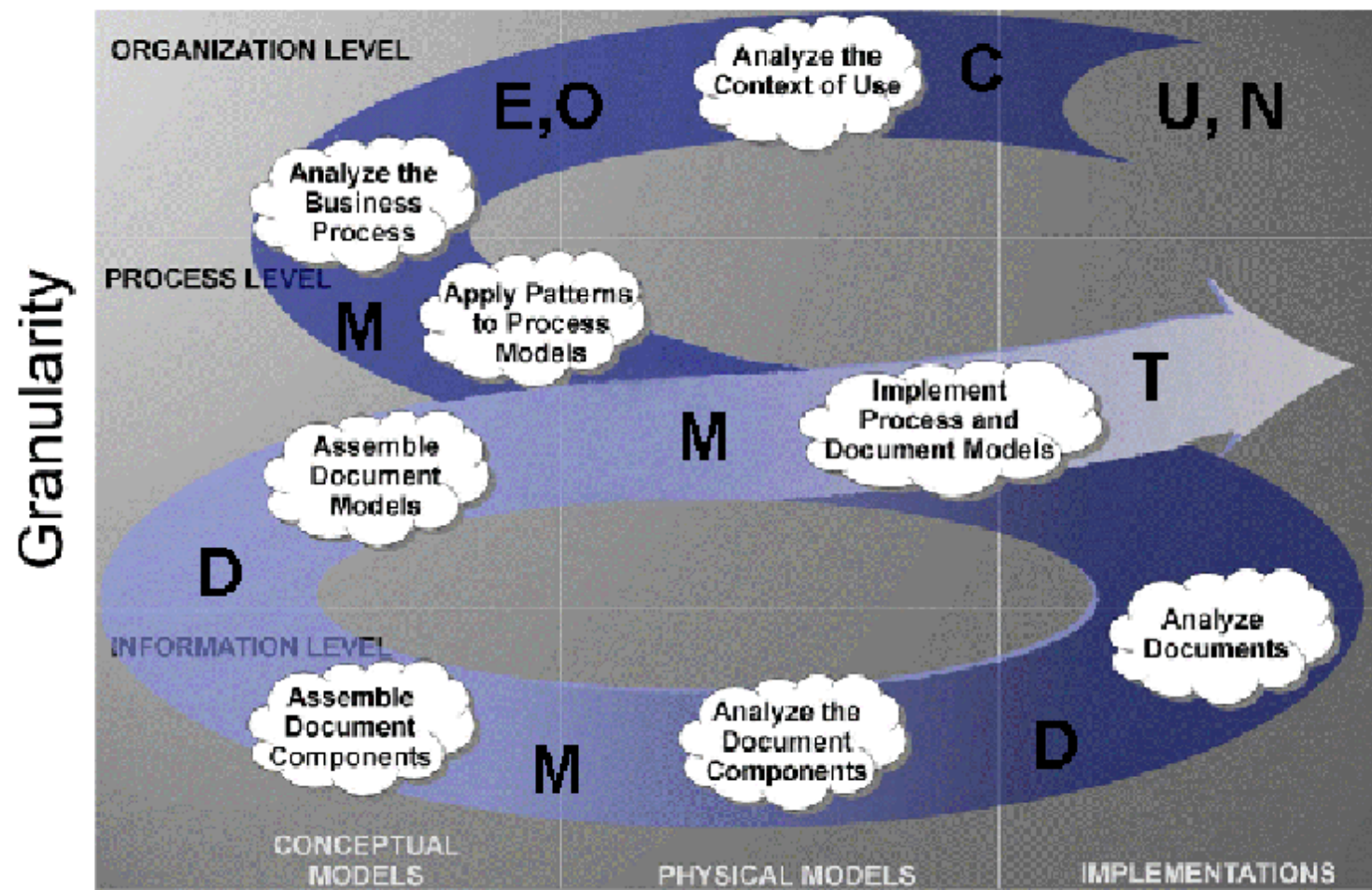
M -- models, patterns, or standards that apply

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