A TECHNOLOGY AND POLICY ANALYSIS FOR GLOBAL E-BUSINESS

by

Hongwei Zhu

B.S. Thermal Eng., Env. Eng., Tsinghua University, 1991
M.S. Env. Eng., Tsinghua University, 1994
M.S. Env. Eng., University of Cincinnati, 1998

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Technology and Policy

at the

Massachusetts Institute of Technology

June 2002

Signature of Author_____
Technology and Policy Program, Engineering Systems Division
May 17, 2002

Certified by _____
Stuart E. Madnick
John Norris Maguire Professor of Information Technology
Professor of Engineering Systems
Thesis Supervisor

Accepted by _____
Daniel Hastings
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program
Chairman, Committee for Graduate Students

A TECHNOLOGY AND POLICY ANALYSIS FOR GLOBAL E-BUSINESS

by

Hongwei Zhu

Submitted to the Engineering Systems Division
on May 17, 2002 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Technology and Policy

Abstract

We introduce an e-business analytical framework that focuses on transaction flows, including information, physical goods, and services. Within this framework, global e-business involves transaction flows that cross both organizational and national boundaries. Many challenging technology and policy issues arise from this trans-boundary characteristic of global e-business. These issues are analyzed using web aggregation as an example global e-business application.

We start the analysis by introducing web aggregation services and their enabling technologies. Our survey of current status of web aggregation indicates that most services are still operated regionally despite their global presence. Although benefits of web aggregation have been realized in regions with extensive use of information aggregation, little is done at the global level. Our case study on worldwide price distribution of a nearly homogeneous consumer electronics product indicates great potential for global aggregation to bring information and efficiency to the global market.

In addition to lack of global integration, we identified other deficiencies of web aggregation. Technological challenges and possible solutions to overcoming these deficiencies are discussed. However, having technological capability for trans-boundary information flow does not solve all problems in global aggregation. National policies often prohibit such flow into nations that have different policies, especially in database and privacy protection areas. We analyze these policy issues and propose future research on international policy harmonization.

Thesis Supervisor: Stuart E. Madnick
Title: John Norris Maguire Professor of Information Technology &
       Professor of Engineering Systems

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1   INTRODUCTION

With its increasing connectivity and capability, the World Wide Web is becoming the platform for global e-business, which involves intertwining social, economics, technology, and policy issues. Before we delve into a detailed analysis of these issues, let's look at a simple online shopping scenario.

1.1 Motivational Example

Imagine for the moment you are from Sweden and interested in buying a pocket sized digital camcorder. After some research on the Web you decide to buy a SONY DCR-IP5, which records video in MPEG format for easy editing on computers and weighs only 12 ounces (i.e., 336 grams). So you launch your favorite comparison shopping engine to find the best deals and it returns information as shown in Figure 1.

Produkt	Butik	Märke	Pris	Lev.tid	Fraktpris	Totalpris	
DCR-IP5 Visa mer	FOTO Butiksinfo	SONY	kr 19,495	2-5 D	kr 115	kr 19,610	Mer
DCR-IP5 Visa mer	Lars Bengtsson Ljud Video Butiksinfo	SONY	kr 19,900	3-7 D	kr 75	kr 19,975	Mer
Sony DCR-IP5 Visa mer	alfadigi.com Butiksinfo	SONY	kr 17,983	1-3 D	kr 99	kr 18,082	Mer
Sony DCR-IP5 (E) Visa mer	CYBERPHOTO Butiksinfo	SONY	kr 19,115	1-3 D	kr 99	kr 19,214	Mer
Digital videokamera DCR-IP5. Visa mer	OnOff Butiksinfo	SONY	kr 21,994	4-5 D	kr 95	kr 22,089	Mer

Sökresultat: Klart — 5 produkt(er) funna 5 butik(er) funna

Sortera genom att klicka på titlarna:

Sortera resultatet efter: Produkt, Butik, Märke, Pris, Fraktpris, Totalpris

Figure 1. Prices for DCR-IP5 in Sweden.

Among the five vendors, you decide to go with the one that has the lowest price, 18,082 Swedish Krona. Right before you click the button to make a deal, you remember what your product manager said at a recent company meeting – "Think globally". So you ask yourself: *is this the best deal on a global basis?*

Then you launch another shopping engine popular in the U.S. and find a vendor who offers $1,099.99 including shipping, as shown in Figure 2.

Figure 2. An Offer from the U.S.

Between 18,082 Swedish Krona and $1,099.99, where would you buy? A seemingly simple question once you figure out that 1 US dollar is about 10 Swedish Krona. The Swedish offer is 64% more expensive than the U.S. offer. That might be well worth a few extra days delay for shipping. You should have that choice. But why doesn't your favorite shopping engine include the information shown in Figure 2? Can it tell you the price in your own currency once it has a global coverage? You need to give out personal information to the company you buy the product, how will they use this information, especially in the case where the company is outside your country? The list of questions could go on and on. These are the typical issues you would encounter in global e-business.

We need a framework that captures a broad range of issues and helps structure our analysis of global e-business.

1.2 Analytical Framework for Global E-Business

Many global e-business studies have focused on infrastructures on a region-by-region basis (Wolcott et al., 2001) or e-business within a region or an organization (Westland and Clark, 1999). Findings in these studies are helpful in determining the capabilities of conducting e-business at each region. As more and more regions become "e-ready", we need to understand how we can effectively conduct e-business across these regions, where issues like those in the motivational example will arise and should be addressed.

As we see in the motivational example, information goes between entities that are engaged in e-business. Once a deal is made, flows of funds transfer and goods delivery occur. It is natural to

focus our analysis on these various flows in e-business transactions. By looking at whether these flows cross organizational and national boundaries, we can categorize e-business into four categories as shown in Figure 3.

	Domestic e-business: Most online shopping today	Global e-business: Comparing prices of domestic and foreign vendors
Inter-organization		
Intra-organization	Domestic internal e-business: Cross selling between domestic divisions	Global internal e-business: Exchanging information between branches in different countries
Transaction flow: • Information • Physical goods	Intra-nation (domestic)	Inter-nation

Figure 3. Categories of E-Business

The lower left quadrant represents e-business where the transaction flows stay within the organizational and national boundary. Issues involved here are domestic and internal. As we move up to the upper left quadrant, transaction flows cross organizational boundaries. Issues of dealing with external organizations arise. Similarly, moving right to the lower right quadrant results in flows that cross national boundaries but within the same organization. Here international differences may affect such flows. And finally the upper right quadrant represents global e-business that has both organizational and national trans-border flows. Had the Swedish buyer bought the product from the U.S. vendor, the transaction would include trans-border flows and constitute global e-business.

Each type of e-business has a set of unique issues that deserve detailed analysis. Take domestic internal e-business for example. Ownership of data and incentive systems within an organization may affect the flow of information (Van Alstyne, *et al.*, 1995). As we move away from domestic internal e-business into other types of e-business, new issues arise and complexity increases. This framework captures these characteristics of e-business and allows us to analyze technologies and policies that support, and sometimes limit, such transaction flows.

Instead of discussing each type of e-business, we will focus our analysis on global e-business using Web aggregation as an example. The comparison shopping engine used by the hypothetical Swedish user is an example of regional comparison aggregation. Lack of a global aggregation tool makes getting useful information from global sources a difficult task. In the example, the Swedish user has to know the existence of multiple regional comparison aggregators and use them to perform global comparisons by manually integrating results. Global aggregation will automate this process, help the user "think globally", and bring many benefits to the global market.

The rest of the thesis will analyze various technology and policy issues for global aggregation. Section 2 describes web aggregation, including types of aggregation, enabling technologies, current status. Section 3 presents worldwide price dispersion phenomenon, which to some extent indicates the inefficiency of the global market. Section 4 discusses the benefits of aggregation and global aggregation. We devote section 5 to the analysis of deficiencies, technology and

policy challenges to global aggregation. Finally we conclude in Section 6 and discuss future research.

2 WEB AGGREGATION

Web aggregation is a service that provides consolidated information by transparently collecting and analyzing information from multiple web sources (Madnick and Siegel, 2002). As an e-business application, web aggregation can be used in any of the four types of e-business identified in section 1. For example, the COntext INterchange (COIN) project at MIT has an intra-organizational aggregator that consolidates schedules of computer science related seminars posted at various departmental websites. This application involves information flow within the same organization (i.e., MIT) and among different divisions (e.g., AI Laboratory, Laboratory for Computer Science, IT Group of Sloan School, etc.). We can also classify web aggregation according to its primary functionality, such as comparison aggregation and relationship aggregation (Madnick and Siegel, 2002). We will illustrate both types of aggregation using examples next.

2.1 Comparison Aggregation and Relationship Aggregation

Comparison aggregation services are also known as comparison shopbots. Most comparison aggregators perform multi-attribute comparison. Depending on the nature of the goods or services, the attributes being compared vary. Common attributes for consumer goods include vendor rating, base price (some also include estimated tax and shipping cost), and availability. See Figure 4 for an example.

Retailer Price Comparisons

	Retailer	Price	+ Approx. Tax	+ Approx. Shipping*	Resort By = Bottom Line Price	Stock	Merchant Rating
Shop	CompDirect.com Retailer Info	$1098.00	No Tax	$49.00	Your Best Price $1147.00	See Site	★★★★☆ 181 Reviews
Shop	Priceit4Less Retailer Info	$1139.00	No Tax	$19.99	$1158.99	Yes	★★★★☆ 52 Reviews
Shop	CAMERAT→PIA 1-888-507-0010 Retailer Info	$1144.44	No Tax	$44.40	$1188.84	Yes	★★★★☆ 86 Reviews
Shop	dbuys.com Featured Merchant Retailer Info	$1149.00	No Tax	$24.90	$1173.90	Yes	★★★★★ 443 Reviews
Shop	TriStateCamera.com Retailer Info	$1149.99	No Tax	$39.90	$1189.89	Yes	★★★★☆ 28 Reviews
Shop	DIGITAL TAILER Retailer Info	$1155.00	No Tax	$24.95	$1179.95	Yes	★★★★☆ 356 Reviews
Shop	IbuyDigital Retailer Info	$1158.00	No Tax	$24.90	$1182.90	Yes	★★★★★ 210 Reviews
Shop	Abe's of Maine Retailer Info	$1199.95	No Tax	$71.99	$1271.94	Yes	★★★☆☆ 49 Reviews
Shop	MP SUPERSTORE Retailer Info	$1244.99	See Site	See Site	See Site	Yes	★★★☆☆ 53 Reviews
Shop	TheCameraZone.com Retailer Info	$1249.95	No Tax	$50.00	$1299.95	See Site	★★★★☆ 62 Reviews

Figure 4. Example of Comparison Aggregation (screenshot from PriceGrabber.com).

The multi-attribute comparison capability is important because consumers use more than just price to make purchasing decisions (Smith and Brynjolfsson, 2001).

Relationship aggregation is a service that collects information from multiple relationship accounts on the user's behalf. These accounts often include various rewards program accounts, financial accounts, online bills, etc. Figure 5 gives an example of financial account aggregation.

Figure 5. Example of Relationship Aggregation (screenshot from www.ewise.com.au).

2.2 Current Status of Web Aggregation

Comparison and relationship aggregation services emerged in the U.S. a few years ago and have been widespread around the world. Comparison aggregation has been more widely adopted than relationship aggregation, but we began to see a quick spread of relationship aggregation since 2001. Despite their worldwide presence, aggregation services are primarily operated regionally with little trans-border information flow.

2.2.1 Comparison Aggregation

We surveyed comparison aggregators in the U.S., Europe, and Asia and listed a few examples in Table 1. Few of them qualify as a global aggregator according to our framework. AddAll.com, a book comparison aggregator, does qualify to be a global aggregator in that it queries bookstores from a handful countries and displays prices in more than a dozen currencies. Aggregators that have an international presence often have regional operations and aggregate information from regional sources. For example, most information sources of PriceGrabber are from North America, while PriceRunner and Kelkoo are more Europe oriented. Even though mySimon and Dealtime have a presence in both North America and Europe, there is little overlap of information sources between the two regions. More interestingly, little overlap is found in European aggregators unless a vendor has presence in multiple European countries. This indicates that information flow for comparison aggregation seldom crosses national boundaries, which qualifies their services to be regional, rather than global, according to the framework we described earlier.

Table 1. Comparison Aggregators and Their International Presence.

Aggregator	Country
mySimon	U.S., France, Germany, U.K.
DealTime[1]	U.S., U.K.
PriceGrabber	U.S., Mexico, Brazil
PriceRunner	Denmark, France, Sweden, U.K.
Kelkoo	Denmark, France, The Netherlands, Italy, Norway, Spain, Sweden, U.K.

Capabilities are also different among aggregators, and sometimes among different regional sites of the same aggregator. Most of the aggregators do not provide easy search for the same item across different countries, i.e., a user has to repeat similar search steps for each country. Only PriceGrabber allows users to see the prices for the same item in a different country by clicking the country link. We should point out that PriceGrabber aggregates one set of information sources in the U.S., then presents them to Mexican and Brazilian users with estimated country specific tax and shipping costs. This is essentially not a true cross-country search.

Most of the international aggregators maintain similar user interfaces across their sites, however, mySimon does have different ones for each country. On the U.S. site, one can search by keywords across all categories, but this is not offered on sites outside the U.S. Figure 6 shows the search interface for its French and U.K. sites, with one having no keyword search at all and the other having search by model name after you have navigated to a particular category. These differences indicate the disconnected nature of current international aggregators and pose difficulties in achieving global aggregation.
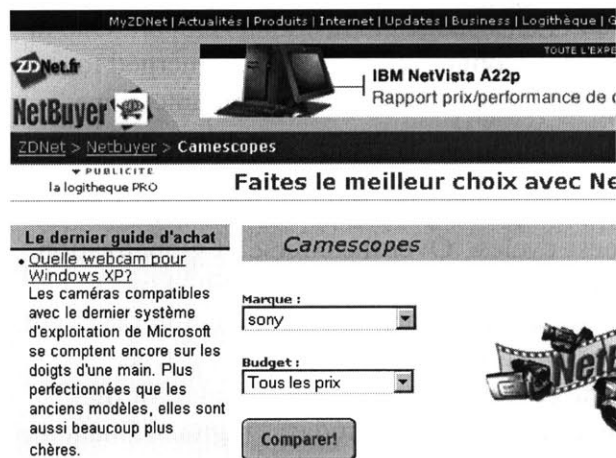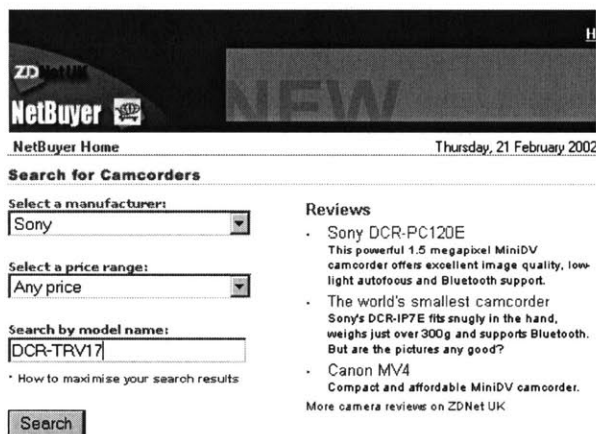
Figure 6. Different Capabilities Across Sites: (a) French Site; (b) U.K. Site.

Another differentiating factor is the comprehensiveness of coverage in terms of product categories and vendors. Most comparison aggregators cover multiple categories. Some started

---

[1] In addition to the U.S. and the U.K., Dealtime had operations in Germany and Japan in 2001.

with a single category and gradually expanded their coverage. For example, PriceScan initially focused on consumer electronics but now covers books, health, sporting goods, and a few other categories. There are still aggregators that focus only on one category, e.g., AddAll.com aggregates books only and Shopper.com focuses solely on computers/electronics. These differences may have some impact on company's revenue models. Aside from the effect of different levels of complexity, these differences may have little impact on the migration towards global aggregation.

2.2.2 Relationship Aggregation

Maxmiles is an example of a global relationship aggregator that aggregates rewards programs from more than 30 airlines from around the world and over a dozen hotel, rental car and other rewards programs from different countries. This service makes it easier for domestic as well as international travelers to manage their accounts from one single point.

Financial account aggregation, on the other hand, has remained largely as a regional service[2]. This service first started in the U.S. in 1999 by a number of technology companies. Today, many web portals and over 100 financial institutions in the U.S. are offering this service. Meanwhile, we begin to see such service or announcement for providing such service outside the U.S. The first few were either provided by pioneering U.S. technology companies or by the foreign branches of U.S. financial institutions. VerticalOne (now merged with Yodlee) was the first U.S. technology company to provide service outside the U.S. when it offered its solution in Australia to AMP Bank in September 2000, followed by CitiBank, which launched aggregation service in the U.K. in October 2000 (Celent Communications, 2002).

In addition to the diffusion of U.S. based providers, aggregation technology companies started to emerge in many countries. For example, eWise and Parker's Edge in Australia, Account One in Japan, Tarang in India, BigOnTheNet in Sigapore, and Accountunity in the U.K. emerged to provide aggregation technology and services to regional customers (Celent Communications, 2002; company websites).

The world economy is becoming more integrated as evidenced by the growing number of multinationals and recent global resonance of business cycles. Over the course of globalization, we will see increasing international flow of information, goods, capital, and people. For example, global gross capital flows quadrupled in a decade and reached $7.5 trillion in 2000[3]. It is conceivable that more people will establish relationship accounts beyond frequent flyer programs with organizations outside their home country. This should create demand for multi-category global relationship aggregation in the future. In other words, today's regional financial account aggregators will be providing global account aggregation services tomorrow.

---

[2] Merrill Lynch has global account aggregation only for its high net worth clients.
[3] Christopher Farrell, "Riding the Global Business Cycle", *Business Week* Online, March 29, 2001, citing International Monetary Fund.

2.3 Aggregation Technology

Comparison and relationship aggregation collects information from hundreds of thousands of sources that are owned by different organizations. The Internet, especially the Web, has created inter-linkages among these information sources. Although there are still aggregators relying on other technologies, such as faxes, to collect information, most of today's aggregators leverage the connectivity of the Web and use automated integration tools to collect information with, or often without, prior arrangement with organizations that own the sources. We will discuss the commonly used technologies and how aggregators use them in the following sections.

2.3.1 Screen-Scraping Technology

Screen-scraping is a technique that extracts useful information from semi-structured or text files. In an effort of turning heterogeneous web sources into relational databases, our research group developed any early demonstration of this technique in 1995. Firat *et al.* (2000) documented the evolution and other related efforts in this area. Figure 7 illustrates the mechanisms of this technique in an improved implementation called Cameleon (Firat *et al.*, 2000).
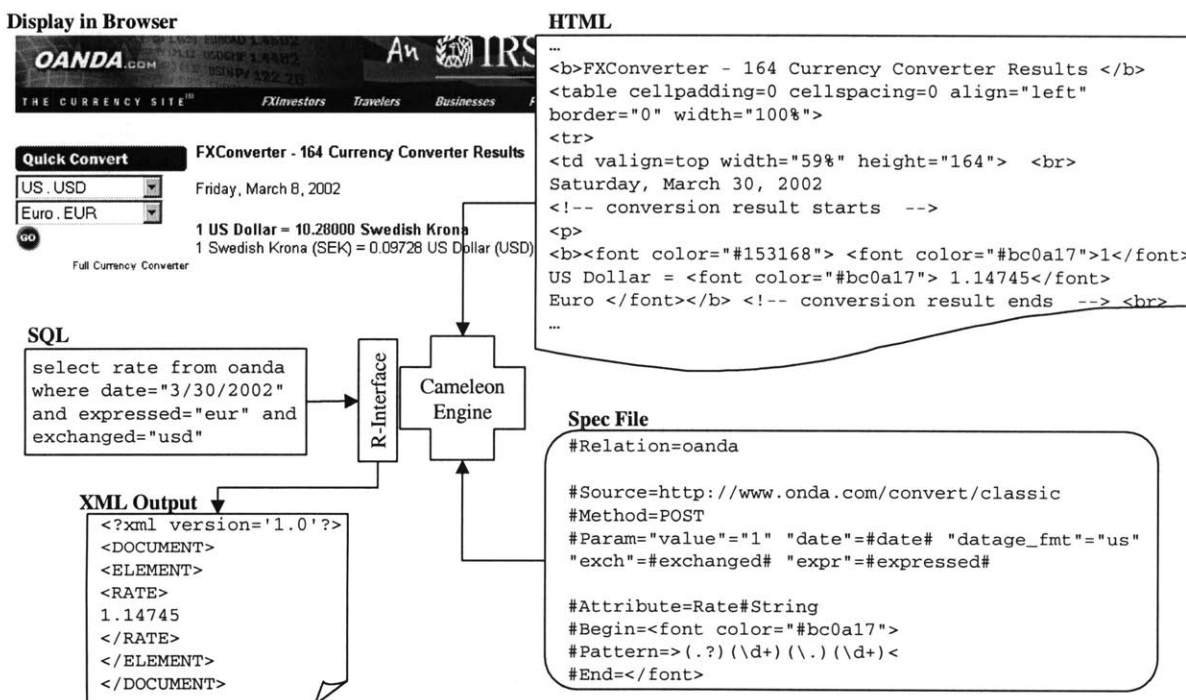
Figure 7. Screen-Scraping Technology.

The upper left of Figure 7 shows how a user would normally see a web page via a browser. To its right is an excerpt of the underlying HTML file that generates this display. The Cameloen screen-scraping engine uses a set of declarative rules predefined in a spec file shown at the lower right to sift through the HTML file and find specific information element of interest. With a relational interface, Cameleon can accept SQL queries and return results in either a table format or as an XML document. Figure 7 shows that with such a tool, an online historical currency conversion web site becomes a relational database from the user or software client's point of

view. That is, the web site is defined as a database with a simple relation called *oanda* with four attributes: *rate, date, expressed,* and *exchanged.*

Variations and improvement can be made to the technology. These include the development of efficient parallel query processing, database connectivity middleware, automated spec file generation, etc. For example, Parker's Edge in Australia developed a graphical tool to allow non-programmers to pick and choose the content to "scrape" (Celent Communciations, 2002). According mySimon, they developed a technology called Virtual Learning Agent that can be trained by mimicking human behavior to extract certain content. In addition, aggregators, especially comparison aggregators that do not need login credentials from individual users, also use information discovery technologies (i.e., web crawlers) to find online information.

This simple technique becomes a powerful information aggregation tool when it is implemented with parallel processing capability to simultaneously submit and process queries to multiple sources. Note that technically as long as an aggregator has spec files that match the sources (and is granted access to the sources by its users in the case of relationship aggregation), the tool can collect useful information without any special arrangement with the information providers. With these capabilities, information such as prices or account balances from any online vendors or service providers can be aggregated into a consolidated view.

While screen-scraping is flexible and effective even with non-cooperative sources, it is error-prone and can be costly to maintain. All spec files have to be kept updated. The overall computation model is not efficient, which hinders the scalability of the technology. These drawbacks can be overcome with standards-based information sharing, which we will discuss next.

2.3.2 Standards-based Data Exchange

Traditionally organizations that engage in information interchange agree to a common coding standard and often require dedicated connections and special applications, such as Electronic Data Interchange (EDI). The implementation and operation of such data interchange technology are expensive. The emergence of Extensible Markup Language (XML) a few years ago made it possible to develop standards that are easier to implement.

The financial industry has developed a number of XML based standards, such as Open Financial Exchange (OFX), Interactive Financial Exchange (IFX), Quicken Interchange Format (QIF), and many others. Participating organizations can publish their data using one or multiple standards via their web services. Aggregators can retrieve such data more efficiently and accurately.

Although standards-based data exchange technology solves some of the problems of screen-scraping, it has its own drawbacks. First, there are too many of them[4], which makes it difficult for organizations to follow or very costly to support multiple standards. Second, it requires agreement to a standard by multiple organizations, which requires some coordination efforts.

---

[4] See, for example, Mary Knox from Gartner Group, "Commentary: XML Standards in the Financial Industry", November 8, 2001, news.com.com/2009-1001-275607.html; and John Rymer's "Why 90% of XML Standards Will Fail" at http://www.zdnet.com/filters/printerfriendly/0,6061,2689765-84,00.html.

Sources that do not join the agreement will continue to require screen-scraping technique for aggregation.

2.3.3 Technology Adoption by Aggregators

Most comparison aggregators use automated discovery and extraction tools to collect online information, sometimes with manual assistance. PriceRunner's FAQ section explains how a typical aggregator collects information:

"Some aspects of the information-gathering are automatic thanks to our 'price agent technology' (methodical data collection), while other aspects demand manual work. Our software runs daily to crawl the web updating prices. We have an army of Pricerunners manually collecting data everyday..."

Although none of the comparison aggregators we studied disclose how "extraction" is done or if they use "screen-scraping", the techniques they use follow the general principle we described earlier.

Some aggregators seem to largely rely on manual data collection. Here is an excerpt from PriceScan's FAQ:

"We obtain pricing information from magazine ads, vendor catalogs and vendor web sites. Some vendors email or fax us daily with price changes and others have no direct contact with us."

Some provide mechanisms for vendors to directly input their prices, e.g., Shopper.com has such practice (Baye *et al.*, 2001).

For relationship aggregators an automated discovery tool is less critical because they cannot aggregate until users provide their login credentials for the organizations with which they have relationship accounts. Initially, most financial account aggregators used screen-scraping data extraction technology. With the emergence of XML based standards, about two thirds of financial accounts are aggregated using screen-scraping, with the rest using one of the industry standards (Celent Communications, 2002). To gain back some of the control over data, the financial industry is in favor of standards-based aggregation and recently has proposed a framework for next generation aggregation based on FAST: Financial Agent Secured Transactions[5].

Relational aggregation requires an initial configuration that collects login credentials to multiple accounts. Most aggregators store such information centrally on their servers. An Australian company, eWise (www.ewise.com.au), developed a technology that allows users to store their login information locally on their PC or other devices, similar to desktop aggregation software such as Quicken and Microsoft Money. This distributed solution is promising for aggregation in regions that require a secure login device to access account information.

---

[5] See the project description for FAST-based aggregation framework published on February 28, 2002 by FSTC (Financial Services Technology Consortium) at www.fstc.org. Also see "Online Banking: Banks, Vendors Plot Alternative to Screen-Scraping" by Steve Bills, American Banker Magazine, March 14, 2002.

3 GLOBAL PRICE DISPERSION

In the motivational example of section 1 the Swedish consumer can save over 40% on the product if it is purchased from the U.S. vendor. Is this a special case? The answer to this question deals with worldwide price dispersion, a topic that economists have been studying for years. In this section, we will review some empirical studies. Since comparison aggregation is an excellent tool of collecting price information, we used a number of regional aggregators to conduct a case study on price distribution of the Sony camcorder we mentioned in the motivational example. The results of the case study are presented in this section. Then we discuss contributing factors of price dispersion and point out that some of the factors can be improved using a global comparison aggregation to increase market efficiency.

3.1 Empirical Studies on Price Dispersion

When speaking of price differences around the world, people often refer to the infamous BigMac index compiled by The Economist magazine. A recent compilation shows that a $2.54 BigMac in the U.S. costs $0.83 in South Africa in contrast to $3.74 in Switzerland. As pricing is such an important topic in economics, studies on price dispersion go far beyond BigMacs.

Using OECD STAN database, Bakus et al. (1997) studied price dispersion of 28 non-overlapping categories of manufactured goods in the G7 from 1973 to 1994. They observed that prices often deviated from the mean by more than 10% over the two decades and there had been no sign that the deviation would diminish over time. Crucini et al. (1998) investigated price dispersion in Europe based on a Eurostat (the Statistical Office of the European Communities) survey conducted in mid 1980s for over 2,000 commodities, ranging from long-grained rice to man's shoes to suburban garage parking. To illustrate the huge dispersion, they compared the total costs of a 225-commodity basket fulfilled locally in a country versus buying at the lowest prices in the 13 surveyed countries: one can at least save 14% if he is in Luxemburg or save as much as nearly 75% if he is from Greece! A more recent survey in the EU (EU Economic Forum, 2001) shows that in the fresh food market "high price countries are often two times more expensive than countries with minimum prices"; even in the consumer electronics market, one country could be over 50% more expensive than another for a particular product. Different from the previous studies, which used aggregate survey data, this study relied on three consultants to collect prices of specific products, such as potatoes and SONY 4-head VCR.

Price dispersion in the U.S. online market has been studied recently. Brynjolfsson and Smith (2000) studied price dispersion for books and CDs using data from a few dozen of online and physical stores. With over 8,500 price observations collected over 15 months, they found that online store prices are 9-16% lower than prices in conventional stores, depending on whether taxes, shipping, and shopping costs are included in the price. They also observed online store prices differ by an average of 33% for books and 25% for CDs; each study consisted of 20 titles. Using an automated program, Baye et al. (2001) collected listing prices for 1000 of the best-selling consumer electronics products on Shopper.com, a comparison aggregation service provided by CNet. Based on their four million observations over an eight-month period between August 2000 and March 2001, they found that on average prices differ by about 40%. Using a

similar technique of querying comparison aggregation sites (DealTime and PriceScan), Clay et al. (2001) collected nearly 200,000 price observations for 399 books covering categories including New York Times and computer book best sellers, former best sellers and random books. They found that the inter-store price difference could be as high as 27.7% of average selling price.

It is worth noting here the tremendous benefits of comparison aggregation in helping us better understand price dispersion phenomenon. Comparing the two groups of the studies. Without comparison aggregation, one has to either use aggregate data, therefore losing specificity, or rely on costly, time consuming, and error-prone manual processes. With comparison aggregation, one can study the issue from aggregate level (e.g., book industry) down to the finest level (e.g., a specific title) with ease, accuracy, and temporal continuity, if she wishes.

We could only find one study on price dispersion in global online market. Clay and Tay (2001) collected prices of 95 English-language textbooks offered by online bookstores in four counties: US, Canada, UK, and Germany. Again, price dispersion is high. Including shipping and handling, the difference between the least expensive price in the U.S. and the least expensive price in the U.K. for a particular textbook was 42% of the U.S. list price. Interestingly, prices of a same company in two different countries are different. For a bundle of five books, one could save over 33% if purchased from the U.K. branch instead of the U.S. branch of Amazon.

3.2 Case Study – Prices of SONY Camcorder MICROMV DCR-IP5

As there have been few studies on global price dispersion of online market, we conducted an empirical study on the SONY digital camcorder in the motivational example: MICROMV DCR-IP5, which was introduced into the consumer electronics market in early 2002. Market prices for such a new product are extremely volatile; Figure 8 shows the dramatic price changes over just a few weeks in the U.S. by an aggregator, PriceScan.com. For this reason, we report statistics on global prices queried within 24 hours between March 8 and 9, 2002.

Time Period: 01/21/02 through 04/29/02
Each tick mark represents one week
Top = Highest Price, Middle = Average Price, Bottom = Lowest Price
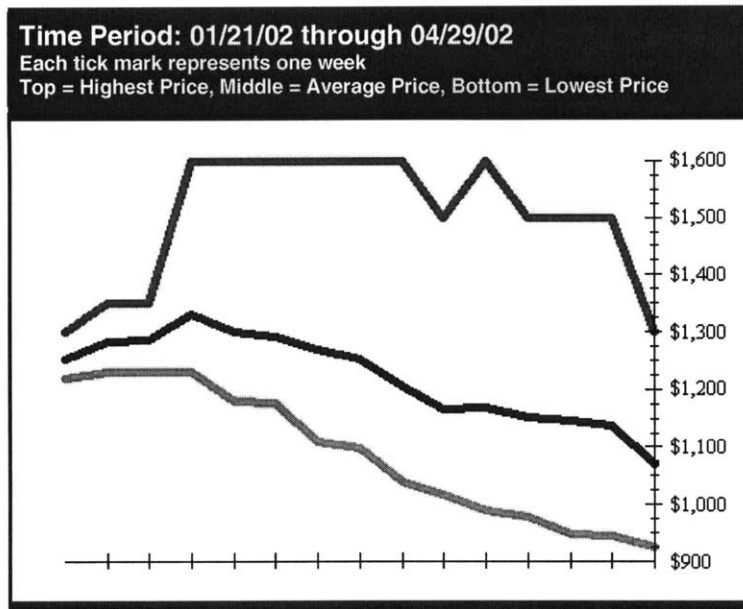
Figure 8. Volatile Prices for DCR-IP5 in the U.S. (Source: PriceScan.com).

3.2.1 Data Collection

We used a number of regional comparison aggregators to manually retrieve the prices for the product. These aggregators include BizRate, mySimon, Dealtime, Shopper, PriceRunner, PriceGrabber, Kelkoo, and Kakaku. Some of these aggregators provide services in multiple countries, as seen earlier. We report our analysis on the unique vendor/price basis within a country. That is, if multiple aggregators in a country report on the same vendor, we treat them as one observation if the prices are the same or within $1 difference. If a vendor has its online and physical stores as two entities, we treat them as two different observations even though both may charge the same price. All prices are listing prices not including shipping charges.

3.2.2 Worldwide Price Dispersion

We collected 172 observations covering US, Mexico, Brazil, Japan, and nine European countries. Because the product is not "officially" available in Mexico or Brazil, prices found in the two countries are offered by US vendors who charge 52.8% and 61.8% import taxes, respectively. Table 2 summarizes the data.

Table 2. Descriptive Statistics of Price Dispersion (in US $).

	Including Mexico, Brazil N = 172	Excluding Mexico, Brazil N = 128
Min	783	783
Max	2558	2254
Median	1700	1569
Mean	1733	1524
Standard Deviation	354	335

Let's focus on the last column of the table – statistics without the 44 observations for Mexico and Brazil. Prices range from $783 to $2254, i.e., the difference is 1.88 times of the lowest price, or the highest price is nearly three times of the lowest price! The average price, $1524, is still much higher than the official price set by SONY for the U.S. market: $1300 (from SONY.com).

Comparing with the statistics that includes Mexico and Brazil, we note that although the price range does not change much, the average price is increased from $1524 to $1733.

Figure 9 shows the histogram of prices excluding Mexico and Brazil. It is obvious that prices are highly dispersed. Most prices are within the range of $1000-2000 and they are nearly evenly distributed in this range. Prices outside this range exist at both ends.
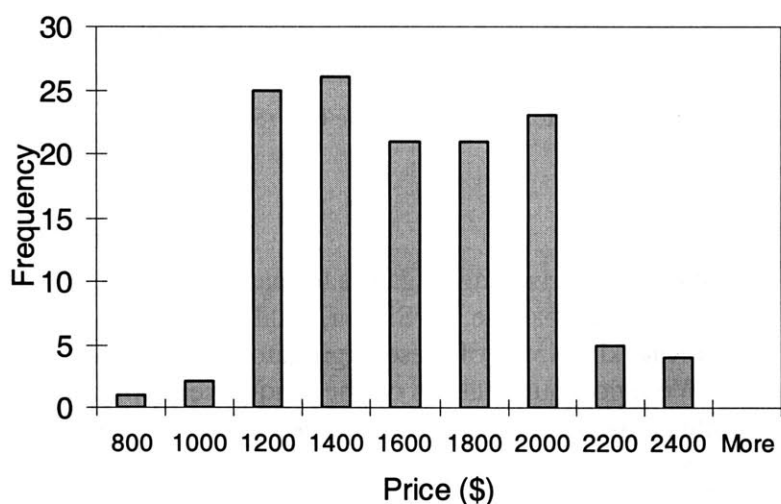
Figure 9. Price Histogram Excluding Mexico and Brazil (N=128).

If each country has a single price but different countries have different prices, can we still have worldwide price dispersion as seen in Figure 9? In other words, can each bar in this figure correspond to each country? To answer this question, we need to compare prices across countries.

3.2.3 Price Dispersion by Country

Figure 10 shows the price distribution for all 13 countries, with the number of observations for each country at the bottom. This is a box plot with each box representing 50% of price observations (i.e., the 25% and 75% quartiles) and the line within the box being the median. Lines stemming out of boxes cover all the other prices except for the extremes marked as solid circles.
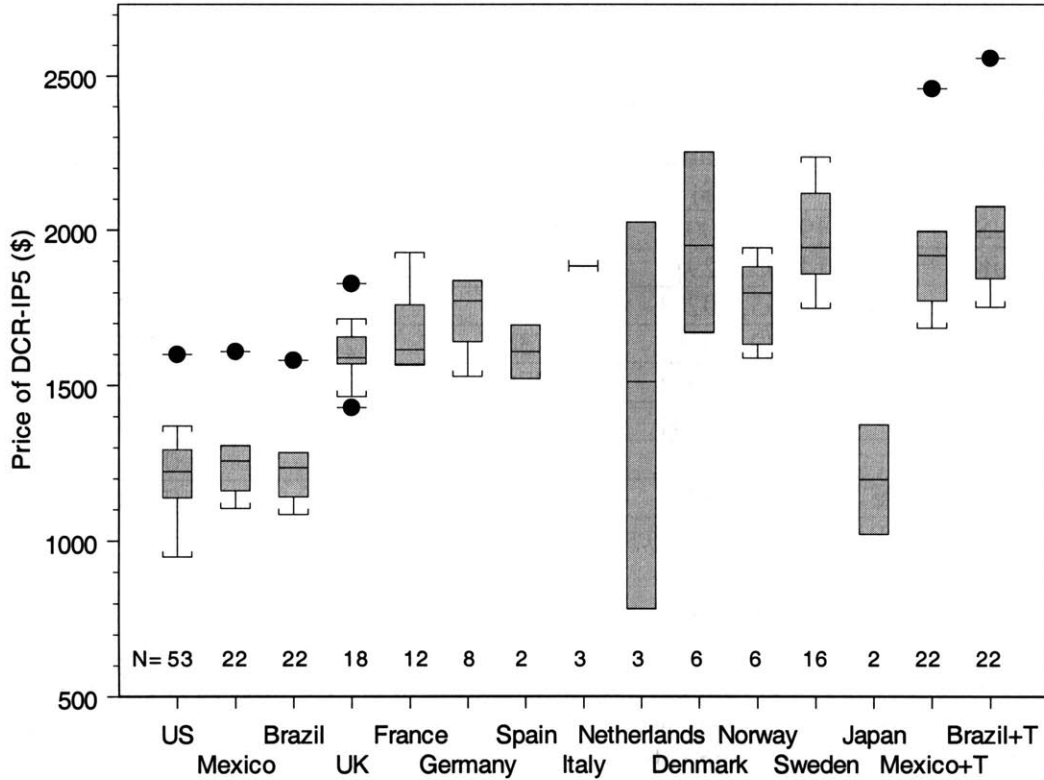
Figure 10. Price distribution in different countries.

As seen from the figure, prices are different between countries. US and Japan have the lowest price levels. Most of continental European countries, except for Italy, have medium high prices. Italy and northern European countries have the highest price levels in our observation. Note for Mexico and Brazil, the price levels are similar to the U.S. before tariff. But because the product is not "officially" available, U.S. vendors sell the product to the two countries with a high tariff, 52.8% for Mexico and 61.8% for Brazil. These high tariffs bring the price level comparable to northern European countries, with exception of one extreme observation for each country (all from one U.S. vendor, who already adjusted its price to $1099 when we rechecked on May 14, 2002). Comparing with the international book price study (Clay and Tay, 2001), which shows that the U.K. has lower book prices, here we find that the U.K. has higher prices for this camcorder than the U.S.

Another phenomenon to observe is that, except for Italy, all countries have some price dispersion for the product. In addition to the visual comparison in the figure, we present in Table 3 two commonly used dispersion measurements for countries with more than 8 observations. Range over min, or the range for short, is the range of price divided by the lowest price in the country. Coefficient of variation is the ratio of standard deviation divided by the arithmetic mean of prices in a country. The larger these two parameters, the more dispersed the prices are in the country.

Table 3. Price Dispersion of Representative Countries.

	Price Range (US $)	Mean Price (US $)	Range/Min	Coefficient of Variation
US	[949, 1600]	1203	0.686	0.092
UK	[429, 1829]	1606	0.280	0.060
France	[1565, 1929]	1677	0.233	0.078
Germany	[1529, 1929]	1734	0.202	0.073
Sweden	[1750, 2237]	1975	0.278	0.077
Denmark	[1671, 2254]	1959	0.348	0.134
Europe	[783, 2254]	1758	1.878	0.132
Worldwide	[783, 2254]	1633	1.878	0.220

Note: Worldwide statistics excludes Mexico and Brazil.

Even though U.S. has the lowest price level, it has the second highest dispersion in term of coefficient of variation, next to Denmark. This value is very well in line with the results of four million observations of 1000 consumer electronics prices in the U.S. (Baye et al., 2001). Note the last row of the table is the worldwide dispersion measurements, which is not the total of each column. Their large values indicate substantial worldwide price dispersion.

Let's look at US prices in more detail, shown in Figure 11. These 53 unique price observations do not include SonyStyle US, Sony's online store in the U.S., and major consumer electronics vendors like BestBuy and CircuitCity, which offer the produt at the same "official" price: $1299.99. We can see from the figure most prices are at or below this price level. The average price is $1203; if we exclude one extreme observation, the average price drops to $1196. That is, the average price is around $1200, which is 7.7% below the "official" price. This is close to the range of 9-16% lower prices for books and CDs on the Internet estimated by Brynjolfsson and Smith (2000). More importantly, U.S. average price is 26.3% lower than the worldwide agerage.
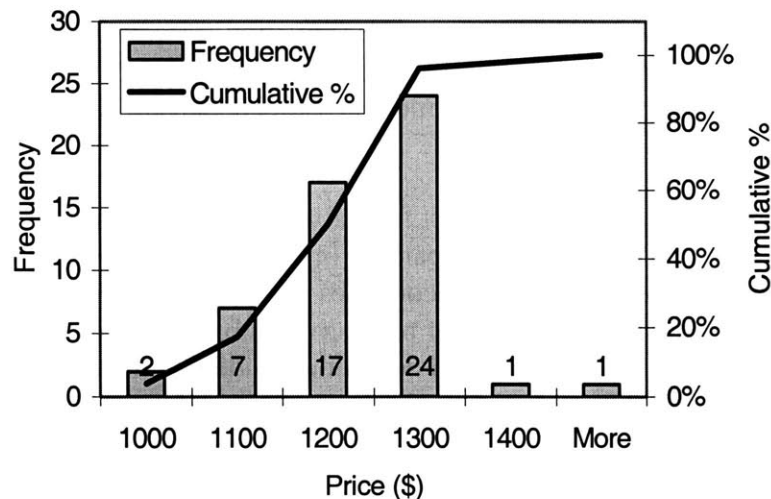
Figure 11. Price Distribution in the U.S. (N=53).

3.3 Explanations for Price Dispersion

Textbook economic theory predicts that under perfect competition (e.g., Bertrand competition) commodity prices converge to one price, the so-called Law of One Price. But real world markets have shown no evidence to support this. The price dispersion phenomenon has been explained as a violation of one of the Bertrand assumptions: product homogeneity, zero search costs, or perfectly informed consumers (Brynjolfsson and Smith, 2000). Other theories are being developed, e.g., Rauh shows that when buyers and sellers make small mistakes in their beliefs about prices a temporary equilibrium exists to result in price dispersion.

In our case study, we looked at prices for one single product. Although it does have two models for video output (i.e., PAL and NTSC), this distinction is marginally important because its MPEG recording format allows for easy processing on a PC, which does not use the video output. In addition, many TV sets support dual video standards. So this product can be regarded as homogeneous. Regional aggregators can help lower search costs, which should lead to convergence of prices (Bakos, 1997). Whether all consumers are perfectly informed about price distribution is in question. Although comparison aggregation has gained some popularity, none of the popular comparison aggregators ever make to the top 50 most visited sites in the U.S. measured by Jupiter Media Metrix.

In domestic e-business, it is possible that the three assumptions are met to some degree. So Brynjoflsson and Smith (2000) turn their explanations to heterogeneity in other vendor characteristics that are not directly observable and suggest trust in vendor brands may play a major role for price dispersion, which they further explored (Smith and Brynjoflsson, 2001). In the context of global e-business, trust certainly plays an important role (e.g., would the Swedish buyer trust the U.S. vendor even though it can potentially save him nearly 40%?). Even the basic assumptions could be violated in the setting of global e-business.

Although in terms of features the camcorder is nearly homogenous worldwide, other factors exist to result in heterogeneity. The product may be assembled in different plants that have different cost structures (e.g., plant in Malaysia vs. plant in Japan). Manufacturers often use different labeling to segment the market, e.g., using different languages for product manuals in different regions. Warranty and other post sales services are often divided into regions, e.g., camcorder sold by the U.S. vendor to the Swedish buyer may not be covered by manufacturer's warranty in Sweden if it is a U.S. only model[6].

We already know that even in domestic e-business, not all buyers use comparison aggregators to reduce search costs. In global e-business, search costs are much higher because few buyers search for worldwide prices. Even worse, most aggregators are regional, making global price comparison a difficult task. We gave a hypocritical situation in the motivational example, but in reality chances are the Swedish buyer does not know any price information in the U.S.

---

[6] When we called the vendor on May 13, 2002, the vendor's price already dropped to $949.99 for U.S. model. We were told that with $100 more we could get an international model with 1-year manufacturer warranty valid anywhere in the world. Including $100 shipping to Sweden, total price of $1150 is still over 36% cheaper (with higher price as base).

These factors, especially the lack of a global comparison tool, contribute to the worldwide price dispersion phenomenon. Although among the sampled countries, U.S. has the highest dispersion in measures of price range and coefficient of variation, the worldwide dispersion is 3.3 and 2.3 times of the U.S., respectively. The following summarizes various explanations for global price dispersion:

- Manufacturers have heterogeneous production costs around the world.
- Vendors have different pricing strategies, e.g., some may offer specials in certain parts of the world to promote sales.
- Buyers involve different search costs and have different preferences, e.g., buyers are not aware of price differences and weigh other factors more than price.
- Fluctuation of exchange rate causes price differences among countries.
- Manufacturer price control via market segmentation and other means of price discrimination, e.g., introducing product at different times or marketing product with different model numbers in different parts of the world.

Although price dispersion will not completely disappear, price transparency resulted from comparison aggregation should help mitigate dispersion and lower overall prices. This effect has been observed in the online market, e.g. average online prices are 7.7% lower than official price of the Sony camcorder and for books and CDs online prices are 9-16% lower than prices in physical stores (Brynjoflsson and Smith, 2000). Further, the U.S. average price for the camcorder is 26.3% lower than the worldwide average and the adoption rate of comparison aggregator is among the highest. Arguably, regional aggregation has helped increase competition and lower the overall price level in the U.S.

In next section, we will examine the benefits of comparison aggregation and explore potentials of global comparison aggregation. Benefits of relationship aggregation will also be examined.

4   BENEFITS OF GLOBAL AGGREGATION

Web aggregation has impacts to all involving parties, including consumers, aggregators (i.e., aggregation service providers), aggregatees (i.e., information source owners), and manufacturers (e.g., Sony in the example) (Zhu et al., 2002). An in-depth analysis of aggregator-aggregatee relationship can be found in Madnick and Siegel (2002). Our analysis here will be focused on the benefits of aggregation with emphasis on consumer benefits.

In many cases the benefits of global aggregation are a realization of those for regional aggregation on the global basis. For example, with global aggregation, increased competition spread from regional to global, benefiting worldwide consumers.

The benefits to consumers from comparison and relationship aggregation services share some commonalities. For example, both provide convenience in information gathering, which can be measured by time saving and value of information in subsequent decision-making. Providing combined aggregation services, e.g., account aggregation and interest rate comparison, can create even more values to consumers (Madnick and Siegel, 2002). However, there are still some distinctions in consumer benefits from the two types of aggregation. For purpose of convenience, we will discuss them separately in the following sections.

4.1 Methodology for Consumer Benefits Estimation

In the motivational example, the value of global aggregation is reflected in cost savings for the Swedish consumer. In the long run, such information will increase competition of global market and lower overall prices. Direct calculation of cost savings will not continue to reflect consumer benefits. We need a methodology to appropriately estimate overall consumer welfare resulted from the long-term effect of global aggregation.

Hausman and Leonard (2001) developed a technique of assessing consumer welfare from the introduction of new products, measured as difference of consumer expenditure functions at the post-introduction utility level. The total effects are decomposed into two components: welfare of the new product and welfare from lower price due to increased competition. Using this technique, Brynjolffson et al. (2001) estimated consumer surplus of obscure books available in the Internet market. In their estimation, they assumed log-linear demand function and neglected income effect. Their derivation leads to the following equations for estimating consumer surplus from the two effects:

$$CV_{New\,Pr\,oduct} = -\frac{p_1 x_1 - p^* x_0}{1+\alpha} \tag{1}$$

$$CV_{Lower\,Pr\,ice} = -\frac{(1-\phi)p_0(1-\phi\alpha)x_0 - p_0 x_0}{1+\alpha} \tag{2}$$

where p_1 and x_1 are post-introduction price and quantity; p^* is a virtual price that effectively set demand to zero, therefore $p^* x_0$ is zero; α is price elasticity for demand, and ϕ is the percentage of price decrease. Equation 1 is for consumer surplus of the new product effect; equation 2 is for the price reduction effect from increased competition.

4.2 Benefits of Global Comparison Aggregation

Global comparison aggregation will be an effective tool for global price dispersion studies. As we discussed earlier, without an aggregation tool, we would have to rely on aggregate data or time consuming and labor intensive market surveys to obtain price data. These methods are poor in terms of level of specificity (e.g., price of a particular product) and timeliness (e.g., current price rather than price of two days ago). It was an aggregation tool that made it possible to collect four million observations on 1000 products over several months of time. Using aggregation I literally collected global prices of the case study within hours. Most of the time was spent on merging the data from multiple regional aggregators to facilitate comparative analysis. Had we have a global aggregation tool, this process would have been automated and instead of having only snapshots, we can collect data over time to facilitate global longitudinal price analysis.

Timeliness is important. New products are introduced to the market on roughly an annual basis due to the effect of Moore's law or marketing strategies. Prices of a certain product therefore change frequently. Over time, some vendors may start carrying a product and likely some vendors may cease to carry that product. This dynamic nature of market is obvious in the online auction market. Timeliness of aggregation allows us to study the dynamics of the market. Table 4 summarizes the changes of vendors and their prices within two hours. Most of the vendors lowered their prices (except for one increased by $20) as reflected in the lower average price of the later sample; one vendor was not listed and four new vendors were reported in the second snapshot. Note the experiment was done on March 3, 2002, one week before the case study. The average price reported by the same aggregator dropped to $1199 at the time of the case study.

Table 4. Vendor and Price Changes within Two Hours at DealtTime, U.S.

	Snapshot	Snapshot after 2 Hours
Number of Vendors	18	21
Overlapping Vendors	17	17
Average Price ($)	1279	1251
Minimum Price ($)	1069	1049

There are several possible reasons for above observation. The lifecycle of product and supply and demand of product all have impact on pricing. Another explanation is that some vendors are possibly using market data that is readily available at the comparison aggregation sites to dynamically price their products. In fact there are already software companies that provide such tools; one of these tool providers claims that its customers will have as much as 20% of gross margin increase (Menduno, 2001).

Similar to online retailers, manufacturers can benefit from web aggregation as well. Take the Sony example. Sony may have control over wholesale prices, based on which provides a suggested retail prices in each region in the world. But Sony would have no idea how this product is priced by vendors all over the world. With global aggregation, Sony can easily study the actual pricing of its products. Together with its sales data at each region, Sony can become

more accurate in estimating consumer demand in each region and optimizes its pricing schemes for different regions.

With comparison aggregation, consumers incur lower search costs. With increased price transparency and competition, consumers can benefit from a lower price level. Among the 13 countries we compared in Figure 10, U.S. has the lowest price; part of the reason is due to increased competition.

Let's try to estimate consumer welfare due to increased competition. We observed the average price for DCR-IP5 is 7.7% below the Sony "official" price; so we use 0.077 for ϕ in our estimation. Suppose this can be generalized to all consumer electronics products. According to Thibodeaux (1998) citing consumer electronics market data, elasticity is about –2 for VCRs, –3 for portably CD players, and –3.5 for portable radios. For estimation purpose, we use –3.5 as the price elasticity for the industry. The Consumer Electronics Association (CEA, 2002) estimated the 2001 U.S. sales of consumer electronics were $93.2 billion. Using equation 2, the consumer surplus in consumer electronics section alone is about $6.4 billion in the U.S.

For all European countries we found their average price for DCR-IP5 is $1758. If global aggregation can lower their prices to the level comparable to the U.S., which is $1203, European consumers would be able to enjoy a great surplus. Assume that all consumer electronics have price differences similar to DCR-IP5 and the sales in Europe are at the same level of U.S., global aggregation could create a surplus of $16.6 billion for European consumers.

This estimate, albeit very rough with assumptions, shows significant potential benefits that a global comparison aggregation can bring to customers. Along with benefits to researchers, vendors, and manufactures, we feel there are great opportunities for global comparison aggregation services. As regional aggregators are becoming more broadly adopted, global aggregation will emerge. However, there are other barriers to international and inter-organizational flow of information and physical goods. We will discuss those obstacles in a later section. With the overcome of all the difficulties, the power of global aggregation will be unleashed.

4.3 Benefits of Relationship Aggregation

Compared with comparison aggregation, the benefits of relationship aggregation are more complicated to analyze because of uncertainties in its early stage and concerns of consumer privacy. In fact, the understanding of it has gone through some dramatic changes from the initial perceived threat to incumbents (information sources being aggregated) to a "got-to-have" service to attract customers and create opportunities for other value added services. This dynamic field has attracted much attention from the industry and popular press, however, little has been found in academic literature on this topic except for a few publications from our research group (e.g., Madnick and Siegel, 2002).

After First Union National Bank withdrew its lawsuit against Paytrust (an early bill payment aggregator in the U.S.) in early 2000, financial institutions changed their view to aggregation and began to realize its strategic value, such as cost reduction, customer retention, potentials for

cross-selling, providing financial advice, and many other post-aggregation value-added services to customers. Some industry analysts such as Celent Communications, Forrester Research, and Datamonitor have been trying to quantify these benefits to gauge the ROI (return on investment) of aggregation[7]. Uncertainties in this still relatively new service have resulted in differences in these analyses. For example, while Celent sees a possible positive ROI primarily from customer retention effect, Forrester Research finds that customer interest in the service is often overstated and a positive ROI within a 3-year time frame is unlikely[8]. Like other IT investment, aggregation benefits cannot be fully realized until it is seamlessly integrated with other services, which often requires appropriate changes in strategy and organizational processes (Madnick and Wang, 1988).

Consumer benefits from aggregation service include the convenience of a single logon and potential better return from optimized investment portfolio when aggregation is integrated into other value-added services. These benefits could be estimated using equation (1) of the methodology discussed earlier, but it is difficult to estimate the parameters in the equation. To get an approximate idea, let's assume the price for the service is equal to its marginal cost, which can be estimated with per user charge by technology providers. A number of sources have shown that per user charge could go from $10 to $99 per year. Similarly, estimated number of users often has a wide range: while Celent Communications (2002) estimates that worldwide users could reach 15 million by 2005, Datamonitor predicts that the number of users in Europe alone will be 35 million by 2005[9] and 60 million with US and Europe combined[10]. The demand elasticity is very difficult to estimate at this stage, for illustration purposes let's assume it is between 3 and 5. Simply plugging these paramters into equation (1), we find that the estimated consumer surplus from this new service to be from $37 million to nearly $3 billion. This wide range indicates great uncertainties and the need for further analysis as the service becomes more mature.

Now let's take a different approach and evaluate the consumer benefits from the time saving effect of the new service. In the U.S., an average person uses 3 to 6 financial institutions[11] and spends about 5 hours per month for managing these accounts[12]. Suppose that aggregation services can save a person 5 hours per year. Recent Bureau of Labor Statistics numbers show that there are a little over 100 million workers in the U.S., and the average hourly wage is around $8[13]. Putting this together, this 5-hour time saving is worth at least $4 billion annually to the U.S. economy. In an effort of estimating benefits of information sharing among financial institutions (account aggregation is one type of information sharing), a study sponsored by

---

[7] See, for example, October 2001 report from Celent, "Finding the ROI from Online Retail Banking Applications"; July 2001 report from Forrest Research, "Account Aggregation: The Elusive ROI"; and June 2001 report from Datamonitor, "Financial Services Aggregation 2001: Drivers, Trends and ROI".

[8] From an article by Mark Bruno, "Questions about Aggregation", US-Banker.com, October 2001.

[9] Datamonitor December 2001 report, "Account Aggregation: The Right Way Forward?"

[10] UNISYS July 30, 2001 news release citing Datamonitor's June 2001 report, "Unisys Teams up with Corillian to Bring New Consumer Service to UK Financial Service Sector".

[11] According to Catherine Allen, CEO of BITS, who presented "BITS Perspective on Account Aggregation" at the 2nd Account Aggregation Conference in April 2001.

[12] Tiffany Bass Bukow, founder of MSMoney.com, "Get Your Financial Life Organized with Less Effort", www.msmoney.com/2001/11/30/billpay.htm. In contrast, a millionaire spends 20 hours a month.

[13] See, for example, ftp://ftp.bls.gov/pub/special.requests/lf/aat39.txt.

Financial Services Roundtable (2000) shows that consumers of the FSR 90 institutions can save $17 billion and 320 million hours per year. Comparing with this study, our estimation is very conservative, yet it still reflects enormous consumer benefits.

As an example of global relationship aggregation, Maxmiles already provides services for international travelers to aggregate their rewards program accounts from different countries. As the world economy gets more integrated, more people will have relationship accounts beyond frequent flyer programs that are located outside of their home country, therefore, we will see increasing needs for global relationship aggregation. We need to evaluate this great opportunity and its benefits as it emerges and grows.

We have shown the Swedish buyer in the motivating example could save substantially using global price information. Using the economics technique introduced in the beginning of the section and a set of assumptions, we further gave a rough estimate of consumer benefits that global aggregation could potentially generate. Next, we are going to analyze technology and policy challenges for global aggregation services to realize these benefits.

5 CHALLENGES FOR GLOBAL AGGREGATION

There are various deficiencies of existing regional aggregation services, most of which are technology related. Moving from regional aggregation to global aggregation, more technological challenges arise from increased complexity. National and international policy issues will also arise in cross-country information flows. We will analyze these challenging issues and briefly discuss other trade barriers related to other types of trans-border transaction flows.

5.1 Deficiencies of Current Web Aggregation

Aside from the fact that few aggregators provide global aggregation services, there are other deficiencies in terms of comprehensiveness, timeliness, and accuracy.

5.1.1 Lack of Comprehensiveness

No aggregator has a complete coverage of all relevant information sources. This limits their effectiveness. For relationship account aggregation, in addition to the single login convenience, users can get a comprehensive view of their personal finances only if the aggregator is comprehensive itself. We tested a couple of major account aggregators in the U.S. by examining their coverage of a few randomly chosen small banks in Cambridge, Massachusetts. None of them cover East Cambridge Savings Bank and Cambridge Trust Company, even though these two banks do have online banking services. For users who have relationship accounts with organizations that are not covered by the aggregator they cannot rely on the aggregator to get a comprehensive view at all.

For comparison aggregation, lack of complete coverage may leave out good bargains the consumers are looking for and possibly be perceived to have bias in choosing vendors. In the Sony DCR-IP5 case study, 11 vendors were listed by both mySimon and DealTime, out of total 20 and 23 vendors, respectively; BizRate and Shopper.com listed only about a dozen stores for the product; in one extreme case, BizRate had only two stores when we searched the aggregator at a different time. Apparently, comparison aggregators often selectively choose what vendors to cover, as described at Kelkoo's website:

"Comparing prices from all shops on the web would be impossible because there are far too many of them. Instead, we find the best choice of shops, with the keenest prices, in each department. How? We don't just include the big brand names, which you would expect to see. Every two weeks we test our own price service, and if we can find any product cheaper outside of Kelkoo, we will contact that shop and aim to include them on Kelkoo (provided they meet our customer service guarantees)."

PriceScan, who lists vendors free of charge, clearly states their selection criteria:

"1. The vendor's prices must be competitive with what's currently in our database.
2. The vendor must honor the prices listed on PriceSCAN."

However, most comparison aggregators levy a fee to all (e.g., Shopper.com, Kelkoo) or some (e.g., BizRate, mySimon) of the listed vendors.

Another aspect of comprehensiveness is in the types of information that is useful for decision-making. Most comparison aggregators provide some, but not all, of the following information regarding a good and its vendors: base price, taxes, shipping charges, whether have international shipping, availability, expected time of delivery, whether accessories and warranty are included, vendor reputation, etc. A purchasing decision often cannot be made without further investigation into vendor's websites to find the missing pieces of information, which is time consuming.

5.1.2 Lack of Accuracy

Most aggregators cache results on their local servers and update aggregated information once a day. In fact, in order to minimize impact to servers of information sources, the Association for Payment Clearing Services (APACS, 20010) in the U.K. suggests specific time of day and data amount of screen-scraping activities be put into a Service Level Agreement between the aggregator and the data providers. This restriction limits the timeliness of information. We have seen price changes within a two-hour time frame; in the case of auction aggregation, prices could change within seconds. Daily updates are not adequate to provide timely information. In the case of global aggregation, the concept of "daily" might be not even appropriate.

When an aggregator and its information sources are out of sync, erroneous information is presented to users. Figure 12 shows an example. An aggregator reported the base price for a particular book at Powell's to be $62.00; we visited Powell's by clicking the link on the aggregator's page and found the offering price was $77.15, which is 24% higher. Three days later when we revisited the aggregator, this error had been corrected.

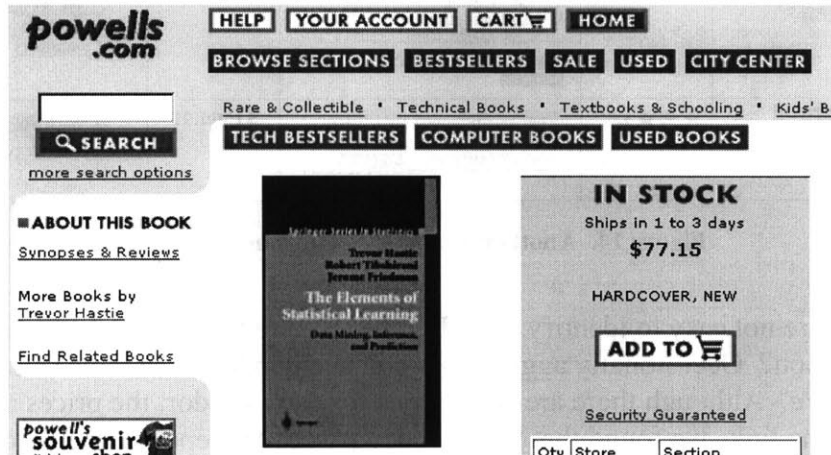| Sort by **Store Name**	Sort by **Store Rating**	Shipping ZIP: 02142 *Change*		Ground ▼
		Sort by **Base price**	+Approx Tax* & Shipping	Sort by =**Total Price** ▼
powell's books Legendary Independent Bookstore	Be one of the 1st to review this store	$62.00	Tax: check site Ship: $2.50	check site
eCampus.com EASY FAST CHEAP Huge selection of books -- UP TO 50% OFF!	★★★★☆ 6 Reviews	$70.45	Tax: check site Ship: $1.99	check site

Figure 12. Example of Errors due to Lack of Timeliness: Price Reported by Aggregator (top) vs. the Source.

Some times the errors are easy to recognize. See Figure 13, $2 is way too low to be a price for a digital camcorder. The aggregator might have mistaken the price for an accessory of the camcorder and mixed it with prices for the camcorder.


Figure 13. Example of Obvious Errors.

Figure 14 shows another type of easy-to-spot error; one of the entries has an incorrect model - DCR-IP7 is a model that is more advanced than DCR-IP5, even though we requested comparison for the DCR-IP5 model (though it is nice to get a better model with a lower price).

Sony DCR-IP5: Compare Prices
· Be one of the first to review this product
· Format Type: micro mv
· LCD Display
· See more product features

There are 22 Store Offers

Sort by: Store Name	Sort by: Store Rating	Product Description	Sort by: Price	
Tri-State 800-221-1926 Your photo specialist with 25,000 items in stock.	★★★☆☆ 22 Reviews	DCR-IP7E,MICROMV CAMCORDER,10X OPTICAL/120X DIGITAL ZOOM,CARL ZEISS,2.5 INCH LCD SCREEN,CVF,I LINK... IN STOCK!	$1579.99	▶ See it at: Tri-State Computers
CAMERA CLUB.com Everyday's a super sale day at CameraClub.com!!!	★★★☆☆ 133 Reviews	SONY DCRIP5 microMV DIGITAL VIDEO CAMCORDER , DIGITAL VIDEO CAMCORDER IN STOCK!	$1599.99	▶ See it at: CameraClub.com

Figure 14. Another Example of Obvious Errors.

Sometimes errors are not easy to identify. See Figure 15. Are there any problems with the displayed information? Occasionally aggregators do have duplicates in their listing, but this is not the problem here. Although there are two entries for each vendor, the prices are not the same. Following the links we found that both vendor's websites reported the second price. For the first vendor it could be an error (e.g., not being able to uniquely identify a vendor during an update so that outdated information is not overridden with the latest); for the second, however, the first price was a special offer to consumers who are willing to give away their email addresses.

TRONICS DEPOT Toll 1 800 549 7299 Merchant Info		sony	SONY DCR-IP5 The DCR-IP5 MICRODV camcorder is extremely stylish and ultra portable. It will record your memories... **Avail:** in stock **Shipping:** see site **LCD:** Yes	$1,069.99 (Buy)
TRONICS DEPOT Toll 1 800 549 7299 Merchant Info		sony	SONY DCR-IP5 The DCR-IP5 MICRODV camcorder is extremely stylish and ultra portable. It will record your memories... **Avail:** in stock **Shipping:** see site **LCD:** Yes	$1,269.99 (Buy)
Tri-State Camera	★★★	Sony	DCR-IP5 MICROMV MPEG2 CAMCORDER,10X OPTICAL/120X ZOOM,2.5 INCH LCD SCREEN,CARL ZEISS LENS,COLOR VIEWFINDER, **Avail:** In Stock	$1,149.99 (Buy)
Tri-State Camera	★★★	Sony	DCR-IP5,MICROMV CAMCORDER,10X OPTICAL/120X DIGITAL ZOOM,CARL ZEISS,2.5 INCH LCD SCREEN,CVF,I LINK (PAL) **Avail:** In Stock **LCD:** Yes	$1,369.99 (Buy)

Figure 15. Example of Complicated Errors.

Sources for errors come from lack of timeliness, deficiency of extraction tool, and other improper handling of information. Manual process could introduce errors, too. Baye *et al.* (2001) audited 171 prices of shopper.com, an aggregator that lets vendors to input their own prices; 96% of them were accurate within $1, meaning about 7 out of 171 prices were wrong by more than $1.

Global aggregation will be more complex; therefore it has more possible sources of error than regional aggregation. These errors are not only misleading to users, but also can cause problems for the development of more advanced features of aggregation, e.g., an agent that automatically purchases goods on user's behalf.

5.2 Technology Challenges

Global aggregation integrates information from a large number of highly autonomous sources and presents aggregated information that is meaningful and personalizable to diverse users. Finding a scalable and extensible solution is a challenging task. An implementation that improves one aspect of the required features may cause negative impacts on others. A set of data semantics related issues, which we call context and will illustrate with examples later, add another layer of complexity to technological solutions.

Take the information timeliness illustrated in Figure 12 where prices are off by 24%. To scale the system to cover a large number of sources and at the same time to be responsive to user requests, caching extracted content is a natural choice (e.g., content caching by proxy servers; Akamai Technology's services and similarly Inktomi Corporation's products for boosting performance via content caching). But caching negatively impacts timeliness and could introduce errors when cached data becomes outdated, as seen in the example.
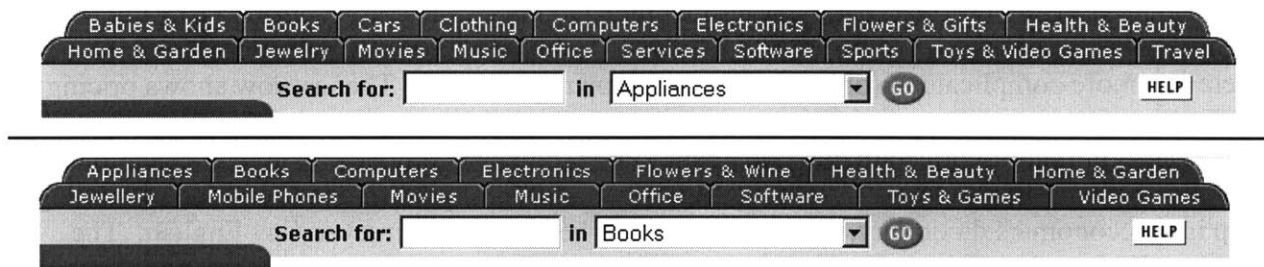
Figure 16. Different User Interfaces for Different Regions: U.S. (Top) vs. U.K. (Bottom).

Another set of contradicting features is between globalization and localization. With globalization, information comes internationally, so do the users. Differences in language, culture, and many other factors among user groups dictate diverse localized services from an aggregator. The simplest case is user interface. Figure 13 shows the interfaces of the same aggregator for U.S. and U.K. users. Both are in English, but the spelling for some words could be different (we hope readers can find one difference). In addition, the grouping of categories is different, e.g., Flowers and Gifts are grouped together on the U.S. site while Flowers and Wine are together on the U.K. site. These differences may seem to be trivial, but in fact they play a

major role in providing a comfortable and convenient interface to users for that locale. As the number of locales serviced by an aggregator increases, this can quickly become a significant effort.

The diversity in the origination and destination of information causes enormous complexity in making aggregated information meaningful and understandable. We have seen in the motivational example that currency needs to be converted to make sensible comparison for the Swedish user. Other issues exist. Let's illustrate these issues with an example of information about laptop computers from multiple sources of Sony, summarized in Table 5. We will ignore language difference in the following discussion.

Table 5. Information from Multiple Sources.

	U.S.	U.K. (in English)	U.K. (in German)
Weight	2.76 lbs	1.26 kg	1,26 kg
Thickness	1.09"	NA	NA
Price	$2,029 plus $25 shipping	1,699.00 GBP incl. VAT	1.699,00 GBP inkl. MwSt.

First to note is that not all information is available at a single source, which is actually one of the motivations for providing aggregation services. In this case the thickness information is not immediately available from U.K. sources (it is buried in a PDF document as one of the three W-H-D dimensions with a value of 27.8 mm). If an aggregator takes the information from the U.S. source and directly reports to its German users, 1.09" probably would not make much sense to the users who are familiar with metric systems for measurement. In addition to different units being used (lbs vs. kilograms, inches vs. millimeters, US dollars vs. British Pounds, etc.) there are other representational differences, such as symbols for thousands separator and decimal point. These differences have to be detected and reconciled before the aggregated information is presented to users in a format that the users can understand.

There is a more complicated problem in the data shown in Table 5. The last row shows pricing information for the product. Aside from representational differences, we note that the components going into price are quite different. Price, however simple as it appears, is in fact a complicated concept that has different meanings from different perspectives. Instead of looking it up in an economics dictionary, let's see how it can be interpreted in everyday English. The Merriam-Webster's online Collegiate dictionary provides a handful interpretations, two of which are relevant to the discussion here:

1. *the amount of money given or set as consideration for the sale of a specified thing*
2. *the cost at which something is obtained*

The distinction of the two is clear. How much an item costs for someone to acquire is often different from how much it is listed for because of other costs that are associated with this transaction, including taxes, duties if it involves international trade, shipping and handling, etc. An accurate calculation for price in the sense of "cost to acquire" could be very complicated in the context of global e-business. Calculation of VAT alone requires lots of additional information because VAT varies depending on the type of product, origination, destination, and

special treaties between regions. The variations range from 0 to 25% of the listing price in European countries (see Appendix 1). The information listed in Table 5 is a hybrid of the two concepts for *price* with some missing components. This makes aggregation and meaningful comparison difficult. McCarthy and Buvac (1994) illustrated this problem with an example of different prices of the same GE aircraft engine perceived by different organizations, such as the U.S. Air Force and U.S. Navy depending on whether the price includes spare parts, warranty, etc.

Another problem not explicitly shown in Table 5 is how the aggregators identify the same product from different regions. In the process of manually composing the Table, we noticed that the model numbers are different between laptops in the U.S. and those in Europe. We recognize their similarity (in this case identity except for the model numbers) by examining the configurations (e.g., CPU speed, hard disk capacity, weights, etc.). The fact that manufacturers often market the same product with different names in different regions makes it difficult for the aggregator to recognize their identity. This problem is best described from the following Camera example from Focuscamera.com:

"... a USA Minolta Maxxum is a Minolta Dynax overseas, the USA Canon EOS Rebel 2000 is an EOS 300 overseas, Pentax IQ Zooms are Pentax Espios overseas, etc."

Conversely, when models with different features are named the same or slightly different in different regions, aggregators sometimes cannot recognize the distinction. In the Sony DCR-IP5 case study we found that some vendors label the product as DCR-IP5E to indicate that it is an international model compatible with the PAL standards rather than the NTSC standards in the U.S. What makes it worse is that most vendors use DCR-IP5 for both the NTSC model and the PAL model. Although this does not cause big problems because of its MPEG recording format, for other types of products this could be an issue. An aggregator should not compare apples with oranges.

The preceding discussions can be summarized into three issues, with increasing complexities in the order shown below:

1. Representation – how do we represent things
2. Composition – what are the components for the thing
3. Recognition – what is the thing we are really referring to

How can we solve these problems so that users will get accurate, consistent, and meaningful aggregated information? Approaches of standardization and tight coupling through software engineering are extremely difficult to implement in the case of global aggregation where sources are highly autonomous due to the fact they are owned by many organizations located in different regions. The mediated architecture (Wiederhold, 1992), logic based context interchange approach (McCarthy and Buvac, 1994; Madnick, 1999; Goh *et al.*, 1999), and the use of ontologies (Fensel, 2001) provide a promising solution.

The mediated architecture achieves scalability by adding an abstract knowledge layer on top of data the data layer. The context interchange approach captures the afore mentioned three issues

into a context knowledge base and an ontology model; the logic programming paradigm allows for more general solutions than traditional software engineering.

Now let's suppose that technologies are developed to enable global aggregation to help people like the Swedish user search for the best buys worldwide. This will involve large amount of trans-border data flow. Can this occur? Next we examine policies that govern information flow and the obstacles from international differences.

5.3 International Policy Differences

In the U.S., account aggregation can be performed without the permission of financial institutions. In contrast, the APACS (2001) of the U.K. recommends a practice of permissible aggregation where all parties should have reached consent on how aggregation is performed before aggregation can take place. This difference in practice reflects fundamental differences in policies governing the flow of information. Policy differences among nations have little impact on domestic and regional aggregation but their impact on global aggregation is significant. Whether trans-border flow is allowed depends on whether the policies of the nations involved are in harmony. In this section, we will discuss major differences in policies that govern information flow between the U.S. and the European Union and analyze their impact on global aggregation. Our analysis will focus on two aspects that are closely related to aggregation: database protection and consumer privacy protection.

5.3.1 Database Protection

As we have seen aggregation enables information flow across organizational boundaries. The technology used allows an aggregator to extract information even from sources whose owners are not willing to share the information. But this practice has been challenged by some new development of policies regarding property rights in databases, or more generally collections of information, which poses obstacles to free flow of information across organizational boundaries. International policy differences consequently create significant frictions for information flow across national or regional boundaries.

In the U.S., the landmark Feist Publications vs. Rural Telephone Service Company (916 F.2d 718) case, where Feist extracted the listings of Rural's white pages to incorporate them into its own phone directory, made it clear that factual information is not copyrightable because of its lack of originality, the touchstone of copyright laws around the world. For this reason, copyright infringement claims were rejected in quite a few recent web content extraction related cases, e.g., eBay vs. Bidder's Edge with the defendant being an aggregator of auction sites. The court decision (NO. C-99-21200 RMW) to preliminarily enjoin Bidder's Edge from accessing eBay's site was made on the basis of trespass theory, whose applicability to the Internet has been controversial[14] (Lemley et al., 2000; Burk, 2000). This indicates that the existing legal

---

[14] Bidder's Edge also counter sued eBay. The two companies settled the case outside the court in early 2001 before the scheduled rehearing. The preliminary injunction was decided on May 24, 2000. Interestingly, in the Ticketmaster vs. Tickets.com case, the court rejected the trespass claim on March 27, 2000; after seeing eBay case, the court reworked its argument to recognize the potential validity of trespass but remained its initial decision of no

framework cannot appropriately address the issues arisen from the reuse of data on the Internet. Four bills were proposed in the U.S. congress over the past few years (Linn, 2000) to seek protection of database creators, but none of them were passed due to the difficulties in balancing the interests among different groups, primarily the investment of database creation and the values created by downstream data reuse. It is in this uncertain legal environment that aggregators emerged and evolved in the U.S.

In contrast, database owners in member countries of the European Union are protected from any significant and systematic extraction of their data under the Database Directive of 1996, which grants database owners a *sui generis* (Greek word meaning "of its own kind") right, or the "database right" in the *Copyright and Rights in Databases Regulations 1997*, the U.K. implementation of the Directive. In a recent case, the British Horsing Board (BHB) exercised their database right against William Hill, a betting service provider who extracted forthcoming racing information compiled by BHB who made it publicly accessible on its website. On February 9, 2001, the High Court of the U.K. ruled that William Hill violated BHB's database right[15].

The rational of establishing database right is to protect incentives of data creation under the "sweat of the brow" doctrine. The subsistence of database right requires "a substantial investment in obtaining, verifying or presenting the contents of the database", according to the U.K. regulation. One can argue that the effort in compiling price databases by vendors is an investment in the sales of products, not in the contents of the data. Data in relationship accounts is often generated by user-initiated activities, e.g., funds transfer generates data for new balances. The "maker of a database" is "the person who takes the initiative in obtaining, verifying or presenting the contents of a database". In this sense, it is the user who "takes the initiative" in creating account balance data. So the user should consequently be the maker of this piece of data and the collective users are the creators of the database. For these reasons, we believe data extraction activity by web aggregators should not violate database right.

This speculation needs to be further tested. The eBay case did not test this speculation because U.S. has not recognized database right. The BHB case did not either; the effort in collecting horse racing information is investment in the data and BHB is the creator of the database.

This speculation may not be contested in court at all because database owners (i.e., vendors who have online pricing information) benefit from a wide market reach by allowing aggregators to list their product offerings, therefore they are not likely to exercise their database rights even if they have such rights. In some cases, e.g., Kelkoo.com (see its website's FAQ section), vendors even need to pay a fee to the aggregators to get listed.

However, uncertainties as to whether our speculation holds exist, which could be a potential obstacle for global aggregation. If database right exists in data being extracted, international differences will affect the development of global aggregation. First, whether or not a global aggregator can extract data from sources in a particular region depends on local policy. Second,

---

injunction because of too little evidence of harm. Details can be found at http://www.gigalaw.com/library/ticketmaster-tickets-2000-03-27.html.

[15] News release of BHB on February 9, 2001. Full text can be found at www.bhb.co.uk/press_release.asp?id=255.

the way in which data is extracted may differ by region. These complicate the implementation of global aggregation and add frictions to international data flow.

4.2.2 Privacy Protection

Since aggregation is still a fairly new phenomenon and little aggregation data flow is trans-border, the differences in database protection have not caused major litigations so far. But things are different in the area of privacy protection, where the differences already have caused troubles for trans-border personal data flow. For example, American Airlines majority-owned Sabre reservation system is prohibited from transferring personal data of Swedish passengers according to the ruling of the Swedish Data Inspection Board in 1995. Even though some of the information that the Sabre system would like to collect (e.g., health condition for special care or religious belief to avoid serving religiously prohibited meal) is a crucial part of the business, this ruling was held after two unsuccessful appeals and it is pending for final hearing before Sweden's Supreme Administrative Court[16]. More recently in 1999, Microsoft had to pay $60,000 to settle charges brought by Spain because Microsoft's disclosure regarding Windows registration information is not clear to Spanish consumers (Kemp, 2001).

What went wrong? The differences in privacy protection policies have caused a "data embargo" situation, where data is prohibited to flow into a country whose privacy protection is regarded inadequate. This guideline is set in article 25 of the E.U. Data Privacy Directive, which took effect in 1998. Lack of privacy protection (in terms of level of protection, not number of privacy laws) in the U.S. has resulted in the data embargo to Sabre and the litigation against Microsoft.

What are the differences? Bennett (1997), who suggested using divergence instead of difference to emphasize the process rather than the status, proposed three aspects of the issue: policy content (the principals and goals), policy instruments (the implementation), and policy outcomes (effects on individual and organizational behavior). Nations are converging on the contents of privacy policy to a set of standards, i.e., the First Principles of information privacy that includes data quality, processing transparency, treatment of sensitive data (e.g., religious belief), and enforcement mechanisms (Reidenberg, 2000). However, divergence exists in policy implementations. The E.U. takes a centralized approach by establishing an independent agency in each member state to oversee privacy. But the U.S. has a highly decentralized system (Glancy, 2000) and largely relies on industry self-regulation. Since there is no effective measure for the outcome of privacy policies, the principle of independent oversight agency may be used as a test of "adequacy" (Bennett, 1997). This seemingly trivial implementation divergence, in fact, has deep root in national differences in the normative roles of the state, the market and the individual, e.g., liberal market-base governance vs. social-protective rights-based governance (Reidenberg, 2000). These deep-rooted differences will stay and require other mechanisms to reconcile.

As a temporary resolution, the U.S. and the E.U. reached the "Safe Harbor" agreement in 2000. But its legitimacy, scope (i.e., only industries under the jurisdiction of the FTC and DOT), and enforceability (e.g., FTC is charged to protect U.S. companies and citizens, not the Europeans)

---

[16] See a summary at www.privacyexchange.org/legal/swedishcase.html. See also Davies (1998), Chapter 7 in Swire and Litan (1998).

are in question (Reidenberg, 2001). A February 2002 review by the E.U. shows that only a trivial number of U.S. companies (129 as of December 2001) are in the agreement; even worse, fewer than 50% of the participating companies comply with all of the Safe Harbor principles[17]. The ineffectiveness of Safe Harbor indicates that the tension of privacy divergence still exists. Effective solutions may take the form of international treaty (Reidenberg, 2000), which will take a long time to form.

What does this tension mean to global aggregation? Remember in the motivational example the Swedish buyer will need to give out personal information to the U.S. vendor if he would like to buy the camcorder from the vendor. In this case, personal information has to flow from Europe to the U.S because it is unlikely that the vendor has a subsidiary or data processing center located in Europe. Is this allowed and what are the implications if it occurs? Strictly speaking, no; if it were allowed, the person should report this transfer to central authority and the U.S. vendor should delete the information as soon as the transaction is completed. If the vendor provides its own warranty, it can keep the record till the warranty expires. Bottom line is that the transaction and the facilitating systems become more complicated. To implement these complicated systems, it has been estimated that it will cost E.U. member states $15 to $20 billion (Davies, 1998) and cost U.S. institutions between $9 billion and $36 billion to avoid data embargo (Hahn, 2001).

This example deals with user-initiated data transfer. Relationship aggregation also deals with user authorized data transfer. We speculate that these types of data flows should not be subject to the restrictions of privacy regulations because the bilateral contractual agreement[18] between the user and the aggregator or the vendor should prevail. Again this speculation involves uncertainties that put global aggregation under potential risk.

5.4 Other Trade Barriers

So far we have been focusing on frictions to the flow of information. As we know, global e-business consists of other trans-border flows. To fully realize the benefits of global comparison aggregation, it also requires a smooth order fulfillment channel that includes at least the delivery of goods.

Like in the case of information flow, obstacles exist for these trans-border flows. In its 2002 annual foreign trade barriers report, the Office of the U.S. Trade Reprehensive (USTR, 2002) identified 10 categories of trade barriers, including importing policies, standards, testing, labeling, certification, etc., which may have certain impact on global e-business. For example, customs reporting and inspection may cause delay in the delivery of goods. A product labeled for sale in one region may need a separate approval in another region. Products from overseas may not be compatible with local standards (e.g., voltage and electric plug of the camcorder designed for the U.S. may be different for other regions).

---

[17] Commission Staff Working Paper released on 2/13/2002.
[18] A relationship aggregation user needs to sign an agreement, often by clicking an "I Agree" button when signing up for aggregation services, which grants the service provider the right of collecting personal information on the user's behalf.

Another type of flow, which we did not include in the framework, is people flow. Although the electronic connectivity greatly enhances interactions between trading partners, which can potentially reduce the needs for traveling, face-to-face meetings cannot be completely replaced due to complexities of some transactions and other factors, such as trust. Enormous obstacles for mobility of people still exist. For example, one can easily buy an airline ticket online with the lowest fare to a foreign country, but the person may not be able to use the ticket because of visa problems. This often happens to travelers from developing countries to the developed world, where visa denials occur for various reasons. The analysis of this issue is out the scope of this study, largely due to its complexity.

6   CONCLUSIONS AND FUTURE RESEARCH

6.1 Conclusions

We introduced a framework to categorize e-business according to organizational and national trans-border transaction flows. This framework captures a wide range of e-business practices and their related issues.   We have focused our analysis on technology and policy issues for global e-business that has both types of trans-border flows using global aggregation as an example.

Despite the global presence of comparison aggregation and the worldwide spread of relationship aggregation, most of the services are offered regionally, not globally. Lack of global information can result in inefficiency in the global market. Our price dispersion case study shows that the worldwide prices for DCR-IP5, a Sony digital camcorder, can differ by nearly three times. A global aggregator can close the information gap and bring benefits to researchers, manufacturers, and consumers. We discussed these benefits and showed how to estimate consumer surplus from a more competitive market.

The major benefits of relationship aggregation are twofold: time savings and potentials for optimization and other value-added services. While this service is still in its early stage, it has been steadily growing within the U.S. and started to spread to the rest of the world. As the world economy gets more tightly integrated, more people will have business relationships with international service providers, creating great opportunity for global relationship aggregation.

While we see a promising future for global aggregation, we also realize the existence of many challenges and obstacles. We identified various deficiencies of existing aggregation services and suggested solutions that involve technologies such as the mediated architecture, use of ontology and Context Interchange techniques. The capabilities of these technologies combined will be demonstrated in the proposed global aggregator prototype discussed shortly for future research.

Another set of obstacles for global e-business come from policies that govern the transaction flows. Directly related to global aggregation are information policies for international data flow. We focused on policies of property rights in databases and privacy protection. We speculated that impact on global aggregation would be limited, but uncertainties and international differences pose frictions to trans-border information flow that is crucial to global aggregation.

6.2 Future Research

Further research is needed for the issues that we are not certain, such as the determination of data ownership and issues related to user initiated transactions. We have focused primarily on information flow; other types of transaction flows deserve further analysis. Issues for intra-organizational transactions also should be further studied, which may include the effects of aggregation on changes in control, incentives, and firm boundaries.

In the near term, we plan to develop a prototype of global comparison aggregation to demonstrate the capabilities of data extraction and context mediation technologies developed

from the COIN project. It can be used as a research tool in further investigation of global price dispersion. We also hope such a demonstration will trigger some entrepreneurial experimentations for global aggregation as the Universal Financial Aggregation demo from our lab did for commercial development of financial account aggregation. As an example of its usage, global comparison aggregation can be a useful tool for tourism agencies to make shopping suggestions to their international tourists. A manufacturer can also use it to find out the actual retail prices of their products around world, with which they can better assess demand and set wholesale and suggested retail prices.

In the long run, we will enhance the core technology for large-scale meaningful system integration to facilitate global aggregation and other global e-business applications. We will apply both legal and economics frameworks to analyze related global policy issues. By combining research in technology, policy, and especially their interactions, we hope to gain a comprehensive understanding of the issues in global e-business.

REFERENCES

APACS (2001) "Best Practice Aggregation Guidelines", December 19, 2001.

Backus, D.K., Crucini, M.J., Telmer, C.I. (1997) "International Price Dispersion in the G7", GSIA Working Paper, Carnegie Mellon University.

Bakos, J.Y. (1997) "Reducing Buyer Search Costs: Implications for Electronic Market Places", *Management Science*, **43**(12), 1613-1630.

Baye, M.R., Morgan, J., Scholten, P. (2001) "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site", Working Paper, Indiana University.

Bennet, C. (1997) " Convergence Revisited: Toward a Global Policy for the Protection of Personal Data?" in Agre, P.E., Rotenberg, M, Eds. *Technology and Privacy: The New Landscape*, The MIT Press, 99-124.

Brynjolfsson, E., Smith, M.D. (2000) "Frictionless Commerce? A comparison of the Internet and Conventional Retailers", *Management Science*, **46**(4), 563-585.

Brynjolfsson, E., Smith, M.D., Hu, Y. (2001) "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety", WISE 2001, New Orleans.

Burk, D.L. (2000) "The Trouble with Trespass", *Journal of Small and Emerging Business Law*, **4**(1).

Celent Communications (2002) "The State of Account Aggregation Industry".

Clay, K., Krishnan, R., Wolff, E. (2001) "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry", *Journal of Industrial Economics*, **49**(4), 521-539.

Clay, K., Tay, C.H. (2001) "Cross-Country Price Differentials in the Online Textbook Market", Working Paper, Carnegie Mellon University.

Crucini, M.J., Telmer, C.I., Zachariades, M. (1998) "Dispersion in International Prices", Working Paper.

Davies, S. (1998) "Europe to U.S.: No Privacy, No Trade", Wired.com, May 1998.

EU Economic Reform (2001) "Price Dispersion in the Internal Market"

The Financial Services Roundtable (2000) "Customer Benefits form Current Information Sharing by Financial Services Companies", December 2000.

Fensel, D. (2001) Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Springer.

Firat, A., Madnick, S., Siegel, M. (2000) "The Cameleon Web Wrapper Engine", *Proceedings of the Workshop on Technologies for E-Services*, September 14-15, 2000, Cairo, Egypt.

Glancy, D. (2000) "At the Intersection of Visible and Invisible Worlds: United States Privacy law and the Internet", 16 *Santa Clara Computer and High Technology Law Journal 357.*

Goh, C.H., Bressan, S., Madnick, S., Siegel, S. (1999) "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information", *ACM Transactions on Information Systems,* **17**(3), 270-293

Hahn, R.W. (2001) "An Assessment of the Costs of Proposed Online Privacy Legislation", available at http://www.actonline.org/pubs/HahnStudy.pdf.

Hausman, J., Leonard, G. (2002) "The Competitive Effects of a New Product Introduction: A Case Study", *Journal of Industrial Economics*, forthcoming.

Kemp, T. (2001) "Privacy Rules Cross The Pond", InternetWeek.com, July 16, 2001.

Linn, A. (2000) "History of Database Protection: Legal Issues of Concern to the Scientific Commnity", ICSU/CODATA report, http://globalchange.gov/policies/linn.html.

Madnick, S.E. (1999) "Metadata Jones and the Tower of Babel: The Challenge of Large-Scale Semantic Heterogeneity", *Proceedings of the 1999 IEEE Meta-Data Conference*, 1-13, April 6-7, 1999.

Madnick, S.E., Siegel, M.D. (2002) "Seizing the Opportunity: Exploiting Web Aggregation", *MISQ Executive*, **1**(1), March 2002, 1-12.

Madnick, S.E., Wang, R. (1988) "A Framework of Composite Information Systems from Strategic Advantage", *Proceedings of the 1988 Hawaii International Conference on System Sciences*, 35-43, Hawaii.

McCarthy, J. and S. Buvac (1994) "Formalizing Context (Expanded Notes)", Stanford University.

Menduno, M. (2001) "Priced to Perfection", Business 2.0, February 2001.

Rauh, M.T. (2001) "Heterogeneous Beliefs, Price Dispersion, and Welfare-Improving Price Controls", *Economic Theory*, **18**, 577-603.

Reidenberg, J.R. (2000) "Resolving Conflicting International Data Privacy Rules in Cyberspace", *Stanford Law Review*, **52**, 1315-1376.

Reidenberg, J.R. (2001) "E-Commerce and Trans-Atlantic Privacy", *Houston Law Review*, 717-749.

Smith, M.D., Brynjolfsson, E. (2001) "Consumer Decision-making at an Internet Shopbot: Brand Still Matters", *Journal of Industrial Economics*, **49**(4), 541-558.

Swire, P.P, Litan, R.E. (1998) *None of your Business: World Data Flows, Electronic Commerce, and the European Privacy Directive*, The Brookings Institution.

Thibodeaux, T. (1998) "Pricing Plots Products' Destinies", Consumer Electronics Vision, July/August 1998.

USTR (2002) "The 2002 National Trade Estimate on Foreign Trade Barriers", Office of the U.S. Trade Representatives, www.ustr.gov/reports/nte/2002.

Van Alstyne, M., Brynjolfsson, E., Madnick, S. (1995) "Why not one big database?" *Decision Support Systems*, **15**(4), 267-284.

Westland, J.C., Clark, T.H.K. (1999) *Global Electronic Commerce: Theory and Case Studies*, The MIT Press.

Wiederhold, G. (1992). "Mediators in the Architecture of Future Information Systems." *Computer*, **25**(3), 38-49.

Wolcott, P., L. Press, W. McHenry, S. Goodman and W. Foster (2001) "A Framework for Assessing the Global Diffusion of the Internet." *Journal of the Association for Information Systems*, **2**, article 6.

Zhu, H., Siegel, M.D., Madnick, S.E (2001) "Information Aggregation – A Value-added E-Service", *Proceedings of the International Conference on Technology, Policy, and Innovation: Critical Infrastructures*, The Hague, The Netherlands, June 26-29, 2001.

Zhu, H., Madnick, S.E., Siegel, M.D. (2002) "The Interplay of Web Aggregation and Information Policy", Working Paper.

APPENDIX

Appendix 1. VAT in EU Countries (Source: Amazon.com U.K.)

Country	Books	Audio-books	Music	DVD & Video	Electronics	Software and PC & Video Games	Mixed Media	Gift Wrap	Delivery
Austria	10%	20%	20%	20%	-	20%	20%	20%	VAT rates vary, depending on contents
Belgium	6%	21%	21%	21%	-	21%	21%	21%	VAT rates vary, depending on contents
Denmark	25%	25%	25%	25%	-	25%	25%	25%	25%
Finland	8%	22%	22%	22%	-	22%	22%	22%	VAT rates vary, depending on contents
France	5.5%	19.6%	19.6%	19.6%	-	19.6%	19.6%	19.6%	VAT rates vary, depending on contents
Germany	7%	16%	16%	16%	-	16%	16%	16%	VAT rates vary, depending on contents
Great Britain and Northern Ireland	0%	17.5%	17.5%	17.5%	17.5%	17.5%	17.5%	17.5%	VAT rates vary, depending on contents
Greece	0%	17.5%	17.5%	17.5%	-	17.5%	17.5%	17.5%	VAT rates vary, depending on contents
Ireland (Eire)	0%	21%	21%	21%	21%	21%	21%	21%	VAT rates vary, depending on contents
Italy	4%	20%	20%	20%	-	20%	20%	20%	VAT rates vary, depending on contents
Luxembourg	3%	15%	15%	15%	-	15%	15%	15%	VAT rates vary, depending on contents
Netherlands	6%	19%	19%	19%	-	19%	19%	19%	VAT rates vary, depending on contents
Portugal	5%	17%	17%	17%	-	17%	17%	17%	VAT rates vary, depending on contents
Spain	4%	16%	16%	16%	-	16%	16%	16%	VAT rates vary, depending on contents
Sweden	6%	25%	25%	25%	-	25%	25%	25%	VAT rates vary, depending on contents