Legal Challenges and Strategies for Comparison Shopping and Data Reuse

Hongwei Zhu
Stuart E. Madnick

Working Paper CISL# 2010-05

September 2010

Composite Information Systems Laboratory (CISL)
Sloan School of Management, Room E62-422
Massachusetts Institute of Technology
Cambridge, MA 02142

LEGAL CHALLENGES AND STRATEGIES FOR COMPARISON SHOPPING AND DATA REUSE

Hongwei Zhu
College of Business and Public Administration
Old Dominion University
Norfolk, VA 23529, USA
hzhu@odu.edu

Stuart E. Madnick
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02142, USA
smadnick@mit.edu

ABSTRACT

New technologies have been continuously emerging to enable effective reuse of an ever-growing amount of data on the Web. Innovative firms can leverage the available technologies and data to provide useful services. Comparison-shopping services are an example of reusing existing data to make bargain-finding easier. Certain reuses have caused conflicts with the firms whose data has been reused. Countries in the European Union have implemented the Database Directive to provide legal protection for database creators, but the impact and the interpretation of the new law are unclear and still evolving. Lawmakers in the U.S. have not decided on a policy concerning database protection and data reuse. Both data creating and data reusing firms need to develop strategies to operate effectively in this uncertain environment. Comparison-shopping and other data reuse services face similar legal and strategic challenges. Thus we address these challenges in the broader data reuse context. We use economic reasoning to formulate strategies in anticipation of the likely policy choices and interpretations of existing legislation. Both data creating firms and data reusing firms should focus on innovative ways of using or reusing data to create differentiated products and services. For firms that gather data from multiple sources, they can also use the insights gained from integrated data to provide other value-added services.

Keywords: comparison shopping, data strategy, data reuse, database protection, innovation

An increasing number of innovative applications have been developed to take advantage of the large amount of data on the web. These applications, such as comparison-shopping bots and various "mashups", make relevant data from multiple websites easily accessible at a single website. For example, Bidder's Edge allowed the user to search and compare auction data on over five million items from more than 100 auction sites, such as eBay and many others, as easily as the user could search one auction site. Similarly, Kayak lets the user compare airfares by searching more than 140 travel sites, such as Expedia and Travelocity, to find the best fares available. With these applications, no more is it necessary to visit individually multiple sites and manually "integrate" the data from these sites. These applications do all that for the user automatically. They extract and reuse relevant web data, often in very innovative ways, to make the data more valuable to the user.

As Tim Berners-Lee, inventor of the Web, once said, "the exciting thing is serendipitous reuse of data: one person puts data up there for one thing, and another person uses it another way" [Frauenfelder 2004]. It is with that view that computer scientists and technologists have been developing various techniques to enable easy data reuse on the web. Comparison-shopping bots [Wan 2009b], also called shopbots or comparison aggregators, are typical examples of reusing Web data to provide value-added services.

Despite the enthusiasm about "serendipitous reuse of data" on the web among innovators and users of such applications, the firms whose data has been reused often have tried hard to control who can use "their data" and how it is reused. eBay sued Bidder's Edge [eBay, Inc. v. Bidder's Edge, Inc. 2000]. Expedia sent a "cease and desist" letter to Kayak [Johnson 2004]. Bidder's Edge stopped searching eBay per a preliminary injunction and later ceased operation. Kayak now does not incorporate Expedia and four other sites' data in the search result. Searches to these

sites are performed in separate popup windows, thus it now takes a more cumbersome manual process to compare their airfares with those that are automatically extracted and organized by Kayak.

There have been discussions on how comparison-shopping services might affect retailers [Borzo 2004; Fowler 2009], who fear to face increased price competition and reduced profit. Empirical studies show that price dispersion is persistent on the Web and there are other factors such as brands [Smith & Brynjolfsson 2001], product category [Lindsey-Mllikin & Grewal 2006], and maturity of the Internet and electronic market [Bock et al. 2007] that affect pricing. As discussed in [Wan & Liu 2008; Madnick & Siegel 2002], there have been different reactions to comparison-shopping and other data reuse services. However, legal challenges to these services still remain and there are other issues to be addressed. Do firms like eBay and Expedia actually own the data on their websites so that they control who can reuse the data and how it is reused? From the two examples above, it may seem so. But the reality is less clear and more complex. How can someone own the data that is openly accessible via the web and control its uses? What is the strategy for those who think they own the data? What is the strategy for those who want to reuse data on the web? We will address these important research questions in this paper using legal case analysis, business case analysis, and economic reasoning. Since comparison-shopping and other data reuse services face similar issues, we will address them in the broader context of data reuse.

1. Comparison Shopping and Data Reuse

As a shopbot gathers prices and other data from many Web sites to facilitate comparison, it essentially reuses data from these sites. There are other services that also reuse data from multiple Web sources. For example, account aggregators such as Yodlee, Mint.com, and MaxMiles (now InsiderFlyer.com) can act as a user agent to retrieve data from various accounts of the user and provide an integrated view of these accounts with a single sign-on. Such data reuse can be with or without prior arrangement with the data sources. For convenience, let us call those who reuse data from other sources the *data reusers* and those whose data is reused the *data creators*. Over the years, both the technology for data reuse and the dynamics between creators and reusers have evolved.

The first shopbot was BargainFinder [Krulwich 1996], created in 1995 using screen-scraping technology that exploits patterns in semi-structured Web pages. The advantage of the technology is that it does not need cooperation of the Web site owner. But despite years of research and development in data extraction, the technology requires significant manual efforts and is not completely reliable [Chang et al. 2006; Firat et al. 2001].

Initially, certain data creators were protective of their data and tried both technical and legal means to protect their data from reusers. For example, eBay tried to block queries sent from Bidder's Edge identified by the computer's IP address. But Bidder's Edge circumvented this obstacle by rotating a pool of IP addresses to query eBay's site. Certain other creators viewed comparison and other data reuse services as a means to reach customers so they embraced these services. The sentiments of certain protective creators later changed. For example, financial institutions that initially resisted account aggregation later offered the service to their customers. In 2004, the Financial Services Roundtable, a consortium of 100 of the largest financial institutions in the U.S., issued a guideline for banks to provide standard-based data feed to make account aggregation more secure, accurate, and efficient [BITS 2004]. In 2005, eBay acquired shopbot shopping.com (which owns dealtime.com) for $634 million to provide their sellers a tool for selling to more buyers [eBay 2005].

The advent of Web services technology provides a more efficient alternative to screen scraping [Fasli 2006; Li et al. 2009]. The technology allows a data reuser to obtain data by remotely executing programs of data creators that provide Web services. For example, Amazon provides Web services to allow others to obtain price data in its databases for purposes such as price comparison. This technology requires cooperation between data creators and data reusers.

Even though data creators can benefit from data reuse and some of them may have become less restrictive of reusing their data, the legal and strategic issues mentioned earlier still remain unresolved. As more data becomes available on the Web and new technologies continue to emerge to further enhance the capability of data reuse, these issues become even more acute for both data creators and data reusers.

2. Legal Challenges

2.1. Lawsuits and Legal Threats

As mentioned earlier, the growth of comparison-shopping and other services involving data reuse was not without legal challenges. A number of lawsuits were filed by data creators against data reusers. Extensive discussions of these lawsuits from legal perspectives can be found in papers such as [Lipton 2003; Gibson 2005] and in law review journals. A discussion of legal cases from an agent computing perspective can be found in [Wan & Liu 2008]. We will not repeat discussions of these legal cases here. Instead, we briefly discuss the tactics used by data creators: lawsuits and legal threats.

Lawsuits. It is interesting to observe four different outcomes of lawsuits between a data creator and a data reuser: (1) partial loss; (2) withdrawal; (3) preliminary (and controversial) victory; and (4) victory due to intimidation.

(1) In the case of Ticketmaster v. Tickets.com [2003], both companies sold event and entertainment tickets to consumers. On its website, Tickets.com provided hyperlinks to Ticketmaster web pages for tickets not available at Tickets.com. This practice is generally known as deep-linking. To enable this feature, Tickets.com also employed data extraction techniques to extract event information from Ticketmaster. The action of using computer agent to repeatedly visit Ticketmaster's computer led to the allegation of "trespass to chattels", which is based on tort law for real properties where the infringing party has interfered with another person's property. Initially, the court dismissed the trespass claim because "it is hard to see how entering a publicly available web site could be called a trespass, since all are invited to enter." [Ticketmaster v. Tickets, 2000a]. After seeing the decision of the eBay case, the court withdrew its initial ruling and recognized the potential validity of trespass to chattels, but it did not issue a preliminary injunction because no evidence of harm was found [Ticketmaster v. Tickets, 2000b]. In 2003, the court reaffirmed that tangible interference must be present to attract trespass to chattels. The court also dismissed copyright infringement claims that were based on temporary "copying" of a Web page for data extraction purposes and deep-linking to particular pages. The case left open contract issues for potential future litigation.

(2) In the case of First Union Corp v. Secure Commerce Services, Inc. [1999], the defendant was an account aggregator who operated Paytrust that allowed customers to pay bills and view balances of multiple accounts. But later First Union withdrew its lawsuit when it decided that it was necessary for it to also offer account aggregation services.

(3) In the case of eBay v. Bidder's Edge [2000], eBay made numerous claims ranging from trespass to chattels, to false advertising and to unjust enrichment. The queries sent by Bidder's Edge represented 1.53% of traffic on eBay's servers. The court issued a preliminary injunction based on the reasoning that if such reuse were allowed, it could eventually interfere with eBay's service. This is different from the opinion of the court that decided the Ticketmaster case, which insisted that material harm must exist to apply trespass law. Furthermore, there has been debate about the applicability of trespass law in cyberspace [Hunter 2003; O'Rourke 2001].

(4) In the case of mySimon v. Priceman [2000], Priceman reused data from other shopbots including mySimon to provide a more comprehensive comparison. Being a very small company with little financial resources, Priceman ceased operation in fear of costly litigation.

Legal Threats. Like Priceman, many innovative data reusers have limited financial resources. Sometimes, without actually filing a lawsuit, a creator can deliver sufficient threat using a cease and desist letter to stop a reuser from reusing its data. We mentioned earlier that this tactic worked for Expedia to stop Kayak. There was a similar case between Intershipper, which provides shipping price comparison service, and one of the carriers whose data was reused by Intershipper [Madnick & Seigel 2002]. A letter from the carrier's corporate counsel was enough for Intershipper to stop reusing its data. However, the carrier changed its mind six months later and Intershipper readmitted the carrier into its price comparison listing.

Regardless of tactics and outcomes observed so far, it is clear that there have been substantial legal challenges to data reusers. The preceding discussions also suggest that it is unclear what legal principles may be applicable to data reuse on the Web. In recent years, the European Union and the U.S. have taken different approaches to legislation for data reuse. We will discuss these approaches next.

2.2. Legislation and Uncertainties

It is without any doubt that data is one of the important assets of a firm. The firm "owns" the data when it can fully control who can access the data and how the data should be used. But when the firm makes the data accessible to the general public on the web, its "ownership" to the data will be determined by a set of legal rights generally known as intellectual property rights. Most jurisdictions recognize four forms of intellectual property rights: trade secret, trademark, patent, and copyright. Here we consider databases that contain mundane facts, such as airfares, and have been made publicly accessible, therefore the database creators cannot claim trade secret, trademark, or patent protection. What about copyright?

No Copyright for Factual Data. Both eBay and Ticketmaster alleged that their copyright was infringed in their lawsuits. The courts rejected the allegation based on the principle that copyright only protects the original selection and arrangement of factual data, but not the data itself or the effort in compiling the facts. This principle was established in 1991 by the U.S. Supreme Court in the appeal case *Feist Publications* v. *Rural Telephone Co*. [1991]. Feist reused 1,309 of approximately 7,700 Rural's White Pages listings in creating its phone book covering a large area that included the service area of Rural. The Supreme Court decided that copyright should be used to reward originality and originality requires "some minimal degree of creativity". The decision explicitly rejected the so-

called "sweat of the brow" doctrine that attempts to use copyright to reward "the hard work that went into compiling facts."

Although there are differences in the originality requirement of copyright law internationally [Zhu & Madnick 2009], it is quite uniform that one cannot claim copyright protection for individual entries of facts stored in a database. Thus copyright cannot be used to prevent others from reusing the individual facts of a database when the database is openly accessible via the web. Furthermore, copyright has been evolving and initiatives such as Creative Commons advocate copyright that is less restrictive and more conducive to creative reuse of digital contents [Guterman 2009].

Database right in the EU. The European Union (EU) introduced the Database Directive in 1996, requiring member states to implement laws to grant database creators a *sui generis* right, which is called the *database right* in the U.K. This right lets the database creator prevent unauthorized extraction of the whole, a substantial part of, or systematic extraction of an insubstantial part of database contents. Under this new law, the British Horseracing Board (BHB) sued William Hill, an online betting company who, on its website, reused the lists of upcoming horse races created by BHB. William Hill was found to have violated BHB's database right initially in 2001 [BHB v. William Hill 2001]. But the decision was reversed after the European Court of Justice (ECJ) issued its opinion in 2005 [ECJ C338/02]. The protection afforded by the *sui generis* right is much narrower than that it had been expected earlier. It only prevents the reuse whose purpose is to reconstitute a substantial part of the database contents.

In a more recent case [ECJ C-340/07] between Directmedia Publishing GmbH and Albert-Ludwigs-Universität Freiburg (University), Directmedia consulted the list of 1,100 important German poems published by the University on the Internet and selected 856 of them to compile a CD-Rom containing 1,000 German poems. In its judgment issued in 2008, The ECJ determined that an on-screen consultation followed by reusing certain contents of a database constitutes an "extraction" as defined in the Database Directive. Contrary to the view expressed in [Nettleton 2009], the judgment did not intend to expand the scope of the *sui generis* right.

Following the EU's introduction of the Database Directive in 1996, the U.S. Congress considered six bills, all of which failed to pass into law. These bills, as summarized in Table 1, proposed varying degrees of protection for database creators. As far as these bills are concerned, firms (e.g., eBay and Expedia) that compile collections of data are database *creators*, and firms (e.g., Bidder's Edge and Kayak) that use the creators' data are *reusers*.

Table 1: Proposed Bills for Database Protection in the U.S.

Year	Legislation
1996	HR 3531: Database Investment and Intellectual Property Piracy Act. Similar to EU Database Directive.
1998	HR 2652: Collections of Information Antipiracy Act. It offers the database creators criminal or civil remedies if the reuser causes or has the potential to cause harm to the creator.
1999	HR 354: Collections of Information Antipiracy Act. Similar to HR 2652.
1999	HR 1858: Consumer and Investor Access to Information Act. It disallows verbatim copying of a database.
2003	HR3261: Database and Collection of Information Misappropriation Act. It disallows free-riders from creating functional equivalent in the same market as the creator database to reduce the creator's revenue.
2004	HR 3872: Consumer Access to Information Act. It prevents a free-rider from engaging in direct competition that threatens the existence or the quality of the creator database.

As a result, there remain certain legal uncertainties in data reuse and database protection. What should a firm do if it is a database creator? What should a firm do if it wants to reuse someone else's data? Although we cannot predict how future regulations (if any) will evolve or how future data reuse related litigation will be determined, we have found several useful principles for answering these important questions by applying economic reasoning. This approach is useful because the lawmakers and the court often use economic reasoning when they introduce new laws and interpret the laws that involve activities such as data reuse. Thus we can formulate data strategies for the creators and the reusers in anticipation of the most likely policy choices.

3. Insights from Economic Analysis of Database Protection

An economic model has been constructed to analyze the fundamental issues related to database protection that concern law makers [Zhu et al. 2008a]. The model considers a database creator, which incurs a fixed cost to create a database, and a database reuser, which extracts a certain amount of data from the creator database to create the

reuser database. It assumes that the reuser possesses the technologies such as data extraction, integration, and semantic reconciliation [Chang et al. 2006; Firat et al. 2001; Goh et al. 1999; Zhu et al. 2008b] that allow for easy reuse of data on the Web so the reuser's fixed cost of creating its database is significantly lower than that of the creator. The two databases can be different in various aspects, e.g., data coverage, functionality, and quality. The users choose a database according to their preferences. Inefficiencies in any transaction are represented as transaction cost. When considering data reuse related issues, the lawmakers and the court are expected to make choice so that the society as a whole benefits.

The results from the model indicate that these choices depend on three factors: (1) the level of differentiation between the creator database and the reuser database, (2) the cost of creating the initial database relative to the database's market value, and (3) the transaction cost measured by transaction efficiency. The choices in relation to the first two factors are illustrated in Figure 1.

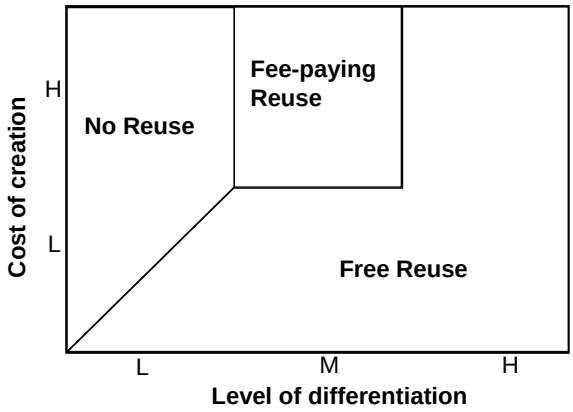

Figure 1: Database protection and data reuse policy choices.

No reuse. When the level of differentiation is low, not allowing reuse is reasonable because such reuse adds little value and the intense competition can drive the price so low that the creator cannot have enough revenue to offset the cost. Verbatim copying of an entire database is a typical example of this scenario.

Free reuse. When the level of differentiation is moderate or high, free reuse should be allowed in two scenarios: low creation cost or high differentiation regardless of creation cost. With moderate differentiation, competition is not as intense as in the case of low differentiation. The softened competition allows the creator to make enough revenue to offset its cost. With high differentiation, there will be little competition between the creator and the reuser. In other words, the data reutilization has little impact on the creator. Although in both cases the reuser could be required to pay the creator a fee, this is not desirable because there is transaction cost involved in money transfer. The fee benefits the creator, but it does not create any extra value and the society as a whole incurs a transaction cost.

Fee-paying reuse. When the level of differentiation is moderate but the cost of creation is high, the reuser should pay a fee to the creator. Without a fee the reuse would cause market failure, but with a fee the creator can sustain. Since the creator may not be willing to license its data to the reuser, a compulsory licensing provision should be in place.

Using the insights from the economic analysis and the framework of strategic analysis [Porter & Millar 1985], we develop data reuse strategies for database creators and data reusers.

4. Strategy for Database Creators

It is always good to be vigilant about reusers who extract data to recreate the same copy of the creator database. But most reusers are unlikely to do this for two reasons. First, although the U.S. still does not have database law similar to that of the E.U., the economic analysis earlier indicates that verbatim copying of a database should be disallowed to preserve the incentives for creating the database in the first place. Second, the "law" of economics shows that the reuser will not benefit from "copying" a database because the price competition will drive the price to the marginal cost of database, which can be near zero. Overall, rational reusers will not engage in verbatim data copying activities, which bear abundant legal risks and little economic benefits.

But if the reuser uses technology to create a significantly different database that has richer data and functionality, there doesn't seem to be any effective legal means to protect the creator database against such reuse.

What else can a creator do? Think *data repurposing*: that is, find innovative ways to make the data valuable to broader market segments that are not served by the original purposes of the data.

There are at least two possible innovative ways of data repurposing to enhance its value: (1) sell the "private" data that is related to the openly accessible data, and (2) be a reuser but offer better service by leveraging the creator's unique capabilities.

Sell "private" data. Most likely a firm that offers open access to certain data also has certain other "private" data that makes the combined dataset unique. Like other types of information goods, a unique database has much less competition and can be sold with a good profit margin. For example, from eBay's website, one can only obtain the current bidding prices of various items. The actual transaction prices (and historical prices for similar items) are not openly available from eBay, neither are the quantities sold at different prices. Such information is vital to people who want to analyze online auction market for purposes such as auction mechanism design and auction trend analysis. This data is generated as a necessary by-product of the transactions on eBay's website. In other words, the initial purpose of the data is a necessity of the online auction business. Realizing that the data combined with auction prices can serve other purposes such as market research, eBay is now selling its data via licenses directly as well as through resellers such as DataUnison. Similarly, as the horse racing authority of the U.K., the BHB is responsible for creating the lists of upcoming races and ensuring the eligibility of the participants. To ensure eligibility, BHB has to collect and verify the information about the identity of the person entering the horse, the characteristics of the horse, the identity of the owner and the jockey. This data is not given out for free; instead, BHB sells it via licenses.

A database creator should strategically decide what part of the database would be free to the public, and what should be kept private so that it can be sold. The combined data (free plus private data) is more useful because it represents a complete dataset. Further, the free database can often help increase the demand for the private database. This happens when there is network externality between the free database and the private database, in which case, the private database is perceived more valuable as more people use the free database. For example, when more people look up the upcoming horse races, they are more likely to be interested in knowing about the horses and the jockeys entering the races. Since this information can help people predict the winner, they will value the information highly. As a result, a high price can be charged for the private database. In this case, the gain from the increased demand and valuation of the private database can more than compensate the profit forgone with the free database. This business model is sometimes known as "Freemium" because the free products help the firm to charge a premium for products that are for sale [Wilson 2006]. A formal analysis of this strategy in the more general domain of digital goods can be found in [Parker & Van Alstyne 2005].

For purposes of getting across the idea, in the above discussion we assumed that BHB has a choice whether to give away the upcoming horse racing list for free. But BHB has to give away the data because it is the authority organizing horse races. Similarly, eBay has to give away data of auction items and their bidding prices because eBay is an online auctioneer. In this situation, a creator should try to identify data complementary to the data that has to be given away. If the complementary data does exist yet, the creator should invest in collecting and vetting the data. For eBay the quantity and selling price data collected from auction transactions is a good compliment of the auction price data. For BHB, the data about the horses, their owners, and the jockeys is a good compliment of the upcoming racing lists. In fact, BHB invested approximately £4 million a year on maintaining the private database [ECJ C338/02].

Become a reuser. Now let us look at the strategy for a creator to be a reuser also. This strategy is different from the ordinary "me too" or "if you cannot fight them, join them" reactions. When a creator also becomes a reuser, the creator should examine its unique capabilities and leverage them to improve its core business or expand its business scope [Venkatraman 1994]. Before it was acquired by Barnes and Noble, the online book seller Books.com displayed its price along with those of its main competitors – to highlight that it was offering good bargains. To do this, it reused the price data from the competitors in a similar way that Bidder's Edge reused price data of online auction sites. Books.com guaranteed to have the best price or it would lower its price. With the intelligence gained from reusing data of competitors and the guaranteed lowest price, Books.com had the potential to establish itself as the best discount book seller online. This practice should send a strong signal to the consumers about its confidence in the product offerings. Had it continued to build up its market leadership, it could have exercised its bargaining power to negotiate a better price from the book suppliers so that the lowest price guarantee would not erode its bottom line.

The online travel booking agencies could use this approach to establish and maintain their market leadership. They indeed often use their bargaining power to secure special rates from the suppliers. Unfortunately, some of the online booking agencies have taken the opposite approach by refusing comparison and even threatening others such as Kayak from doing it. We do not believe this is effective. Information cannot hide –most online consumers have become sophisticated and they will research before buying. The effort that a firm spent on trying to be excluded

from being compared is pretty much spent in vain. For a data creator (e.g., a vendor), it is better to focus on leveraging its unique position to give customers the best offers, the convenience to locate its offers, and even the confidence that its offers are the best.

Being a reuser sometimes has opportunity to learn about "non-customers" and why they choose the competitors. For example, UPS funded iShip, a data reuser that compares shipping charges of DHL/Airborne, FedEx, UPS, and USPS. Through iShip, UPS could learn about how consumers choose shipping carriers, including its competitors. This offers valuable business intelligence that cannot be obtained easily from conventional approaches. With the integrated data, UPS cannot only learn about its customers shipping desires, but also its non-customers – the customers of its competitors.

Sometimes a creator can also find opportunities of expanding business scope when it becomes a reuser. UPS is a shipping carrier. In addition to this core business, UPS also offers shipping management consulting and system integration services through its iShip subsidiary. The revenue generated through these services and the business intelligence gained help increase revenue streams and improve the overall competitiveness of UPS.

5. Strategy for Data Reusers

Data reusers often possess the technology that allows for efficient reuse of data from database creators. By extracting and integrating data from multiple databases, data reusers can add value to the existing data. This is because it not only provides the convenience of easy access to otherwise disparate data, the integrated data also provides insights otherwise invisible with disparate data.

Although with technology the reuser probably can reconstitute the entire database of the creator, this should not be done for reasons discussed earlier. Instead, focus on using the technology to construct a unique database that will add value to existing data. In practice, it is always good to consult legal counsel about the possible need for obtaining licenses with reasonable terms from the database creators.

Differentiate. Differentiation can be in various dimensions. Most price comparison sites reuse data from a wide range of vendors. The immediate value they add is improved efficiency in searching for best deals. They can also further differentiate their databases by improving the quality of the data and adding functionally to the database. AddAll.com reuses price data internationally. It presents data in a uniform currency chosen by the user even though the original data may be in various other currencies. Therefore, its data has higher value to the user because it is easily interpretable and more usable than the data in original sources. Interpretability is an important data quality dimension when data quality is broadly defined as "the fitness for use" [Wang & Strong 1996]. There are several ways of improving the functionality of the reuser database. For example, realizing that the airfares are sensitive to dates, several airfare comparison sites such as AirfareWatchdog search for fares within certain ranges of the specified dates. This feature provides convenience to leisure travelers who are often flexible in traveling dates. As another example, both iShip and its competitor InterShipper provide the capability of integrating with the user's internal logistics management and accounting systems.

The differentiation strategy helps to minimize legal risks, but it may bear certain market risks. As discussed in [Wan 2009a], sometimes it can be risky to be radically different because certain features (especially user interface related features) may have become de facto standards expected by users. These risks must be managed by careful usability tests and market research. In fact, with increased capabilities of information technology, firms are now able to introduce innovative features quickly, incrementally, and at a very low price through frequent, fast, and real-time experiments [Brynjolfsson & Schrage 2009]. This allows the reuser to introduce differentiated, value-added services quickly with minimal market risks.

Analyze the data. In addition to leveraging technology to drive innovation in developing differentiated databases, a reuser should also look into ways of leveraging another unique asset that it possesses: the integrated data itself! A reuser can use the data to make better decisions or offer services to help others make better decisions. Books.com used the integrated price information to dynamically adjust its price. Although initially it tried to offer lowest prices on all items at all times, it could employ sophisticated pricing strategies to offer lowest prices on some items at various times to improve its profit margin. This strategy has been commonly used by many conventional brick-and-mortar businesses [Varian 1980]. A reuser can also help others make better decisions by offering services based on after-aggregation analyses [Madnick & Siegel 2002]. For example, BizRate is a price comparison site. It also analyzes the searches and click streams to its price comparison database to produce market analysis that can be sold to market research firms and retailers. Maxmiles is a relationship aggregator that produces personalized integrated databases for individuals to see their rewards program accounts such as frequent flyer accounts in one place. Aside from providing the convenience of integrated view of multiple accounts with a single logon, Maxmiles also analyzes the data to identify potential missing miles and miles that are near expiration. Similarly, Cadence Network, acquired by Avista Corp in mid 2008, integrates utility billing data for multi-site companies such as

various chain stores. In addition to providing an easy access to the integrated data, it also analyzes the data to identify various cost-saving opportunities for their clients [Zhu et al. 2001].

6. Conclusion

Whether realized or not, once a firm puts up a website, the chances are that the firm has made a database available for potential reuses by others. A firm may own the database, but it is difficult to own its contents after they have been made publicly accessible, unless the firm finds innovative ways of using the contents. This is because currently there is no existing law in the U.S. to protect the factual contents that have been made public. Even the EU Database Directive only provides limited protection to the contents. The recent ECJ judgments of data reuse cases [ECJ C-203/02, C-338/02, C-444/02, C-46/02, and C-304/07] in the EU indicate that the Database Directive does not prevent someone from reusing data unless the reuse reconstitutes a substantial part of creator's database content, neither does it protect the investment in maintaining the databases if the reuse concerns data that is generated as a necessity of the creator's business. This is good news to innovative reusers.

The economic analysis shows that if there is law to be introduced in the U.S. for database protection, most likely it will be on the side of the innovators. That is, the law will not limit those who can find innovative ways of reusing the data to create value. So when a firm has data, use it, innovatively – otherwise, most likely someone else will. When a firm reuses data of other firms, add value. Serendipitous data reuse has made the web an "exciting" place for innovation, and will continue to make the web a valuable platform for delivering the value created by innovative reuses. Comparison-shopping and other value-creating data reuse services will continue to evolve to leverage the increasing availability of data and capability of emerging technologies.

In future research, we will develop economic models to examine the effect of network externalities of comparison-shopping services on service providers and data creators. The analysis should reveal further insights about policy choices. We will also conduct theoretical analysis to produce guidelines for data creators to determine complimentary datasets.

REFERENCES

BITS "BITS Voluntary Guidelines for Aggregation Services," January 2004, available at http://www.bits.org/downloads/Publications%20Page/bitsaggguide2004.pdf, last accessed April 30, 2010.

Bock, G.W., S.Y. Lee, and H. Li, "Price Comparison and Price Dispersion: Products and Retailers at Different Internet Maturity States," *International Journal of Electronic Commerce*, Vol. 11, No. 4:101-124, 2007.

Borzo, J. "Lost in Traffic: The Boom in Comparison-Shopping Sites Threaten to Squeeze out Smaller Businesses; Here is How the Little Guy can Survive," *The Wall Street Journal*, July 26, 2004, R9.

British Horseracing Board Ltd. and Others v. William Hill Organization Ltd. Case No. CHANI/2001/0632/A3, available at http://www.bailii.org/ew/cases/EWCA/Civ/2001/1268.html, last accessed April 28, 2010.

Brynjolfsson, E. and M. Schrage, "The New, Faster Face of Innovation," *The Wall Street Journal*, August 17, 2009.

Chang, C.H., M. Kaye, M.R. Girgis, and K.F. Shaalan, "A survey of web information extraction system," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 10:1411–1428, 2006.

eBay "eBay Completes Acquisition of Shopping.com," August 30, 2005, available at http://investor.ebay.com/releasedetail.cfm?releaseid=171732, last accessed April 30, 2010.

eBay, Inc. v. Bidder's Edge, Inc. 100 F. Supp. 2d 1058, N.D. Cal., May 24, 2000.

ECJ (European Court of Justice), http://curia.europa.eu/jurisp/cgi-bin/form.pl?lang=en, cases can be retrieved using case numbers C-203/02, C-338/02, C-444/02, C-46/02, and C-304/07, last accessed April 30, 2010.

Fasli, M. "Shopbots: A syntactic present, a semantic future," *IEEE Internet Computing*, Vol. 10, No. 6:69-75, 2006.

Feist Publications v. Rural Telephone Co. 499 U.S. 340, 1991.

Firat, A., S. Madnick, and M. Siegel, "The Cameleon web wrapper engine," Proceedings of the Workshop on Technologies for E-Services, Cairo, Egypt, 14–15 September, 2001.

First Union Corp v. Secure Commerce Services, Inc. No. 3:99CV519H, W.D.N.C., December 30, 1999.

Fowler, G.A. "Auctions Fade in eBay's for Growth," *The Wall Street Journal*, May 26, 2009, A1.

Frauenfelder, M. "Sir Tim Berners-Lee," *Technology Review*, Vol. 107, No. 8:40-45, 2004.

Goh, C.H., S. Bressan, S, Madnick, and S, M. Siegel, "Context interchange: new features and formalisms for the intelligent integration of information," *ACM Transactions on Information Systems*, Vol. 17, No. 3: 270–293, 1999.

Guterman, J. "Does Current Copyright Law Hinder Innovation?" *MIT Sloan Management Review*, Vol. 50, No. 2:14-15, 2009.

Hunter, D. "Cyberspace as Place and the Tragedy of the Digital Anticommons," *California Law Review*, Vol. 91, No. 2:439-520, 2003.

Johnson, A. "Cheap-Tickets Sites Try New Tactics," *The Wall Street Journal*, October 26, 2004, J.

mySimon, Inc. v. Priceman, LLC. No. 5:99cv20939. N.D. Cal., February 29, 2000.

Krulwich, B. "The BargainFinder Agent: Comparison Price Shopping on the Internet," in J. Williams (Ed.), Bots, and Other Internet Beasties, pp. 257-263, Indianapolis: SAMS.NET.

Madnick, S. and M. Siegel, "Seizing the Opportunity: Exploiting Web Aggregation," *MISQ Executive*, Vol. 1, No. 1: 35–46, 2002.

Li, X., S. Madnick, H. Zhu, and Y. Fan, "Improving Data Quality for Web Services Composition," Proceedings of the VLDB Quality in Databases (QDB) Workshop, Lyon, France, August 24, 2009.

Lindsey-Mullikin, J. and D. Grewal, "Imperfect Information: The Persistence of Price Dispersion on the Web," *Journal of the Academy of Marketing Science*, Vol. 34, No. 2:236-243, 2006.

Nettleton, E. "ECJ Rules on Acts of 'Extraction' that Infringe Database Right," *Computer Law & Security Review*, Vol. 25, No. 2:181-184, 2009.

O'Rourke, M.A. "Is Virtual Trespass an Apt Analogy?" *Communications of the ACM*, Vol. 44, No. 2:98-103, 2001.

Parker, G. and M. Van Alstyne, "Two-Sided Network Effects: A Theory of Information Product Design," *Management Science*, Vol. 51, No. 10:1494–1504, 2005.

Porter, M.E. and V.E. Millar, "How Information Gives You Competitive Advantage," *Harvard Business Review*, Vol. 63, No. 4: 149-160, 1985.

Smith, M. and E. Brynjolfsson, "Customer Decision Making at an Internet Shopbot: Brand Still Matters," *The Journal of Industrial Economics*, Vol. 49, No. 4:541-558, 2001.

Ticketmaster Corp. v. Tickets.com, Inc. CV 99-7654 HLH (BQRx), C.D. Cal., March 27, 2000a.

Ticketmaster Corp. v. Tickets.com, Inc. C.D. Cal., August 10, 2000b (Court opinion not intended for publication), available at http://pub.bna.com/ptcj/ticketmaster.htm, last retrieved July 14,2010).

Ticketmaster Corp. v. Tickets.com, Inc. WL 21406289, C.D. Cal., March 7, 2003.

Varian, H.R. "A Model of Sales," *American Economic Review*, Vol. 70, No. 4:651-59, 1980

Venkatraman, N. "IT-Enabled Business Transformation: from Automation to Business Scope Redefinition," *Sloan Management Review*, Vol. 35, No. 2:73-87, 1994.

Wan, Y. "Social Aspects of Agent Design," First Monday, Vol. 14, No. 7, July 6, 2009a.

Wan, Y. (Ed.) *Comparison-Shopping Services and Agent Designs*, IGI Global, Hershey, NY, 2009b.

Wan, Y. and Y. Liu, "The Impact of Legal Challenges on the Evolution of Web-based Intelligent Agents," *Journal of International Commercial Law and Technology*, Vol. 3, No. 2: 112-119, 2008.

Wang, R. and D. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information System*, Vol. 12, No. 4:5-33, 1996

Wilson, F. "The Freemium Business Model", A VC Blog, March 23, 2006, available at http://www.avc.com/a_vc/2006/03/the_freemium_bu.html, last retrieved May 9, 2010.

Zhu, H. and S. Madnick, "One Size does not Fit All: Legal Protection for Non-Copyrightable Data," *Communications of the ACM*, Vol. 52, No. 9:123-128, 2009.

Zhu, H., S. Madnick, and M. Siegel, "An Economic Analysis of Policies for the Protection and Reuse of Non-Copyrightable Database Contents," *Journal of Management Information Systems*, Vol. 25, No. 1:199-232, 2008a.

Zhu, H., S. Madnick, and M. Siegel, "Enabling global price comparison through semantic integration of web data," *International Journal of Electronic Business*, Vol. 6, No. 4: 319-341, 2008b.

Zhu, H., M. Siegel, and S. Madnick, "Information Aggregation – A Value-Added E-Service," 5th International Conference on Technology, Policy and Innovation, The Hague, The Netherlands, June 26-29, 2001.