View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2002 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2002

DOES YOUR DATABASE ENGENDER TRUST?

Bruce Rollier
University of Baltimore

Danielle Fowler
University of Baltimore

Follow this and additional works at: http://aisel.aisnet.org/amcis2002

Recommended Citation
Rollier, Bruce and Fowler, Danielle, "DOES YOUR DATABASE ENGENDER TRUST?" (2002). *AMCIS 2002 Proceedings*. 307.
http://aisel.aisnet.org/amcis2002/307

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2002 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

DOES YOUR DATABASE ENGENDER TRUST?

Bruce Rollier and Danielle Fowler
University of Baltimore
brollier@ubmail.ubalt.edu dfowler@ubmail.ubalt.edu

Abstract

In B2C electronic commerce, trust is dependent on the customer's experiences in interactions with the computer system. A well designed, reliable online site that provides customers with a sense that they are being given full information and treated fairly, courteously, and with sincere concern for security and privacy builds trust and repeat business. Careful design of the system and the underlying database are crucial in providing such an experience. This article will explore the factors in the design of systems and databases that contribute to a higher degree of trust among online customers.

Introduction

Trust has always been an important component in the relationship between a company and its customers. Consumers have concerns about such matters as fairness in pricing, merchandise quality, willingness to accept returns, and the validity of guarantees. They are most likely to shop at stores where they can feel confident in the honesty and trustworthiness of the company. In a traditional store where the merchandise can be seen and touched and where there can be interaction with human clerks and managers, the database system is a relatively minor factor in building a high level of trust. The system must ensure that pricing and billing is accurate, that inventory data is reliable, and that discounts are properly applied, and a reliable database system will ensure such factors, but they are usually important in building trust only in a negative sense, when errors are made. The customer can see and handle the products; they can see the posted prices and compare them with other products in the same store, with printed advertisements, or with other stores. They can try on the clothes for proper size and appearance, see and compare colors, feel the fabric, and read the fine print on the label. They can converse with the salesperson about prices or the relative merits of different products, or with the manager about store policies. Trust is built over time, largely from these human interactions and from physical proximity to the merchandise.

For an online company selling products through the Web, the database and the system that maintains it assumes a much more critical role. The products for sale cannot be physically examined or tried on for size. There are no humans available to answer questions or to offer sales advice. Some level of customer trust may already exist, if the brand is familiar or the company itself is well known for its physical stores, but one erroneous transaction can destroy this in an instant. The system must provide means for the customer to examine the products virtually; it must provide information and images which substitute for being able to see and feel the actual merchandise. The customer must be assured that the system can provide the products that are ordered at the correct price and delivered to the correct address in a timely fashion. The system must be able deal intelligently with questions with no human intervention for all but the most exceptional cases. It must cope with huge transaction volumes with very short response times and with error rates approaching zero.

Trust is difficult to earn, but easy to lose. A great restaurant may have painstakingly built its reputation over many years with delicious food and superlative service, but the public's trust can be lost virtually overnight with a single case of food poisoning. The system that has earned a well-deserved reputation for being extremely reliable may cease to work well as volumes increase, as business rules change, or as the emphasis on data quality is given less attention. The system must be scalable and adaptable, with well trained personnel who have a passion for quality. This article will explore the factors in the design of database systems that contribute to a high degree of trust among online customers.

Existing Studies on Trust

Trust is a subject that has been addressed in many discipline areas, including economics, marketing, psychology, and management. In the information systems arena, interest in trust has grown with the advent of eCommerce systems, as companies have struggled to understand how to successfully woo customers in a virtual environment.

Existing studies relevant to business have investigated trust at several levels. Many studies have looked at antecedents or factors influencing an individual's trust in an organization / institution or brand (of most interest in B2C research) (e.g. Jarvenpaa et al. 2000; Stewart, 2000). Others have looked at factors influencing an organization's trust in their trading partners (e.g. Ratnasingham and Kumar, 2000).

While this research is revealing in determining how trust in an organization may initially be built, or sustained, our interest is at the application / system level. Studies in this area have focused primarily on user interface-related factors that affect trust in a particular web site: for example, perceived usefulness or attractiveness of the site (van der Heijden, 2000), or on security-related trust issues, generally at the transaction-level: authentication, confirmation, non-repudiation, availability, etc.

While these are clearly important areas, the data in a system and the architecture that supports them can also influence the trustworthiness of a system. Relatively little attention, however, has been given to database design and management issues. Our interest is therefore in the impact of database design issues on system trust. At what point will data inaccuracy cause a user to lose faith in an online ordering system, or a manager in the results of a data warehouse query? Will an update that takes 24 hrs to propagate through a data warehouse or data mart affect trust in the results of a query?

We are concerned with the factors that build trust, and also those that can break it. Once a user/customer has established a trust relationship with a system, what qualities must the system have in order not to lose it? Are the antecedents for establishing trust the same as for maintaining trust in a system? Decisions about what data is to be collected from the customer, how the database will be organized, and the tools available for data analysis are crucial in answering such questions.

Building a Trustworthy Online System

In preparing to design an online system that will facilitate the building of customer trust, the first requirement for the system developers is a thorough understanding of the company's business model. With an online system, especially one for an organization that does all of its business online, it is probable that the chief system developer will play a key role in the development of the business model. The system will BE the company, for all practical purposes, and if the system fails so will the business. Does the model make sense? Can a system be specified that will represent it? What are the volumes it will have to handle? Is the existing hardware powerful enough for these volumes? Does software exist that is powerful enough and sophisticated enough for this model? If new software is needed, or if there is a need to customize existing software, are personnel available with the necessary skills? Is there adequate financing to provide the resources for this project in a sufficient time schedule?

The main objective of any B2C system is to attract customers who will purchase products, both first-time patrons and returning ones. First-time customers may visit the site because they are familiar with the brand, or the company's reputation, or because of effective advertising or a specific promotion. These factors may have been affected by the database through online analytical processing (OLAP) or data mining. For example, promotions may be targeted toward specific classes of potential customers through the results of data analysis. If the customer then visits the site, the database is a vital factor in convincing the customer to make a purchase. It is critical that the database, in concert with the site interface, be organized to facilitate finding the desired article. Once it is found, the database must provide a simulation of the product realistic enough to substitute for being able to actually see and touch the article. A database is more than text; it may also include pictures or other graphic objects. The DBMS must be able to locate the relevant picture from hundreds of thousand of other pictures and present it in full color and high fidelity.

During the customer's visit, clickstream data (Kimball and Merz, 2000) is collected for the database which can provide the basis for further analysis, to attract other new customers. If the customer's initial experiences at the Web site are easily performed, if they are satisfied with all aspects of the transaction, then they are likely to return. The experience can be negatively impacted if they cannot easily find what they are looking for, if it is not sufficiently visualized, if there is some kind of error associated with the transaction, or they perceive an infringement on their privacy.

In designing and implementing a trustworthy system, the developers must consider the entire system, not just the Web interface. This includes data capture (what elements should be collected, how they might be obtained; what controls should be established

to ensure their accuracy), data storage (backup, recovery, mirroring, data types, characteristics of the DBMS), network specifications, decisions about data distribution and distributed processing, and so forth. Will there be an inventory? If so, will each item be stored in many places or in only one? Will the new system share a database with an already existing system? These questions and many others must be thought through carefully and the answers agreed to by the CEO and all the principal parties before development starts. A project manager experienced in the successful implementation of large projects should be appointed or hired. Skills of project personnel should include both database and Web expertise. All project members should be determined to build a system with the highest possible quality and reliability.

The first step is to design the data model (Simsion, 2001; Reingruber and Gregory, 1994; Silverston, 2001). The designer must have both a thorough understanding of the business and its industry as well as the technical knowledge to correctly prepare the model. Data modeling for an online operational database is essentially the same as for more traditional ones. To test the model for completeness, the designer should consult users, or potential users, to determine what questions they might ask and whether the model has the correct information to answer their questions. All of these steps are crucially important, and if the model is not well designed the system is likely to be a failure. Instructions for defining queries and reports will be difficult to formulate and inefficient to execute, and performance of the system may be painfully slow.

A company engaged in electronic commerce is likely to require a very large database, with hundreds of tables including some with 50 to 100 columns and several millions of rows. The tables actually used on the Web may be a subset of the entire company database, but it cannot be independent; it must be recognized that the databases must be completely and seamlessly integrated. Going online means that the resulting database will be larger than it was previously. The larger the database, the more important is proper organization (Mallach, 2000). If a database is small, say 50 megabytes, and there is only one user, summing the amount fields in twenty different categories would take only a fraction of a second on a desktop computer. Today's commercial databases are often measured in terabytes, and the data may be distributed to many different locations, and shared by thousands of concurrent users. Such volumes cause many serious problems that would be trivial for a small database. Bugs and data integrity violations that might occur only "once in a million" may occur a thousand times in a billion-row table (Bustamente and Sorenson, 1994). A missing join in a single embedded query could have a disastrous impact on performance of the entire system. A large number of disk drives may be required to store all these data, with mirroring and other sophisticated storage techniques to provide redundancy (Bernstein and Newcomer, 1997), backup and recovery procedures, checkpoints, and a high level of indexing. Indexing in turn complicates database updating.

When the data model is completed and the database tables have been designed, the design must be implemented. In a new company that is completely online such as Amazon.com, the process is straightforward, though enormous. A new company does not have to compromise their design to accommodate decisions made in previous years for the existing databases. It is a huge task, nevertheless. To prepare their initial database of a million books, a million authors, prices, inventory numbers, and all the associated tables, (Spector, 2000), Amazon could utilize existing files provided by the publishers rather than keying everything in. There was a universal standard for identifying each book, the ISBN. Still, all the data had to be integrated into a single set of tables. Formats, field lengths, and other factors probably varied significantly. They had to develop a system to ensure that the warehouse location of a book was accurately connected to each book. An established company such as Barnes & Noble has somewhat different problems, but they are just as formidable. Bustamente and Sorenson (1994) point out that "whenever an organization sets out to siphon several years of data into a single platform, data integrity problems are encountered. When file and data item formats changed throughout the years, not all changes were rippled through all the historical data. Logical errors when creating the historical data were not discovered until the data were loaded." (p. 237).

In deciding whether to trust the company enough to buy the offered merchandise, the customer is not considering these vast volumes, but is interested in only one or two items. As far as that customer is concerned, what is important is that this tiny transaction be done correctly; if it is, trust will rise accordingly. If it is not correct, that customer will likely never return even though the error may have been only one in a million.

Factors Affecting Database System Trust

Collecting Information

Building productive, ongoing relationships with customers requires some degree of personalization. The technique called "automatic collaborative filtering" (ACF) selects content to be presented to the Web visitor based on characteristics shared with previous visitors (Kimball and Merz, 2000). This may enable the company to group similar customers together and perhaps recommend collateral products that may interest them. "Cookies" are frequently used to identify re-visits. Many Web sites ask customers for personal information that is retained on the database. Such techniques, however, if not handled tactfully, can be

damaging to the customer's level of trust. Cookies can be both a privacy and a security risk (Sit and Fu, 2001). Kimball and Merz (2000, p. 127) suggest the following guidelines to build user trust when collecting personal data:

- Avoid asking for information that is not specifically required for a bona fide business purpose.
- Ask only for the information needed. Establish the trusted relationship first; postpone asking for demographic data until that happens.
- Don't ask for e-mail addresses unless you need them. Don't spam. Provide a simple method for the user to get off mailing lists.
- Never include personal information in clear text in a cookie.
- Subscribe to privacy auditors like TrustE and the Better Business Bureau, and follow their guidelines in your privacy policies.
- Use encryption on all forms requesting personal information.

Accessibility

A user will be more likely to trust a Web site that has a well-designed layout that makes it easy to find what they want. The entire site should have a standard page design that is similar to other major Web sites so that it seems familiar and intuitive in navigation. The products they search for should be accessible with as few clicks as possible. A correctly designed database can greatly facilitate this, with the product table organized in a logical hierarchy. An easy-to-use search capability should be provided that is designed to fit the database structure. The hierarchy should facilitate searching for a general category (e.g., Men's Clothing) with a drill down capability so that the user could find, in quick succession, an increasing specific product (e.g., men's shirts, then shirts in a particular collar size, sleeve length, color, brand, etc.).

Another aspect of the database is visual images such as photographs, drawings, video clips, or animation. Some of the automobile Web sites have images that can be rotated, so that the customer can see the car from all angles, and the ability to change colors. Some clothing sites provide the capability for the customer to see how the clothing item looks on a model of the same dimensions as themselves: a virtual fitting room. Most commercial sites provide at least a picture of the item being offered. These images must be stored in the database and retrieved for display along with the item's textual attributes. The storage requirements are massive, and there is a significant negative effect on retrieval times. Relational database systems with object-oriented capabilities have recently been introduced by the major database vendors which make these images easier to handle (Stonebraker and Brown, 1998). The new systems provide for abstract data types and allow them to be defined in terms appropriate to the medium, facilitating analysis and retrieval.

High quality images, rapidly retrieved, and especially facilities for virtual manipulation of the images can be of great significance in establishing a high level of trust. The closer the designers can come to simulating the look and feel of the product that a customer can have in a physical store, the more comfortable they will be with shopping online. As computers become ever faster and more powerful, as bandwidth increases, as computer memories get larger and faster, and as software becomes more sophisticated, these images will become increasingly realistic, thus facilitating trust. It can be expected that virtual reality techniques will enhance this capability in the future.

For now, however, providing appropriate images with fast response times is extremely challenging. The images must be stored in the database, accessible via a mouse click, and displayed in full color. For an adequate display, the user may require a fast Web connection and a large computer memory. The images must be identifiable with a specific product number. They consume a large amount of storage in the database; a one-inch square image in high resolution color might consume 12 megabytes or more of storage, and for large electronic commerce sites there might be hundreds of thousands of images.

Security

Security is another critical factor for inspiring trust. The system must provide foolproof security for the customer, and the customer must be convinced that it is foolproof. Secure servers are a necessity for processing orders and credit card payments. Credit card numbers must be stringently protected. Shipping addresses are critical also; there have been instances of hackers breaking into a database and diverting shipments to other addresses. Security of the customer's demographic information is also very important.

Feedback and Analysis

Trust can also be enhanced with "reputation systems" (Resnick, et al, 2000), for collecting feedback from outside parties and making it available to potential customers for evaluation of the company's products. eBay has a very effective system called the Feedback Forum so that buyers and sellers can rate each other. Books offered on Barnes & Noble or Amazon.com have available ratings and comments from readers of the books; this is an effective substitute for being able to browse through the book at a physical store. For many readers, this feedback is a valuable, time saving service which can help them select an appropriate book on a particular topic and to pick one that is highly regarded by previous readers. Such systems must be designed and implemented as a regular segment of the database so that ratings can be retrieved along with the other attributes of the item.

A major capability for analysis is a data Webhouse (Kimball and Merz, 2000). Another term for this is Web farming (Hackathorn, 1999). A Webhouse is a data warehouse with the addition of "clickstream" data. Essentially, every mouse click is captured and stored for analysis. The Webhouse is designed in the same way as a regular data warehouse, with dimensions and fact tables in a star schema (Mallach, 2000). It is distinct from the transaction database, and is used for data retrieval only, but it can be a powerful analysis tool and can greatly facilitate the building of customer trust. The clickstream can be used to analyze customer characteristics, their typical Web behavior, their favorite products, etc. Like a data warehouse for a bricks and mortar business, the Webhouse is mass loaded (rather than updated a transaction at a time) from the transaction database at periodic intervals. There must be a rigorously controlled ETL procedure (extract, transform, load) during which the data is carefully cleaned and transformed into standard formats, data from various sources are integrated, and appropriate aggregates and snapshots are created to facilitate rapid retrieval. Another major difference from a production database is that historical records are included, often comprising two or three years of transaction data.

The greatest system challenge is the enormous data volume that must be stored and made available for retrieval. Already a number of data warehouses are in the terabyte range, and some such as Wal-Mart are approaching a petabyte in volume (Westerman, 2000). Adding the clickstream data greatly increases the storage volumes. However, since data warehouses and Webhouses are used for analysis by decision makers, not for customer access, the retrieval times can be somewhat slower. The analyses can help to build customer trust by providing a clearer picture of what the users want. The data can also improve inventory management and thus minimize stockouts, which can have a major negative impact on customer satisfaction and trust. Valuable data mining studies can be performed with these analytical databases, finding interesting patterns of customer behavior on the Web (Witten and Frank, 2000; Han and Kamber, 2001).

A Framework For Investigation

Our ultimate objective is to develop an overall framework of database-related factors affecting trust. Our initial investigations will target the following summary propositions, which will be tested via controlled laboratory experiments and focus groups. Results from existing studies will be incorporated into the framework synthesis process.

1. That data richness is positively associated with B2C ecommerce system participation from both first-time and returning customers. Data richness includes the precision of the data in defining and describing the product, and the variety and fidelity of photographs or other objects related to the product. Data richness enhances the customer's perceptions and understanding of the product without actually seeing and touching it.

2. That data depth is positively associated with B2C ecommerce system participation, by facilitating more effective advertising and targeted promotions to attract B2C ecommerce customers. Data depth involves having a sufficient variety of attributes for each entity to make data mining and OLAP analyses more effective in identifying customer characteristics for advertising and promotion. Clickstream data from the Web site and software to analyze it can greatly enhance these analysis activities.

3. That such factors as data accuracy, timeliness, completeness, and consistency are positively associated with system participation. These factors contribute significantly to the customer's perceptions of trust and to the likelihood of returning to the site for a subsequent purchase.

4. That collection of data which the customer considers private has a negative association with continuing system participation. Cookies are particularly vulnerable to perceptions of mistrust.

Conclusions

As the Internet and the Web have expanded, the world has become increasingly interconnected. In accordance with Metcalf's Law, the more connections there are the more valuable the participation. Business firms in the global economy have little choice but to join this electronic commerce environment, but the process of building efficient, reliable sites that customers trust requires the full-time vigilance of the executives and employees of the firm. It must be remembered that trust in the firm's Web site is also dependent on the quality and reliability of related sites such as the firm's suppliers, and all the systems that feed information to the site.

Many aspects of a system influence consumer trust: the interface, the performance and reliability of the hardware, software, and networks, and above all the integrity of the data and how that data is used. Developers need to consider trust issues throughout the system development life cycle. Trust features need to be built in from the beginning, not added later. Changing a system after it has been implemented costs much more than designing the changes in at the beginning, and often that makes it prohibitively expensive to add them.

The Web site should be as informative as possible, including clear, complete, and explicit policies and feedback from previous users. A well-thought-out system design that fits the business model may be expensive to implement, but it can pay great dividends if it contributes to a higher degree of customer trust. Lack of trust carries a heavy cost also; consider the charity which receives bad publicity about an illegal diversion of funds; donations are sure to drop drastically. If customers do not trust the system, they will not use it. MIS research in this area should focus on emerging technologies (wireless, XML, etc.) and the opportunities they offer for enhancing trust. Greater bandwidths may in the near future make virtual reality much more feasible, giving customers a greater sense of actually interacting with the system and the merchandise.

References

Bernstein, P. A., and Newcomer, E. (1997). Transaction processing. Morgan Kaufmann, San Francisco.

Bustamente, G. G., and Sorenson, K. (1994). Decision support at Land's End -- an evolution. IBM Systems Journal, 33(2), 228-240.

Hackathorn, R. D. (1999). Web farming for the data warehouse: Exploiting business intelligence and knowledge management. Morgan Kaufmann, San Francisco.

Han, J., and Kamber, M. (2001). Data mining: Concepts and techniques. Morgan Kaufmann, San Francisco.

Jarvenpaa, S.L., Tractinsky, N., Vitale, M. (2000) Consumer Trust in an Internet Store. Information Technology and Management, 1(1-2): 45-71.

Kimball, R., and Merz, R. (2000). The data webhouse toolkit: Building the Web-enabled data warehouse. Wiley, New York.

Mallach, E. (2000). Decision support and data warehouse systems. Irwin McGraw-Hill, New York.

Ratnasingham, P. and Kumar, K. (2000) Trading partner trust in Electronic Commerce participation. Proceedings of the 21st International Conference on Information Systems (ICIS'2000), Sydney, Australia, Dec 10 – 13, pp 544-552.

Reingruber, M. C., and Gregory, W. W. (1994). The data modeling handbook: A best-practice approach to building quality data models. Wiley-QED, New York.

Resnick, P., Zeckhauser, R., Friedman, E., and Kuwabara, K. Reputation systems. Communications of the ACM, 43(12), 45-48.

Silverston, L. (2001). The data model resource book, revised edition, Volume 1: A library of universal data models for all enterprises. Wiley, New York.

Simsion, G. C. S. (2001). Data modeling essentials, 2nd ed. Coriolis Group, Scottsdale, AZ.

Sit, E., and Fu, K. (2001). Web cookies: Not just a privacy risk. Communications of the ACM, 44(9), 120.

Spector, R. (2000). amazon.com - Get big fast: Inside the revolutionary business model that changed the world. Harperbusiness, New York.

Stewart, K. (2000) Transference as a means of building trust in world wide web sites. Proceedings of the 21st International Conference on Information Systems (ICIS'2000), Sydney, Australia, Dec 10 – 13, pp. 459-464.

Stonebraker, M., and Brown, P. (1998). Object-relational DBMSs: The next great wave (2nd ed.). Morgan Kaufmann, San Francisco.

van der Heijden, H. (2000) The impact of perceived web site characteristics on website traffic. Proceedings of the 13th International Bled Electronic Commerce Conference, Bled, Slovenia, June 19-21, pp. 414-425.

Westerman, P. (2000). Data warehousing: Using the Wal-mart model. Morgan Kaufmann, San Francisco.

Witten, I. H., and Frank, E. (2000). Data mining: Practical machine learning tools and techniques with Java implementations.