

# Improving the Quality of Corporate Household Data: Current Practices and Research Directions (ICIQ)

Stuart E. Madnick, Richard Wang, Frank Dravis, Xinping Chen

CISL WP #01-10  
November 2001

## Improving the Quality of Corporate Household Data: Current Practices and Research Directions

**Stuart Madnick**  
M.I.T.  
smadnick@mit.edu

**Richard Wang**  
Boston University  
rwang@bu.edu

**Frank Dravis**  
Firstlogic Inc.  
frankd@firstlogic.com

**Xinping Chen**  
Bose Corporation  
xinping\_chen@bose.com

**ABSTRACT:** Corporate household data not only refers to the strict hierarchical structure about and within the corporation, but also the variety of inter-organizational relationships. It is becoming increasingly important for many purposes ranging from CRM and ERP applications, to risk management, supply chain management, and marketing. We propose conceptual definitions for corporate household, corporate household knowledge, and corporate household knowledge processor. After describing research challenges and conceptual definitions, we summarize current practices and approaches. We then present a two-part plan: (1) continue our qualitative research to describe the various different sources, views, and purposes for corporate household data, including the rules used in each case; (2) apply the context interchange theory to represent the corporate household data and underlying knowledge and enable the context mediation technology to correctly understand and reason about both the context of the sources and the context of the user's query about corporate household data.

### 1. Introduction

How corporate structure and corporate relationships are interpreted and used depends on the context. Additionally, they evolve dynamically in the rapidly changing business environment. The ability to capture, manage, and use knowledge of corporate structure and relationships are fundamentally critical tasks underpinning many important activities, such as marketing promotion, financial risk analysis, and supply chain management. The problem goes beyond existing data quality research such as those found in [2, 6, 10-12, 14-17]. A recent literature search of ABI/INFORM from 1986 to 2001 [8] reveals that there is no corresponding concept for *corporate household data* to that of *individual household data*, and the terms *corporate household* and *corporate householding* have not been mentioned. Although there are many papers on "corporate structure" [1, 9, 13], that term does not capture the rich phenomena we are exploring. By *corporate household* we not only mean the strict hierarchical structure within the corporation, but also the variety of inter-organizational relationships.

To put this issue in perspective, consider a traditional household. As family structure evolves, such as the increasing number of single families, families with no children, or husband and wife with different last name, it becomes more difficult to define and identify "household", [7]. For example, are grandparents or visiting cousins living at same address to be considered part of the same household? Are two unmarried people living together a household? There is no single "right" answer; the answer depends upon the intended purpose of the question.

Similarly, a corporate household must be conceptualized within certain scope, content, and context, with the relationships identified within a corporation, between suppliers and the corporation, and between the business customers and the corporation. Conceptually, the corporate structure would also be different depending on different contexts such as a financial perspective, legal perspective, and the reporting structure. Identifying those contexts and

representing the right structure for the right task can provide competitive advantage.

Our primary research methods are semi-structured interviews via telephone, face-to-face, case studies, and surveys. From these results, we explore possible solution approaches and new technologies. Our long-term research goal is to define the concepts of *corporate household* and *corporate householding* and develop mechanisms to capture the metadata and business rules representing the semantics of *corporate household data*, and the *corporate household knowledge processor* that would produce the appropriate corporate household that would be fit for use depending on the context. We begin with expository example problems.

## 2. Corporate Household Knowledge Challenges

Corporate household knowledge serves various purposes, which in turn necessitate different interpretations of the information. Consider the list of organization names in Figure 1. What is the relationship among these names? As it turns out, these are all names that are in some way related to each other and International Business Machines Corporation (the name at the top of the list). These names include abbreviations (such as IBM), divisions (such as IBM Microelectronics division), wholly owned subsidiaries (such as IBM Global Financing), partially owned subsidiaries (such as IBM de Colombia, S.A.), companies that were acquired by IBM (such as Lotus Development Corporation), companies that were acquired and then later sold by IBM (such as SoftwareArtistry, Inc.), companies in which IBM has a minority joint venture interest (such as the Dominion Semiconductor company), and companies that IBM has a majority joint venture interest in (MiCRUS). It even includes IBM's original name, Computing-Tabulating-Recording Company.

What is the significance? Consider a rather simple question: "How many employees does IBM have?" In a recent study of a major insurance company, this was an important question asked in setting premium rates for business owner protection insurance [18]. Which entities listed above for IBM should be included in this count? How to avoid double counting? The answer depends upon the purpose of the question. The important and subtle issue is: "When is one entity to be considered part of the another entity?" Such corporate household knowledge is used for many different purposes: (1) financial risk, (2) account consolidation, (3) marketing (multiple divisions & subsidiaries), (4) customers & supplier consolidation, (5) customer relationship management, (6) regional and/or product separations, (7) legal liability in insurance, (8) conflict of interest & competition, and (9) ad hoc/temporary structures. In some cases the two entities should be combined and in other cases the two entities should not be combined.

As noted before, corporate household changes over time; thus, the context also changes over time. For example, at one point Lotus Development Corporation was a separate corporation from IBM. When doing a historical comparison of growth or decline in "number of employees"

International Business Machines Corp IBM IBM Microelectronics Division IBM Global Services IBM Global Financing IBM Global Network IBM de Colombia, S.A Lotus Development Corporation Software Artistry, Inc. Dominion Semiconductor Company MiCRUS Computing-Tabulating-Recording Co.
---

**Figure 1. List of organization names**

of IBM, should current Lotus employees be counted in a total as of today? Should the Lotus employees in 1980, when it was a separate corporation, be added with the IBM employees of 1980 to make a meaningful comparison?

Corporate household knowledge could be applied in many other areas. An executive of a global manufacturing company had concerns, in the global sourcing context, in identifying a manufacturing site that could produce a particular product with the lowest costs. A big part of manufacturing cost is raw material cost, and therefore identifying and maintaining relationship with material vendors are critical in reducing costs. However, due to localized systems, different manufacturing sites may have different, independent relationships/contracts with the same vendor for the same material. Inconsistencies between systems make it difficult to understand a vendor globally or know how much of a raw material is used on the global basis. As a result, the company could neither take advantage of nor negotiate low prices across all of its manufacturing sites [8].

### 3. Conceptual Definitions

*Random House Webster's Unabridged Dictionary* defines *corporation* and *household* as:

**Corporation:** (1) an association of individuals, created by law or under authority of law, having a continuous existence independent of the existences of its members, and powers and liabilities distinct from those of its members, (2) any group of persons united or regarded as united in one body.

**Household:** (1) the people of a house collectively; a family including its servants, (2) of or pertaining to a household: household furniture, (3) for use in maintaining a home, esp. for use in cooking, cleaning, laundering, repairing, etc., in the home: a household bleach, and (4) common or usual; ordinary.

The term "householding," although not in most dictionaries, has been used in an increasing number of contexts. For example, it was recently used in notices, such as the one below, sent to hundreds of thousands of people and organizations as a result of a recent SEC rule.

#### HOUSEHOLDING ELECTION

This notice has been placed in this mailing on behalf of your Broker or Bank. In December 2000, the Securities and Exchange Commission enacted a new rule that allows multiple shareowners residing at the same address the convenience of receiving a single copy of proxy and information statements, annual reports and prospectuses if they consent to do so. This is known as "Householding." Please note that if you do not respond, Householding will start 60 days after the mailing of this notice. We will allow Householding only upon certain conditions. Some of those conditions are:

- The Issuer agrees to have its documents Householded,
- You agree to or do not object to the Householding of your materials,
- You have the same last name and exact address as another shareowner(s),
- Consistency with your Broker's or Bank's practices.

If all of these conditions are met, and Securities and Exchange Commission regulations allow, your household will receive a single copy of proxy and information statements, annual reports and prospectuses. . . . Your affirmative or implied consent to Household will remain in effect until you revoke it by calling the telephone number listed in the HOUSEHOLDING ELECTION paragraph.

In terms of the scope and content of a corporate household, we found at least three types of important entities: the *corporation*, *suppliers/vendors*, and *customers*. A corporation includes relationships, functions, and people within the entity of the corporate, such as the one represented by organizational chart. We therefore propose the following conceptual definitions:

**In the dynamic and rapidly changing e-business environment, any group of persons united or regarded as united with the corporation, such as suppliers and customers whose relationships with the corporation must be captured, managed, and applied for the purpose of activities such as marketing promotion, financial risk analysis, and supply chain management in their entirety forms a corporate household. Note that there might be multiple overlapping but distinct corporate households, depending upon the precise set of relationships that are important for the task at hand.**

**The knowledge developed for such purposes are termed corporate household knowledge.**

**The algorithms and corresponding software system that produces the appropriate corporate household knowledge fit for use for the task at hand is called a corporate household knowledge processor.**

With example problems and conceptual definitions described, we next explore solution approaches.

#### **4. Inter-entity Relationships: D&B Family Tree**

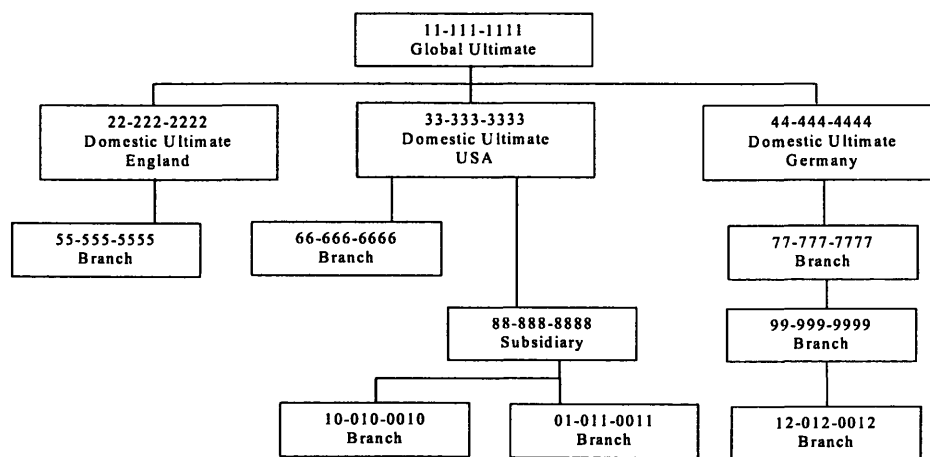
Dun & Bradstreet (D&B) has developed a representation of corporate structure. D&B's Data Universal Numbering System [3], D-U-N-S Number, is a unique nine-digit non-indicative identification number assigned to every business entity in D&B's databases. It widely used for keeping track of millions of corporate family structures and their relationships worldwide. The D&B Family Tree is comprised of linkages and business relationships. Linkage, in general terms, is the relationship between different companies or specific sites within a corporate family. Linkage occurs in D&B WorldBase when one business location has financial & legal responsibility for another business location. Other types of family relationships may occur but are not linked in the D&B file because the affiliated company has no legal obligation for the debts of the other company, such as businesses affiliated through common officers or situations where one corporation owns a part or minority interest in another (50% or less).

D&B's corporate family tree is structured with eight types of entities (single location subsidiary, headquarters, branch, division, subsidiary, parent, domestic ultimate, and global ultimate) and two types of relationships (branch to headquarter linkage, and subsidiary to parent linkage). Each entity is uniquely identifies by a D-U-N-S number. Finally, there is one more business relationship, the Organization Parent D-U-N-S, which is the top most subsidiary, which reports to the Global Ultimate; it has been identified so that customers who prefer can link multi-level family trees, subsidiaries & branches to the direct reporting parent of the global ultimate.

For the purposes of linking these relationships to define corporate responsibility, each

family member carries up to four D-U-N-S Numbers: (1) its own Case D-U-N-S Number, (2) the next highest level in the family: parent or headquarter D-U-N-S, (3) the highest level within its country: its domestic ultimate D-U-N-S, and (4) its top global ultimate; global ultimate D-U-N-S.

Each record carries a set of linkage elements which help to identify the type of record it is, as well as its relationship to other records in the family tree. They are the Status Code, Subsidiary Code, Hierarchy Code, Dias Code and Global Ultimate D-U-N-S Number. The status code is a one-digit field which identifies a record as Single Location, Headquarters, or Branch. The subsidiary code is a one-digit field which identifies the record as Subsidiary, or Non-Subsidiary. The hierarchy code is a two-digit field which determines the record's relative position in a family tree by indicating its relationship to other records. Global Ultimates have a hierarchy codes of "01", while subsidiaries have a hierarchy code of one greater than their parents', and branches have a hierarchy code equal to their headquarters'. The dias code is a nine-digit field which sorts a corporate family tree into family sequence. The dias code changes each time the linkage file is updated. In general terms, all branches will be listed directly below their headquarters while subsidiaries will be listed directly under their parents. In a situation where a parent/headquarters has both branches and subsidiaries reporting to it, the branches will be listed first, followed by the subsidiaries. Branches are sorted alphabetically by country, while subsidiaries are sorted alphabetically by company name. Figure 2 presents a typical corporate family tree structure.



**Figure 2: An example of corporate family tree**

Although the D&B corporate family structure is useful for financial and legal purposes within the corporate boundary, there are many other types of corporate relationships, such as those in the global manufacturing context. Those corporate relationships extend beyond the corporate boundary, and are fundamental in facilitating business decisions such as how to provide a single global sourcing capability or consolidate manufacturing plants globally.

## 5. Entity Identification: Knowledge Structuring using Subject Matter Experts

Another important aspect of corporate household data is "entity identification." As a

simple example, it would involve recognizing that “MIT,” “M.I.T.,” “Mass Inst of Tech,” and “Massachusetts Institute of Technology” are all names referring to the exact same entity. In this section, we will review an approach that was developed by Firstlogic, Inc. [4]. The approach is used when an organization must structure its data to represent the desired abstract view, be it a risk aggregation context, supplier context, global customer context, etc. Often a CRM or ERM system is being used to host the data. It is often the initial load/migration operation into the CRM system that provides the impetus.

No matter the type of householding, be it residential-marketing or corporate-financial, a Subject Matter Expert (SME) approach can be used to help identify and build hierarchical structures to represent relationships between two families. The “family” can be either a two-person residential household or a bank with eight hundred legal entities. In any case the organization engaged in householding is seeking to either: (a) identify the entities in their own family structure, the **internal view**, or (b) identify the entities in the family structure of their business target, the **external view**.

The SME approach is applied when an organization attempting to build the external view has data in its computer systems representing the business relationship, has SME’s that are knowledgeable of both the data and corporate goals, but has no architecture in place that contains and represents the data according to desired abstract view of the organization.

What follows is an example corporate householding project in which a Firstlogic consultant assisted a global financial software (GFS) vendor to:

- a. Allow GFS a single view of its customers
- b. Allow GFS to identify the relationships, at the entity level, between its own corporate tree and those of its top corporate customers.
- c. Allow GFS to plan strategies and marketing campaigns to more effectively leverage the relationships between it and it’s top customers.

We will use the example of GFS to draw out and highlight the step-by-step corporate householding entity identification approach employed. The approach is used to first extract the true business goals, then desired views, rules, and then superimpose those rules on the data. Ultimately, the desired representation is achieved by designing a schema that fits the targeted information repository which, in the case of GFS, was a CRM system.

**Step 1 – Establish project goals.** In any project it is crucial to first establish the goals. In the case of GFS the goal was to achieve a single view of their current corporate customers. Unlike some corporate householding operations, GFS was not interested in consolidating data up and down the organizational tree of their customers, but was instead solely interested in consolidating customer accounts across each branch in the hierarchical structure.

While GFS had data and SME knowledge concerning the parent-child relationships of their customers, their customer purchasing decisions were not driven from the top corporate parent, but mostly from the divisional headquarters. For example, GFS knew that IBM was the parent of Lotus and Informix, but Lotus and Informix from GFS’s perspective retained their own purchasing decisions, at least at the price and volume levels at which GFS sold their product. Consequently even though GFS had the data to vertically rollup the corporate structure of IBM, for example, the purchasing patterns of their customers dictated that they instead consolidate

across each level or branch of the IBM corporate tree. Thus their business rules, represented by the match and consolidation rules, focused on obtaining a single view of the purchasing patterns of Informix or Lotus, etc. But, it was still important, and a challenge, to recognize when two instances, such as “Lotus” and “Lotus Development Corp,” were, in fact, Lotus, for example.

**Step 2 - Define applicable terms and gain cross-functional agreement.** For GFS this meant defining a customer, a contact, a confidence level, etc. The terms had different meanings to different people within GFS. In order for the project to move forward under a common understanding, everyone had to at least agree on the project lexicon. Standardization of meta data definitions allows for uniform queries and reporting of information across the enterprise, in addition to identifying data anomalies. It was often found that when people disagreed on a definition, the opposing parties are really saying the same thing just in a different way, and are reluctant to give up the semantic tug of war. Having an objective third party can help steer the participants back to the corporate goals.

**Step 3 - Define the business rules that attain the goals.** In the case of GFS this meant writing the rules out in English and confirming with a cross functional team that the rules supported the organizational goals. The written business rules included:

- Rules identifying duplicates
- Rules identifying duplicate record confidence levels
- Examples of acceptable duplicates
- Examples of confidence levels
- Rules governing consolidation logic
- Examples of acceptable consolidated records
- Process flow. Sequential matching steps needed for multi-level matching
- Special field level consolidation logic. In the case where two company records are to be consolidated, but individual fields contain different or opposing contents, such as phone numbers or account numbers, what should be done.

When the rules were written out significant debate often occurred concerning the language, especially syntax, of the rule. Another point of debate regarded identifying duplicate or redundant rules. For example, GFS submitted the following three rules that identified a duplicate contact person at a corporate client.

1. Duplicate contact = same name at same address
2. Duplicate contact = same name, address, and e-mail
3. Duplicate contact = same name and e-mail

In the three rules above, rule #1 will identify as duplicate all records with the same contact name and address. For those contact records that have the same name, possibly *different* address, but same e-mail, rule #3 will identify duplicates. “Associating” the duplicates found by rule #1 and #3 via contact name will create a complete set of duplicate contacts across the data set. Rule #2 is redundant. While it is the more specific rule, it was too specific for the contact record patterns found in GFS’s data. Rule #2 would identify duplicates, but always less than Rule #1 or #3. As show below:



<u>Name</u>	<u>Address</u>	<u>E-mail</u>
1. Jon Smith	100 2 <sup>nd</sup> St	Smith@abc.com
2. John Smith	100 2 <sup>nd</sup> St	Smith@Abacuscomputers.com
3. J Smith	100 Front St	Smith@abc.com

Within the matching rules defined by GFS, “Jon”, “John”, and “J.” are allowed as the same first name as long as the contacts had the same exact last name. The “2<sup>nd</sup> Street” and “Front Street” addresses are actually the same physical address with the Front St address being a “prestige” address. The “Abacus Computers” e-mail address was derived from the original company name before it was changed to “ABC”. Matching rule #2 would not have identified any of the three records as duplicates. Only the combination of the two more general rules #1 and #3, and the joining (associating) of results via name could duplicates be identified with confidence at or above the thresholds set by GFS. It took a bit of explaining by the consultant to convince the GFS SMEs that rule #2 was not needed, and that it would slow the matching process with no benefit.

The process of writing the rules in English forces the SME’s to think in concrete terms what they want to achieve and how. When the rules are placed on paper the text will often fail to match what the SMEs thought they meant. The crafting of written rules is an important test. If a written rule can not be agreed upon by the parties involved, this is an important indication that either the rule is unfinished or there is a fundamental misunderstanding of the goal driving the rule. In either case, the conflict must be resolved if the desired single view of the customer, according to the applicable context, is to be obtained.

**Step 4 – Create Rule Matrix.** Once the written business rules are agreed upon and signed-off by all of the applicable personnel the match and consolidation rules are stored in a rules matrix. The purpose of a rules matrix is to gather into one table all the householding rules. The business rules will initially be recorded in a project plan, a requirements document, a statement of work, or some other document provided by the client, and can be embedded in many pages of text. Extracting the match and consolidation rules into one matrix makes it easier to view the entire body in context and further evaluate the rules for redundancy, or the existence of flaws in the householding logic.

**Step 5 – Verify Rules Matrix.** The match and consolidation matrix must be verified and accepted by the SME’s. Redundant rules are marked for deletion and new rules to be added are highlighted.

**Step 6 – Create Application Parameters.** The match and consolidation rules in the matrix must be converted to application parameters. In the case of GFS, Firstlogic’s Information Quality Suite was used to perform the match and consolidation. The consultant loaded the rules matrix into the IQ Suite via the application’s graphical user interface. Each row in the matrix represented a set of match or consolidation criteria that the IQ Suite accepted as job control parameters. The complete translation process from business goals to executable match and consolidation criteria of GFS’s corporate householding project is depicted in Figure 3 below.

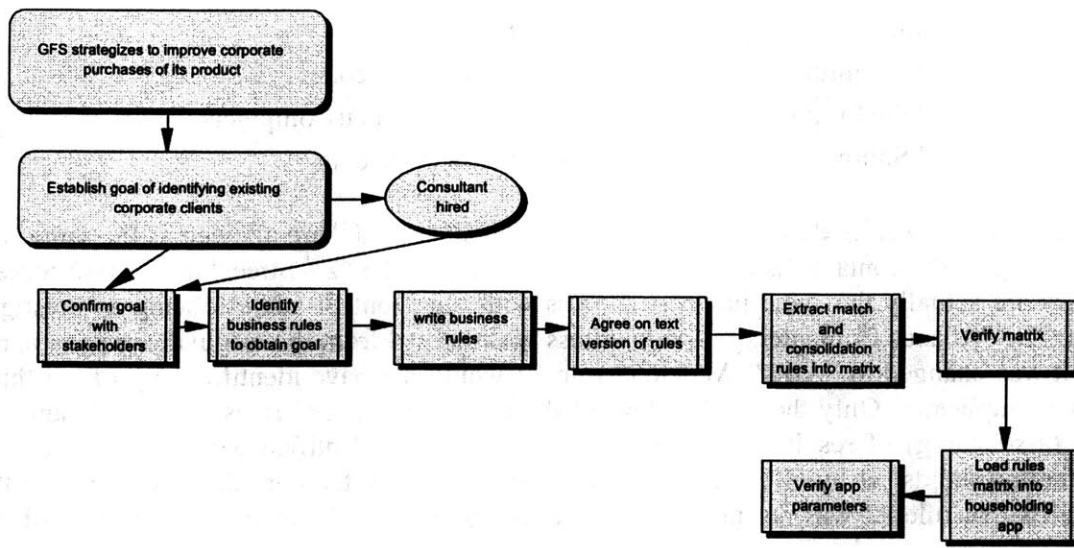


Figure 3. The pathway from corporate goals to application parameters

**Step 7 – Check rules for cohesion.** Using the job file verification feature of Firstlogic’s IQ Suite, the rules are checked for cohesion and operability. Once all warning and error messages were cleared each individual match rule was tested independently of the other rules and the results of the match run approved by GFS. It was during this phase that GFS tweaked their business rules via editing the match parameters as the results of the individual runs highlighted singularities in GFS’s data they were unaware of. Specifically, GFS adjusted the match criteria of firm names, and added *custom* firm names to the IQ Suite firm name parser to allow for the wide distribution of firm name data entry. For example, “Global Scientific” and “Global Sciences Accounting Dept” were entered as valid firm names for the same company. Only after *each* match criteria was verified for expected behavior against live (albeit duplicated) data, would the criteria be signed-off and the next criteria tested.

**Step 8 – Verify entire Match Sets.** Upon complete testing of each match criteria, the entire match sets (combinations of match criteria) were run and the results verified.

**Step 9 – Verify consolidation criteria.** In the process of consolidating duplicate records the opportunity arose to adjust the high-level business rules and even consider modifications to data structures. While match criteria identified records as duplicates, crucial differences existed in field-level data. For example the root corporate records had two child address records one for shipping and one for billing. The marketing department of GFS had been using the billing address record to store the location of the data center where the software was to be used. In reconciling the issue, a third child address record was created, so that there would be one record for each address context: billing, shipping, and data center.

**Step 10 – Run Household Process.** After all consolidation criteria were tested and confirmed for proper operation, the entire householding process was run.

**Step 11 – Completion.** In the case of GFS the data set resulting from the match and consolidation project was used as the initial load into a new CRM system. The investment of conducting essentially a “pre-cleansing” operation prior to loading the CRM package paid

dividends immediately as there were fewer records to load, the records schema had been modified and verified, and the atomic data fields (address, phone number, names) had been cleansed. The pre-cleansing heightened the quality reputation of the information in the CRM system which encourage and accelerated broad use of the system.

In completing the householding project, GFS gained a single view of the purchasing patterns of each self-contained corporate entity at the granularity dictated by their business rules. The approach employed by the consultant in conjunction with the goals and business rules of GFS accomplished the earlier assertion that a good household structure should be able to analyze individual level data within the household environment, and have a structure capable of supporting the various individual-level data as demanded by the context. The three address contexts (shipping, billing, data center) being an example. While the over arching context of GFS's householding project was the vendor-corporate client relationship, the methods employed by GFS would work for any context be it vendor – supplier, risk aggregation, residential marketing, etc.

## **6. Reasoning with Context Knowledge: Context Interchange**

The importance of context in interpreting information has been considered in other research. As a simple example, one source of information might provide length information measured in “meters”, yet the user might require or expect length information in “feet.” How can the “contexts” of these different parties be reconciled? The COntext INterchange (COIN) project [5] has addressed these needs through a mediation approach for semantic integration of disparate information sources. The set of Context Mediation Services comprises a Context Mediator, a Query Optimizer, and a Query Executioner. The Context Mediator is in charge of the identification and resolution of potential semantic conflicts induced by a query. This automatic detection and reconciliation of conflicts present in different information sources is made possible by general knowledge of the underlying application domain, as well as informational content and implicit assumptions associated to the receivers and sources. These bodies of declarative knowledge are represented in the form of a domain model, a set of elevation axioms, and a set of context theories respectively.

The COIN approach allows queries to the sources to be mediated, i.e., semantic conflicts to be identified and solved by a context mediator through comparison of contexts associated with the sources and receivers concerned by the queries. It only requires the minimum adoption of a common Domain Model, which defines the domain of discourse of the application. The result of the mediation is a mediated query. To retrieve the data from the disparate information sources, the mediated query is then transformed into a query execution plan, which is optimized, taking into account the topology of the network of sources and their capabilities. The plan is then executed to retrieve the data from the various sources; results are composed as a message, and sent to the receiver.

The knowledge needed for integration is formally modeled in a COIN framework as depicted in Figure 4. The COIN framework is a mathematical structure offering a sound foundation for the realization of the Context Interchange strategy. The COIN framework comprises a data model and a language, called COINL, and is used to define the different elements needed to implement the strategy in a given application:

- The Domain Model is a collection of rich types (semantic types) defining the domain of discourse for the integration strategy (e.g., “Length”);
- Elevation Axioms for each source identify the semantic objects (instances of semantic types) corresponding to source data elements and define integrity constraints specifying general properties of the sources;
- Context Definitions define the different interpretations of the semantic objects in the different sources or from a receiver's point of view (e.g., “Length” might be expressed in “Feet” or “Meters”).

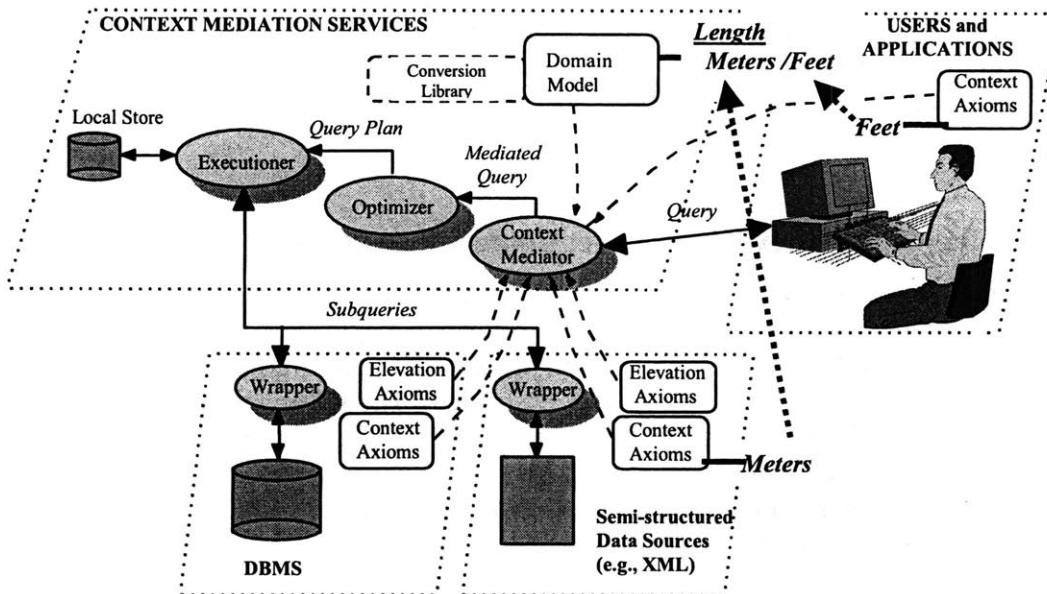


Figure 4. The Architecture of the Context Interchange System

Finally, there is a conversion library which provides conversion functions for each modifier to define the resolution of potential conflicts. The relevant conversion functions are gathered and composed during mediation to resolve the conflicts. No global or exhaustive pairwise definition of the conflict resolution procedures is needed. Both the query to be mediated and the COINL program are combined into a definite logic program (a set of Horn clauses) where the translation of the query is a goal. The mediation is performed by an abductive procedure, which infers from the query and the COINL programs a reformulation of the initial query in the terms of the component sources. The abductive procedure makes use of the integrity constraints in a constraint propagation phase, which has the effect of a semantic query optimization. For instance, logically inconsistent rewritten queries are rejected, rewritten queries containing redundant information are simplified, and rewritten queries are augmented with auxiliary information.

## **7. Research Plan**

Our plan is twofold: (1) Continue our qualitative research to document the various different sources (such as the D&B example), views, and purposes for corporate household knowledge, including the rules used in each case (such as the Firstlogic example); (2) Extend the context interchange framework to be able to represent the corporate household knowledge and rules and enable the context mediation technology to be able to correctly understand and reason about both the context of the sources and the context of the user's query.

Thus, when questions, such as "How many employees does IBM have," are asked, the answer will be the one appropriate to the questioner.

## **8. Conclusion**

Corporate structures and the corporate relationships are changing constantly. The corporate household structure is different under different contexts. Our inquiry in this research has been the understanding of what constitutes a corporate household, how do organizations utilize the concept of corporate household in their business activities, and how they adapt the concept in various tasks. Currently, we are investigating how the concepts and problems that we have identified can be matched with solutions developed in practice, and how research on an innovative technical solution can be developed through extending existing solutions, such as the Context Interchange framework.

**Acknowledgments:** Work reported herein has been supported, in part, by Firstlogic, Inc. and other sponsors of the MIT Total Data Quality Management (TDQM) Research Program.

## References

- [1] Arnold, S., Risk Managers. *Strategic Finance*, 81(12) 2000, pp. 60-64.
- [2] Ballou, D. P., R. Y. Wang, H. Pazer and G. K. Tayi, Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, 44(4) 1998, pp. 462-484.
- [3] Dun & Bradstreet, A Guide to Working with D&B Family Trees.
- [4] Dravis, F. and J. Barash (2001). *Subject Matter Expert (SME): A methodology for building hierarchical structures*. Firstlogic Inc.
- [5] Goh, C. H., Bressan, N., Madnick, S.E., and Siegel, M.D., Context Interchange: New Features and Formalisms for the Intelligent Integration of Information. *ACM Transactions on Office Information Systems*, 1999.
- [6] Huang, K., Y. Lee and R. Wang, *Quality Information and Knowledge*. Prentice Hall, Upper Saddle River: N.J., 1999.
- [7] Kotler, P., *Marketing Management: Analysis, Planning, Implementation, and Control*. 9th Edition ed. Prentice Hall, 1997.
- [8] Madnick, S., X. Chen, J. Funk and R. Wang. Corporate Household Data: Research Directions. in *Proceedings of AMCIS 2001*. Boston, Massachusetts, 2001.
- [9] Putnam, L., The American Keiretsu: America's New Competitive Advantage. *American Business Review*, 16(1) 1998, pp. 113-120.
- [10] Redman, T., The Impact of poor Data Quality on the typical enterprise. *Communications of the ACM*, 41(2) 1998, pp. 79-82.
- [11] Redman, T. C., ed. *Data Quality for the Information Age*. 1996, Artech House: Boston, MA.
- [12] Shankaranarayan, G., R. Y. Wang and M. Ziad. Modeling the Manufacture of an Information Product with IP-MAP. in *Proceedings of Conference on Information Quality*. Massachusetts Institute of Technology: pp. 1-16, 2000.
- [13] Shull, B., The Right Corporate Structure for Expanded Bank Activities. *The Banking Law Journal*, 115(4) 1998, pp. 65-96.
- [14] Strong, D. M., Y. W. Lee and R. Y. Wang, Data Quality in Context. *Communications of the ACM*, 40(5) 1997, pp. 103-110.
- [15] Wand, Y. and R. Y. Wang, Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, 39(11) 1996, pp. 86-95.
- [16] Wang, R., M. Ziad and Y. Lee, *Data Quality*. Advances in Database Systems, ed. A. K. Elmagarmid. Kluwer Academic Publishers, Norwell, Massachusetts, 2001.
- [17] Wang, R. Y., V. C. Storey and C. P. Firth, A Framework for Analysis of Data Quality Research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4) 1995, pp. 623-640.
- [18] Zeng, D. (2001). *Analysis Of XML and COIN as Solutions for Data Heterogeneity in Insurance System Integration*. MIT.