A FRAMEWORK FOR CORPORATE HOUSEHOLDING

(Research-in-Progress)

Stuart Madnick
M.I.T, Sloan School of Management and School of Engineering
smadnick@mit.edu

Richard Wang
MIT & U.C. Berkeley
rwang@mit.edu

Wei Zhang
Boston University, School of Management
zhangwx@bu.edu

Abstract Previous research on corporate household and corporate householding has presented examples, literature review, and working definitions. In this paper, we first improve our understanding of the area by developing a typology of corporate householding tasks and knowledge requirements. We stress the importance of "context" for use to get the appropriate answer for each task. Then we propose an approach to handle these tasks using Communities of Practice (COP) for knowledge acquisition and extensions to the COntext INterchange (COIN) technology for knowledge storage and knowledge processing.

Key Words: Corporate Householding, COntext INterchange, Communities of Practice, Knowledge Management

1. INTRODUCTION

The rapidly changing business environment has witnessed widespread and rapid changes in corporate structure and corporate relationships. Regulations, deregulations, acquisitions, consolidations, mergers, spin-offs, strategic alliances, partnerships, joint ventures, new regional headquarters, new branches, bankruptcies, franchises … all these make understanding corporate relationships an intimidating job. Moreover, the same two corporation entities may relate to each other very differently when marketing is concerned than when auditing is concerned. That is, interpreting corporate structure and corporate relationships depends on the task at hand.

Lets us consider some typical, simple, but important questions that an organization, such as IBM or MIT, might have about its relationships:

[MIT]: "How much did we buy from IBM this year?"
[IBM]: "How much did we sell to MIT this year?"

The first question frequently arises in the Procurement and Purchasing departments of many companies, as well as at more strategic levels. The second question frequently arises in the Marketing departments of many companies and is often related to Customer Relationship Management (CRM) efforts, also at more strategic levels.

These types of questions are not limited to manufacturers and physical goods, a financial services company, such as Merrill Lynch, might ask:

[Merrill Lynch]: "How much have we loaned to IBM?"
[IBM]: "How much do we owe Merrill Lynch?"

On the surface, these questions are likely to sound like both important and simple questions to be able to answer. In reality, there are many reasons why they are difficult, as discussed in the next section.

The paper is structured as following. In Section 2, we present a typology of Corporate Householding problems followed by some additional examples. We then review previous research and industry practice in the area and define working terminologies to facilitate future discussions. Then a proposed approach to improving corporate householding is introduced and explained in Section 4. At the end of the paper, we summarize our ideas and indicate our future research.

2. A TYPOLOGY FOR CORPORATE HOUSEHOLDING PROBLEMS

At least three types of challenges must be overcome to answer questions such as the ones illustrated above: (a) identical entity instance identification, (b) entity aggregation, and (c) transparency of inter-entity relationships. These challenges provide a typology for understanding the Corporate Householding issues, as illustrated in Figure 1 and explained below.

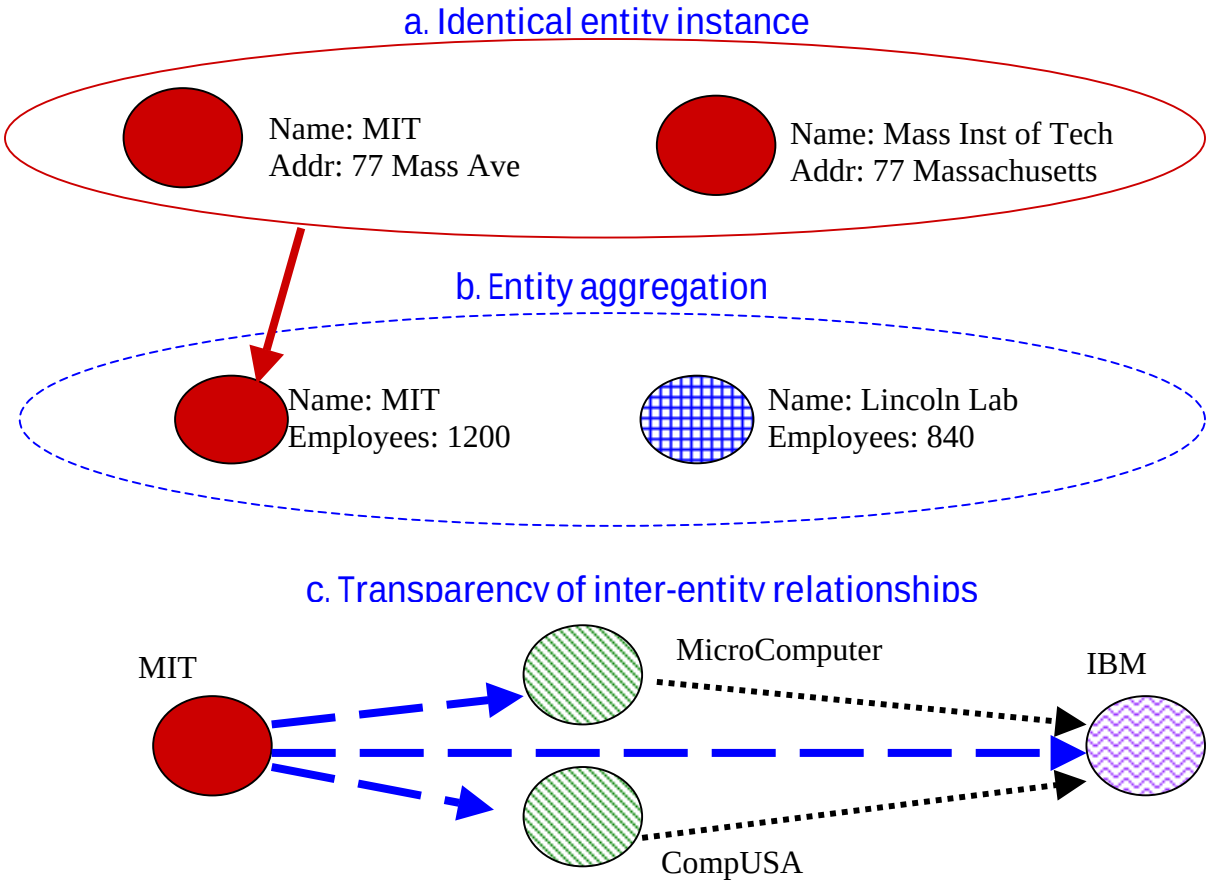

Figure 1. Typology for Corporate Householding

2.1 Identical entity instance identification

In general, there are rarely complete unambiguous universal identifiers for either people or companies. Two names may refer to the same physical entity even though they were not intended to create confusions in the beginning. For example, the names "James Jones", "J. Jones", and "Jim Jones" might appear in different databases, but actually be referring to the same person. Although identifiers such as Social Security numbers are helpful, they might not always be available or feasible. For example, what is the SS# of a French citizen who works in one of IBM's European divisions?

The same problems exist for companies. As shown in Figure 1a, the names "MIT", "Mass Inst of Tech", "Massachusetts Institute of Technology", and many other variations might all be used to refer to the exact same entity. They are different simply because the users of these names choose to do so. Thus, we need to be able to identify the same entity correctly and efficiently when naming confusion happens. We refer to this problem as *Identical Entity Instance Identification* [7]. That is, the same identical entity might appear as multiple instances (i.e., different forms) – but it is still the same entity.

2.2 Entity aggregation

Even after we have determined that "MIT", "Mass Inst of Tech", "Massachusetts Institute of Technology" all refer to the same entity, we need to determine what exactly is that entity? That is, what other unique entities are to be included or aggregated into the intended definition of "MIT." For example, the MIT Lincoln Lab, according to its home page, is "the Federally Funded Research and Development Center of the Massachusetts Institute of Technology." It is located in Lexington and physically separated from the main campus of MIT, sometimes refer to as the "on-campus MIT," which is in Cambridge. Lincoln Lab has a budget of about $500 million, which is about equal to the rest of MIT.

Problem arises when people ask questions such as "How many employees does MIT have?", "How much was MIT's budget last year?", or our original question – for IBM: "How much did we sell to MIT this year?" In the case illustrated in Figure 1b, should the Lincoln Lab employees, budget, or sales be included in the "MIT" calculation and in which cases they should not be? Under some circumstances, the MIT Lincoln Lab number should be included whereas in other circumstances they should not be. We refer to these differing circumstances as *contexts*. To know which case applies under each category of circumstances, we must know the context. We refer to this type of problem as *Entity Aggregation*.

2.3 Transparency of inter-entity relationships

A relationship between entities might involve multiple layers. Under what circumstances should these layers be collapsed? Let us consider our original questions again: [MIT] "How much did we buy from IBM this year?" and [IBM]: "How much did we sell to MIT this year?" As illustrated in Figure 1c, MIT buys computers from IBM both directly and through local computer stores (e.g., MicroCenter and CompUSA). This is the classic case where a seller sells its products to a broker (and maybe directly also), and then the broker sells them to the ultimate buyer. Whether we are interested in the interface between the seller and the broker or the one between the seller and the ultimate buyer (via the broker) also depends upon the context – different answers will be appropriate for different contexts. We refer to this problem as *Transparency of Inter-Entity Relationships*.

2.4 Types of entities and their relationships

In considering the issue of entity aggregation, we need to consider what types of "corporate" entities exist and their relationships. There are obvious examples based on location (e.g., *branches*), scope (e.g.,

divisions), and ownership (e.g., *subsidiaries*). Even these may have variations, such as wholly-owned subsidiaries compared with fractional ownership – sometimes 66%, 51%, and 50% ownerships have different special significance regarding entity aggregation in matters of legal control, taxation, accounting, and bankruptcy.

In addition to these obvious types of entities, there are many others that need to be considered, such as *joint ventures*, which also might be fractional. Referring to our example in Figure 1, what type of entity is MIT's Lincoln Lab and how would one define its relationship with the other parts of MIT? Defining the "atoms" of corporate entities is an important part of this research effort.

2.5 Role of context

We have used the term "context" earlier. To put this issue in perspective, consider a traditional family household. As family structures evolve, 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" [4]. 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? **The important point to note is that there is no single "right" answer; the answer depends upon the intended purpose of the question – which is what we mean by the context.**

Similarly, a corporate household 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 is critical and can provide important competitive advantage.

Furthermore, it is important to note that corporate householding often 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" of IBM, should current Lotus employees be counted in a total as of today? Should the Lotus employees in 1990, when it was a separate corporation, be added with the IBM employees of 1990 to make a meaningful comparison? Thus, temporal context often must be considered.

2.6 Wide range of Corporate Householding applications

There is a wide range of examples of Corporate Householding beyond the few examples used to illustrate the framework above. For example, if an agent is to determine a quote for business owner protection insurance for IBM, he must know how many employees IBM has [5]. To do so, he has to figure out what the rules are to decide what entities are part of IBM as far as business owner protection insurance is concerned. Does Lotus Development Corporation, a wholly-owned subsidiary of IBM, fall under the IBM umbrella? Similarly, if MIT buys a company-wide license for a piece of software, such as IBM's Lotus Notes or Microsoft's Office, does that automatically include Lincoln Lab – or not?

The concerns regarding Corporate Householding play an important role in both purchasing, and marketing activities. In a recent interview, an executive from a global agricultural products supplier gave the following example:

> … On the supply side, *MeatSupplier*[1] is a preferred supplier and they offer us volume discounts. They are owned by *MeatCo*, which also has several other subsidiaries. We do not have preferred status with any of

[1] Names are kept confidential at the request of interviewee.

> these other subsidiaries, so in some instances it's important for us to be able to see the total relationship with *MeatCo* …

We have encountered many other specialized applications in discussing these matters with executives. For example, especially in consulting or auditing practices, you might agree with a client to not also do business with one of its competitors – but how is the "client" defined and how are its "competitors" defined?

3. PREVIOUS RESEARCH AND APPROACHES

Let us suppose that all knowledge of corporate entities and relationships can be captured and stored. Suppose further that one could also capture and store knowledge regarding which set of rules for Identical entity instance identification, Entity aggregation, and Transparency of inter-entity relationships to apply for each context It would then be reasonable to argue that the process of corporate householding could be significantly automated. Unfortunately, although practitioners have been addressing such problems in various ways for a very long time, solutions are usually ad-hoc, limited, and not always effective, due to the inherent complexities of such problems. Research towards a general solution in this area has only recently started [5, 6].

3.1 Definitions

The concept of corporate household is more than companies and their subsidiaries. It is aimed to capture the complicated relationships between corporations in today's fast changing business world. In this paper, we adopt the working definitions from [5, 6]:

Corporate household: a group of business units united or regarded united with the corporation, such as suppliers and customers whose relationships with the corporation must be captured, managed, and applied for various purposes.

Corporate household knowledge: the context-sensitive knowledge of corporate household, i.e., which organization units belong to the same corporate household when a certain task is concerned.

Corporate householding: the process of identifying the corporate household.

Corporate householding knowledge: the knowledge developed for performing corporate householding. Corporate householding knowledge is different from corporate household knowledge. The former is more meta-knowledge. It specifies why certain organization units belong to the same corporate household when a certain task is concerned.

3.2 Examples of Commercial Approaches

In this section, we introduce two approaches used by MIT TDQM Corporate Householding industry sponsors: Firstlogic, Inc. and D&B. Their practices represent state-of-the art approaches. Their willingness to share their knowledge and experiences with us is commendable. In this paper, we use their approaches as starting points and explore ways for further advancement.

Firstlogic has focused heavily on the identical entity instance identification aspects of individual householding, such as for mailing list consolidation, but also can be applied to corporate householding [6]. Rather than totally depending on predefined rules or structures, Firstlogic introduces human intelligence into the job through involving *Subject Matter Experts* (SMEs). SMEs are knowledgeable of both the data and the task domains in which the individual households are concerned. They help clients to establish the rules that identify the entities in their own family structure as well as entities in the family structure of their business targets. The involvement of SMEs makes it possible to perform householding

across task domains, though rendering the process time-consuming and expensive. The Firstlogic tools then allow these rules to be applied across the company's databases.

Dun & Bradstreet (D&B) has developed a tree representation of corporate structure [6]. Their Data Universal Numbering System (D-U-N-S number) assigns a unique nine-digit identification number to every business entity in their database. The numbers consist a tree-structured system that is specifically used for financial and legal purpose. It primarily captures two types of relationships between the entities: branch to headquarter and subsidiary to parent.

None of the existing methods is perfect. D&B is the current industry leader in the area. One of the companies that we recently interviewed uses the D&B family tree. They identified some limitations. First, the application domain of the company is much broader than what D&B family tree is intended. In the same interview with GAC mentioned above, the executive identified their application domain as much broader than merely "financial and legal":

> … Typical scenarios where we have seen householding information deliver value include studying potential acquisition targets, key customer visits (what business do we currently conduct and what other potential business could we conduct), supplier negotiations, and strategic (customer) analysis …

Second, D&B family tree cannot always catch up with the fast changes featured in today's volatile business environment. Third, D&B family tree has its own system for identifying corporations, which is not exactly the same as the data in its customers' the internal databases. For instance, IBM may be expressed as "IBM" in D&B but "I.B.M." in one of its customers' internal databases, and International Business Machine in another internal database. To resolve the discrepancies, the customers have to use a third-party software or other means to try to match the entries. Many potential matches have to be manually reviewed. In fact, in this same interview, it was noted that a centralized operations team has to "match to D&B's data, review key customer and supplier hierarchies, and coordinate resolution of issues with D&B." Fourth, the rather static, rigid D&B family tree does not cater to the local expertise in different departments in the business. As the executive realized, different departments (purchasing, sales, and marketing) all have "pockets of knowledge" about specific companies they do business with. The D-U-N-S family tree embeds only corporate household knowledge, but not *corporate householding knowledge*.

4. A CORPORATE HOUSEHOLDING QUERY PROCESSOR

In this section, we propose an approach for Corporate Householding. As shown in Figure 2 we use a *Corporate Householding Query Processor* (CHQP) to answer corporate household queries and to present the results to the user. The core of a CHQP is Corporate Householding Engine (CHE), which stores and makes use of corporate householding knowledge. It handles the incoming query and its context, processes it, interacts with the target database, and translates and integrates the result to be presented to the user.

To illustrate how CHQP works, let's consider again the query "How many employees does IBM have for purposes of business owner protection insurance". A user simply submits the query to CHQP. Upon receiving the query, CHQP invokes CHE, and CHE uses available knowledge stored inside it to find out all business entities that should be considered as part of IBM as far as the task is concerned. Say, for the purpose of simplification, CHE identifies two and only two corporate entities that should be considered as constituting IBM for the domain task: International Business Machines and Lotus Development Corporation. The CHE could generate two individual queries, one for each of the two corporate entities. The two queries are then executed on the target databases and results fed back into CHE. CHE finally

translates the data, if necessary, and integrates the results from each query and present the final results to the user.

There are many advantages associated with the proposed CHQP architecture. In particular, it moves much of the burden of routine corporate householding from the users to the *Corporate Householding Engine (CHE)*. Users can conduct corporate household queries more easily, more efficiently, and more effectively.

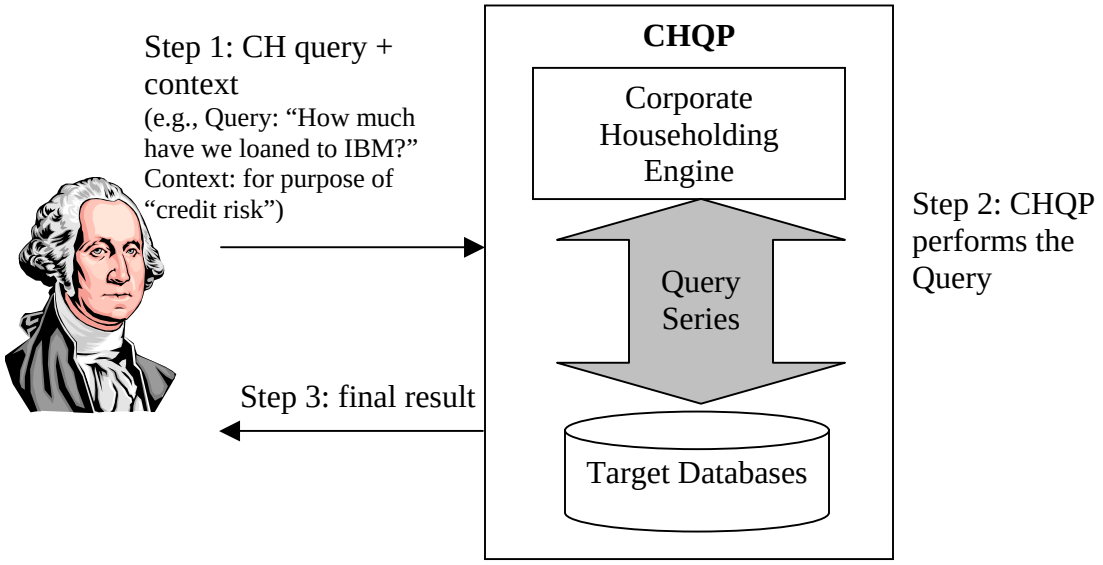

Figure 2: Corporate Householding Query Processor

Below we briefly illustrate a way to implement CHE using Communities of Practice (COP) for knowledge acquisition and extensions to the COntext INterchange (COIN) technology for knowledge storage and knowledge processing.

4.1 Using Communities of Practice (COP) to acquire needed Corporate Householding knowledge

Previous research and practice in general view corporate household problems with an information perspective. Take the D-U-N-S number tree as an example. In the tree, each business entity is uniquely identified by a D-U-N-S number, and carries the D-U-N-S number of its parent entity. With the D&B tree, in some cases, corporate householding is reduced merely to matching business unit names with D&B database records, and finding the links between D-U-N-S records. As we discussed before, such a simplified information view doesn't cover all the complexity of corporate householding. To deal with the problem more effectively, we need a knowledge perspective – we need to know which rules to use in which cases. In another word, it is not enough to simply store and apply the limited amount of corporate household knowledge which may be represented explicitly [8] with a D-U-N-S number tree. What is important is more complete corporate household*ing* knowledge, the rationale underlying the identified corporate relationships to know which set of rules to use in each case.

We reincarnate corporate householding as the process of "*acquiring, managing and applying knowledge for identifying corporate household*", and we look for mechanisms that can facilitate the knowledge acquisition process. Communities of practice (COPs) are informal communities formed by voluntary

members to share individual experiences and knowledge about a common course [9]. Research has shown that they can be valuable tools for organizational knowledge management. In these communities, individual experiences are shared, new knowledge is created, and problems are solved through interactions between community members, sometimes even without the knowledge of the formal organization to which the members belong [1].

Given that COPs are demonstrated to be valuable for knowledge creating and knowledge sharing between organization members, we propose to introduce the concept into corporate householding. Data collected through the earlier aforementioned interview suggests that such a Community of Corporate Householding exist. The interview was with an executive who supervises corporate householding practice in a global agricultural company (GAC). GAC defines corporate householding as "understanding the relationships between the companies we do business with", and acknowledges the importance of corporate householding. In GAC, a centralized operation team matches to D&B's data, reviews key customer and supplier hierarchies, and coordinates resolution of issues with D&B. Some internal experts from purchasing, sales, and marketing also perform corporate householding. They all have pockets of knowledge about specific companies they do business with. Members of the centralized operation team and the internal experts could be the members of a potential Community of Corporate Householding (CCH). Although there is no evidence that this community demonstrates a unique collective identity, each (potential) member is recognized for his/her corporate householding expertise.

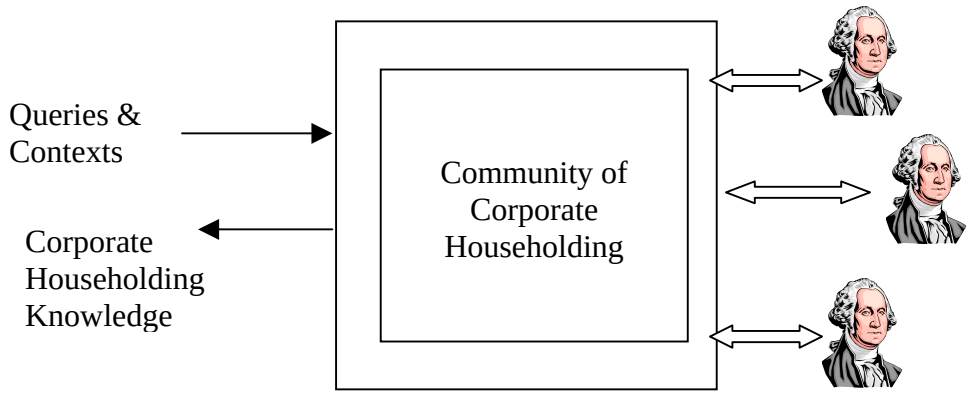

Figure 3: Using Community of Practice (COP) to acquire Corporate Householding knowledge

While more data is needed to reach a conclusion, we feel comfortable to state that a community of corporate householding is emerging in GAC. Potentially, the members of the community can share with each other their individual experiences and help with each other to solve real-world problems [2]. Recognizing its emergence and facilitate its growth would help GAC to collect local corporate householding knowledge globally in the organization, thus improving its corporate householding practice.

In some cases, especially one-time events or very subtle situations, the process depicted in Figure 3 may be most appropriate and sufficient to resolve corporate householding questions. On the other hand, in those situations where similar questions are frequent and the necessary corporate householding knowledge is stable and tacit, but large and distributed, a systematic means of storing the knowledge and processing it as needed becomes important, as explained in the next section.

4.2 Using COntext Interchange (COIN) for storage and processing of Corporate Householding knowledge

COntext INterchange (COIN) [3] is a knowledge-based mediation technology that enables meaningful use of heterogeneous databases where there are semantic differences. For example, attributes that represent money, such as "price", may be expressed in "US dollars" in a USA database but in "Chinese RMB" in a Chinese database. Though the two attributes may have the same name, the semantic conflict has to be addressed before a correct query result involving the attributes can be obtained (e.g., "which price is less expensive?"). We refer to these semantic meanings as being the "context" of each source or *source context*. Furthermore, different users may have different contexts or *receiver contexts* (e.g., and I might want the answer in "Euros"). There are many parallels between the traditional COIN applications and the needs of Corporate Housekeeping where each source has its own Corporation Housekeeping context (e.g., "in this database, data on IBM includes Lotus") and each user's query has a context (e.g., "employee count for liability insurance purposes.")

The COIN approach deals with the problem by providing *Context Mediation Service*, with the help from three components (Figure 4). The *Domain Model* defines the application domain corresponding to the data sources that are to be queried. The *Elevation Axioms* identify the correspondence between attributes in the source and semantic types in the Domain Model. The *Context Axioms* define alternative interpretations of the semantic objects in different contexts. The three components work together to enable the Context Mediator to generate the correct mediated query from the original user query. The research in Context Interchange and COIN framework has been ongoing for some time and a prototype system has been reported to validate the method [3].

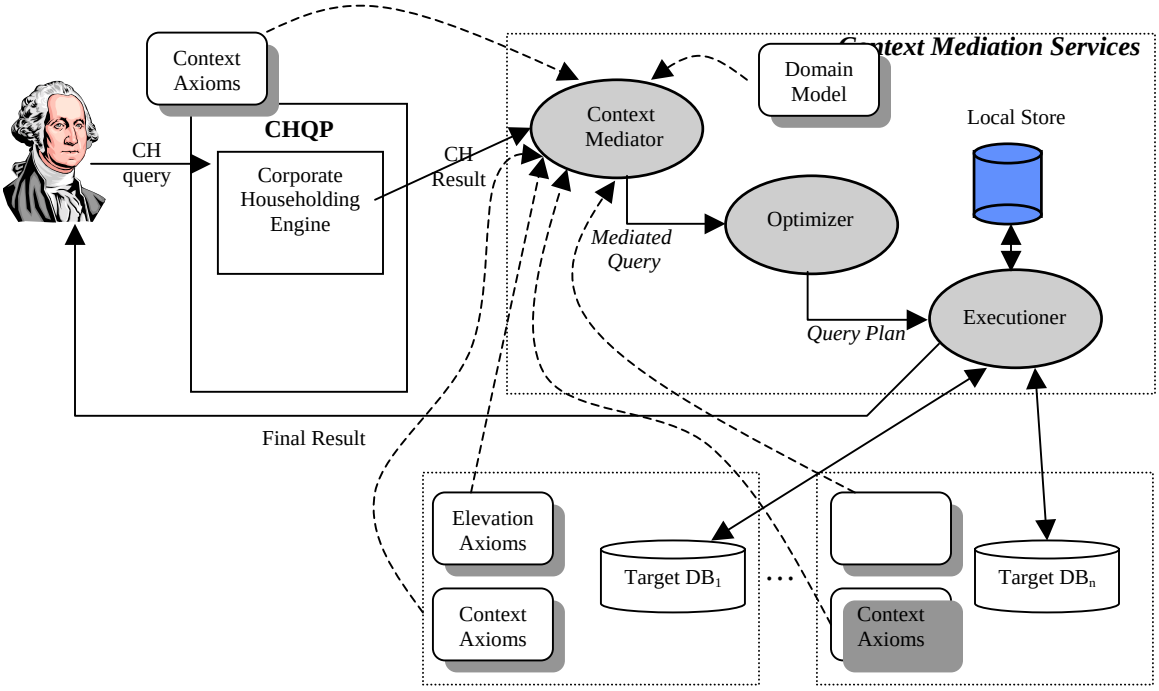

Figure 4: Using Context INterchange (COIN) for Corporate Householding knowledge

COIN was designed originally to address database queries in the face of disparate semantics in different sources. We believe that it can be applied to corporate householding, which – in a certain sense – is to determine which attributes in different databases should be united or viewed as the same. Figure 4 shows

a possible way to implement CHE processing with COIN. In the implementation, the Domain Model stores general corporate householding knowledge. It decides how the relationships between entity instances should be decided when a certain task is concerned. The Elevation Axioms and Context Axioms describe the context associated with each specific database and specific application. The Context Mediator manages the interactions between Domain Model, Elevation Axioms, and Context Axioms. It is the interactions between the three that determine how the data stored in a database can be interpreted in terms of corporate household.

Such an implementation would make it much easier to answer questions such as "What is IBM's total global asset worth for purposes of bankruptcy insurance?", which involves both corporate householding knowledge processing and data semantics knowledge processing.

5. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we present a framework for understanding corporate householding problems. We then proposed that much of the burden of corporate householding could be reduced through the use of a corporate householding engine. We proposed an integrated method to accomplish the goal: using the concept of Communities of Practice to capture existing corporate householding knowledge even though it is distributed across organization departments, and using COntext INterchange (COIN) to store and apply the captured knowledge. COIN builds on previous research, and is intended to maximally automate corporate householding with specially designed software modular – once corporate household knowledge has been acquired. With the concept of Communities of Practice, we address issues related to acquiring corporate householding knowledge.

While we discuss the two steps separately and sequentially, we should emphasize the interactions between the two. With the knowledge captured from the Community of Corporate Householding (CCH), which is stored in COIN, a user can reduce his routine workload by delegating many problems to software based solution; and can spend more time participating in the CCH to create and share corporate householding knowledge with other community members, including exchanging individual experiences on using COIN. The knowledge newly generated, externalized, and verified in a CCH can then be represented and stored in COIN. In other words, the community uses COIN as an effective tool in their corporate householding practice, and the members collectively improve both the tool itself and the use of the tools. In this way, COIN and CCH enhance each other and both grow iteratively.

Our future research plans include the following. First, we will continue to collect field data to determine the types of corporate householding knowledge needed and identify communities of corporate householding. After it is confirmed that a COP exists, we will have a closer look at the identified community to study in depth its role in creating, preserving, and sharing corporate householding knowledge. We are especially interested in the interaction between the knowledge and member behaviors. Moreover, we will explore the role of technologies in empowering and enhancing the communities. For example, is it possible to connect the community members with computer-mediate communication without incurring any advent consequence to the practice?

Second, we will explore the role of COIN in corporate householding. We plan to develop a COIN-based system that facilitates the process of capturing, storing, maintaining, and applying the corporate householding knowledge. In the near future, our target is to develop a specialized corporate householding system that can demonstrate the efficacy of such a system in a particular application domain.

ACKNOWLEDGEMENTS

Work reported herein was supported, in part, by the MIT Total Data Quality Management (TDQM) Program, D&B, Firstlogic Inc., Citibank, Merrill Lynch, MITRE Corp., PricewaterhouseCoopers, Suruga Bank, and the Singapore-MIT Alliance. We also wish to thank the participants of the MIT Summer Data Quality course (15.56s) and the MIT workshops on corporate householding for their helpful feedback and comments on the ideas that lead to the current paper.

REFERENCES

[1] Brown, J.S. and P. Duguid, *Organizational learning and communities of practice: toward a unified view of working, learning, and innovation.* Organization Science, 1991. 2(1): p. 40-57.

[2] Constant, D., L. Sproull, and S. Kiesler, *The kindness of strangers: The usefulness of electronic weak ties for technical advice.* Organizational Science, 1996. 7(2): p. 119-135.

[3] Goh, C.H., et al., *Context Interchange: New Features and Formalisms for the Intelligent Integration of Information.* ACM Transactions on Information Systems, 1999. 17(3): p. 270-293.

[4] Kotler, P., *Marketing Management: Analysis, Planning, Implementation, and Control*. 9th ed. 1997: Prentice Hall.

[5] Madnick, S., et al. *Corporate Household Data: Research Directions.* in *AMCIS 2001*. 2001. Boston, Massachusetts.

[6] Madnick, S., et al. *Improving the Quality of Corporate Household Data: Current Practices and Research Directions.* in *Sixth International Conference on Information Quality*. 2001. Cambridge, MA.

[7] Madnick, S. and R. Wang, *The Inter-Database Instance Identification Problem in Integrating Autonomous Systems*. in *Fifth International Data Engineering Conference* February 1989. Los Angeles, CA.

[8] Nonaka, I., *A Dynamic Theory of Organizational Knowledge Creation.* Organization Science, 1994. 5(1): p. 14-37.

[9] Wenger, E.C., *Communities of practice: The organizational frontier.* Harvard Business Review, 2000. 78(1): p. 139 - 144.