

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

Stuart Madnick (smadnick@mit.edu),
Richard Wang (rwang@mit.edu)
Frank Dravis (frankd@firstlogic.com),
Xinping Chen (Xinping_Chen@bose.com),

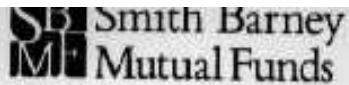
Total Data Quality Management (TDQM) Program

Sloan School of Management
Massachusetts Institute of Technology

Conference on Information Quality (IQ 2001)

<http://web.mit.edu/tdqm>

November 3, 2001; Cambridge, Massachusetts



Your Serious Money. Professionally Managed.™

Important Notice Regarding Delivery of Prospectuses

Dear Shareholder,

In an effort to reduce printing and mailing expenses for the Smith Barney Mutual Fund family, we have adopted a policy known as "householding." Starting in September, shareholders who are members of the same family, share the same address and have multiple accounts in the same Smith Barney Mutual Fund will receive only one copy of a fund's annual prospectus. (Each fund currently mails only one shareholder report to a household if more than one account has the same address.) Please note that this new householding policy does not apply to the account statements you receive. Each shareholder account will continue to receive a separate account statement.

We realize that some of our shareholders may prefer to continue receiving separate mailings for their mutual funds. If for any reason you prefer to receive separate copies of shareholder documents, call Smith Barney Shareholder Services at 1-800-451-2010, and you will be removed from the householding program. We expect that it may take up to 30 days for you to begin receiving separate mailings if you choose that alternative.

Thank you for your continued confidence in Smith Barney Mutual Funds. As always, we welcome your suggestions and look forward to helping serve your financial needs in the future.

Sincerely,

A handwritten signature in dark ink, appearing to read 'Heath B. McLendon'.

Heath B. McLendon

“Household Data”

- **Example: letter from Smith Barney**

“To reduce mailing expense ...

We have adopted a policy known as “householding.”

... shareholders who are members of the same family, share the same address and have multiple accounts in the same Fund will receive only one copy of the annual prospectus.”

- **In general, what is definition of “household”?**
- **What would be corresponding “corporate household”?**

“Household”

Is it:

- **Father, mother and children living at same address?**
- **What if father and mother have different last names?**
- **Father, mother, and children even if living at different addresses?**
- **Include cousin Alice visiting for 6 months?**
- **An unmarried couple living together? (of any sex)**
- **Roommates?**

There can be multiple purposes and contexts with differing answers



Picture of old lady or young lady ?

Corporate Household Example

- **Simple questions:**
 - **“How many employees does IBM have?”**
 - **“How much business do we do with IBM?”**
- **Many reasons for first question:**
 - **Employee casualty insurance**
- **Even definition of “employee” might be ambiguous**
 - **How count part-time employees?**
 - **How count full-time consultants?**
 - **(in university) how count joint appointments or unpaid visiting faculty?**

What is “IBM” ?

What is the relationship among these companies:

- **International Business Machines Corporation**
- **IBM**
- **IBM Microelectronics Division**
- **IBM Global Services**
- **IBM Global Financing**
- **IBM Global Network** (1999-)
- **IBM de Colombia, S.A** (90%)
- **Lotus Development Corporation** (100%)
- **Software Artistry, Inc.** (1998+, 2000-)
- **Dominion Semiconductor Company** (50/50 jv)
- **MiCRUS** (majority jv)
- **Computing-Tabulating-Recording Co.**

Examples of multiple purposes ...

Corporate household /family structure purposes -

- **Financial: Risk (credit - bankruptcy, country)**
- **Accounting: Account consolidation**
- **Marketing (multiple divisions & subsidiaries):**
 - Customers & Supplier consolidation (economies of scale)**
- **Customer Relationship Management (CRM)**
- **Managerial: Regional and/or Product separations**
- **Legal: Liability (insurance)**
- **Relationship: Consultant Conflict of interest & competition**
- **Ad hoc/temp structures**

And these are dynamic ... changing over time

Inter-entity Relationships:

Dun & Bradstreet corporate family tree

Dun & Bradstreet representation of structure as tree with eight types of entities (Each uniquely identified by a D-U-N-S number):

- 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**

Entity Identification: Knowledge Structuring using Subject Matter Experts

Many representations of same entity:

“IBM”, “I.B.M.”, “Computing-Tabulating-Recording Co.”, etc.

Subject matter expert approach, steps:

- 1. Establish project goals**
- 2. Define applicable terms, gain cross-functional agreement**
- 3. Define business rules**
- 4. Create Rule Matrix**
- 5. Verify Rules Matrix**
- 6. Create Application Parameters**
- 7. Check rules for cohesion**
- 8. Verify entire Match Sets**
- 9. Verify consolidation criteria**
- 10. Run Household Process**
- 11. Completion**

General Challenge:

There is not a single right answer

- **As seen with “old woman or young woman” picture**
- **Need to be able to answer question “in context”**
- **This issue appears in many situations ...**

The 1805 Overture

In 1805, the Austrian and Russian Emperors agreed to join forces against Napoleon. The Russians said their forces would be in the field in Bavaria by Oct. 20.

The Austrian staff planned based on that date in the Gregorian calendar. Russia, however, used the ancient Julian calendar, which lagged 10 days behind.

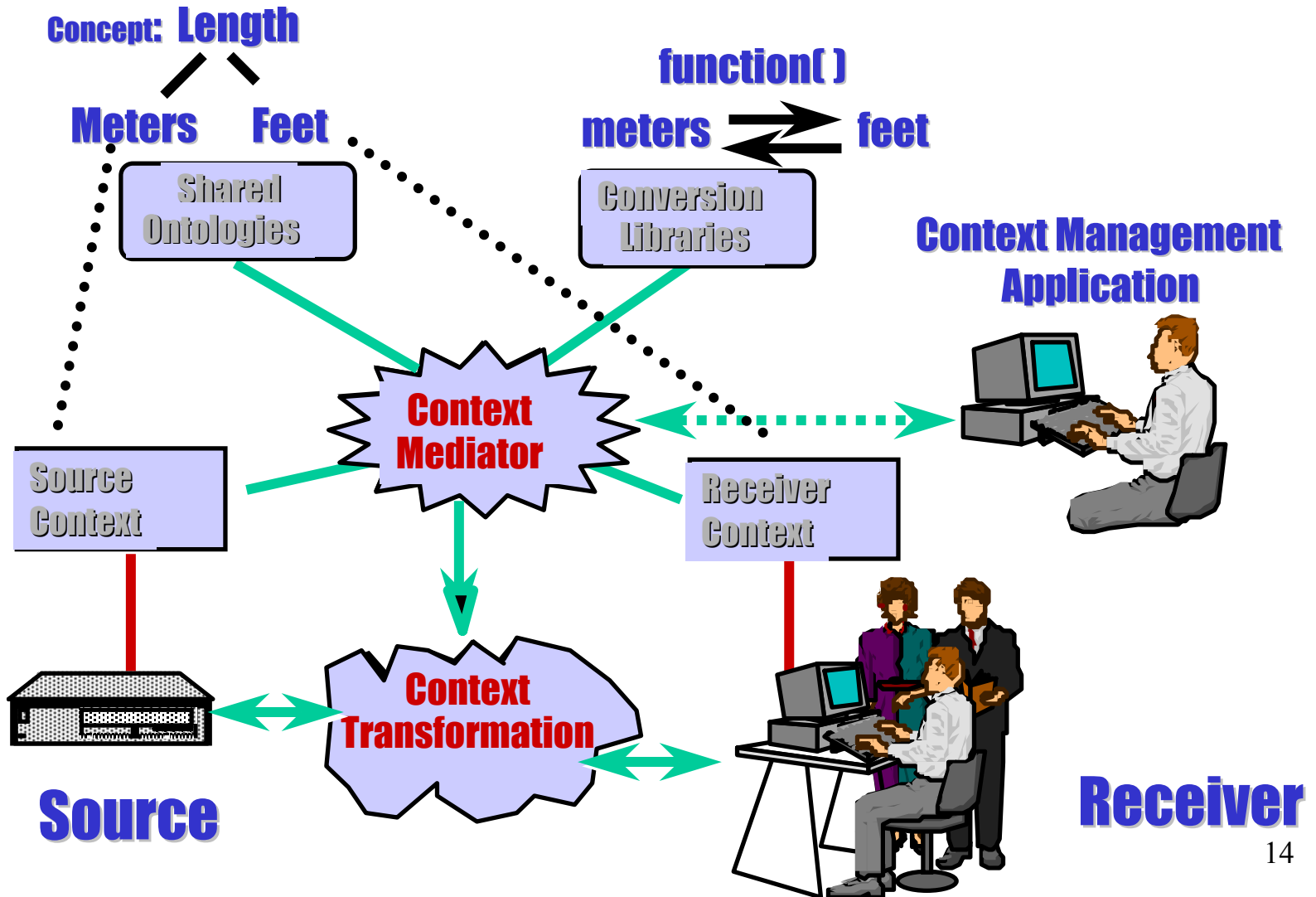
The difference allowed Napoleon to surround Austrian General Mack's army at Ulm on Oct. 21, well before the Russian forces arrived.

Source: David Chandler, *The Campaigns of Napoleon*, New York: MacMillan 1966, pg. 390.

Research Agenda

- **We do not want to build custom solution for each situation**
 - **Need to understand requirements**
 - **Need to understand sources**
- **Process**
 1. **Interviews – describe representative purposes / cases in detail**
 2. **Identify key characteristics (e.g., extend D&B factors)**
 3. **Incorporate into a context knowledge management system**
 4. **Develop reasoning algorithm to map purposes onto data sources automatically (building on prior context mediation work)**

Reasoning with Context Knowledge: Context Interchange Technology



Acknowledgements

Work reported has been supported, in part, by

- **FirstLogic, Inc**
- **MIT Total Data Quality Management (TDQM) Program**

For more information regarding TDQM, go to

<http://web.mit.edu/tdqm>