Silicon Valley
WORLD INTERNET CENTER

SYMPOSIUM

DATA-TO-CASH (D2C): BUSINESS OPPORTUNITIES EMERGING FROM AN OCEAN OF DATA

Tuesday, December 11, 2001

Programs at the Center, including Think Tank Sessions, are possible due to the support of its Sponsors:
Amdocs • Cable & Wireless • Deutsche Telekom • Fujitsu • IBM • SAP • Sun Microsystems

SUMMARY

The Silicon Valley World Internet Center held a Symposium on December 11, 2001, to discuss the factors influencing businesses' ability to turn data into information that can then be used as the knowledge for structuring business decisions – turning data into cash. Following a presentation by SAP, twenty-three participants discussed the current state of data-to-information processes in businesses, what state businesses would like to achieve and the challenges to doing so, as well as possible tools and solutions for those challenges.

Participating in the December 11, 2001, program were representatives from Artloop, Bodha, DCL Logistics, Dejima, iSpheres, KnowNow, Lake Forest Venture Management, Littera, Monterey Network Center, Pacific Cartage and Warehousing, SAP, Stanford Skolar MD, Stanford University, Tony Grant Solutions Sales Consulting, VerticalNet, WebV2, Whizbang!Labs and Wisengine.

Access to relevant information is crucial when making business decisions. Today, more data is available from internal and external sources than ever before, but it needs to be discovered, extracted and turned into information before it can be used for decision support. At this Symposium, SAP Markets' Kaj van de Loo and SAP Labs' Dr. Hartmut Vogler presented two real-world business scenarios. The first depicted how data is currently used to inform business decisions, and the second speculated on how intelligent software agents could help with data gathering as well as in turn data into valuable information. In the current-time scenario, the decision-making process is relatively static. Some data is accessed and assessed as to its accuracy and relevancy; new data comes in, and data is normalized and consolidated. Through these processes, the data becomes information that is used as knowledge about business factors. Knowledge is then applied to business rules and analysis for informing business decisions. Decisions are then made which affect the business. Participants pointed out, however, that most companies currently do not "learn," or analyze back down the process to determine whether or not a decision was a good one, based on the correct assumptions, or whether the information was based on truly accurate or relevant data. Currently, bad business decisions are often repeated in organizations due to the lack a two-way flow. In the second scenario set up by Dr. Vogler, the flow does go back down the chain from the business decision to the data. Was the business decision a successful one? Was it successful in the short term *and* the long term? If it was, then the knowledge, information and data were correctly applied. If the results of the decision were not effective or positive, then the knowledge which informed the decision needs to be reviewed with an eye toward the quality of the information and data (and data sources) from which it stemmed.

CHALLENGES

Data management and analysis is currently a very manual, human endeavor. Participants of the Symposium were challenged to investigate where machines – computers, systems and software – may be applied to aid people in turning data into valuable business decisions. In looking at the broad opportunities and the vast needs in the area of data management, there are several factors that influence the choice of where to begin examining business processes around data and how it is used and quantified. In business today, the heterogeneous landscape is one factor that keeps businesses from being able to access and use data. Because of this variety of systems, the group decided to focus on those processes where existing data and information sources can be integrated most easily between one another and into existing business systems. Integration of electronic data interchange, or EDI, and Internet sources of data was discussed by end users. While one logistics company with several Silicon Valley clients stated that everyone is using the Internet now for data exchange, another logistics executive whose company collects, stores and distributes product from several mid-west-based retailers objected that his clients, indeed, do *not* use the Web for data exchange. In fact, these companies are using both EDI and fax machines to communicate with the logistics company. It was suggested that solutions providers need to find a way to integrate data from such disparate sources

as EDI and the Internet in order to best serve the plethora of businesses which are not yet Internet savvy in their processes.

Another factor the group examined is the cost/benefit analysis of turning data into information. It was noted by more than one participant that turning information into knowledge is currently a very expensive proposition. It is time intensive to look at data and information, evaluate its relevancy and accuracy and report it out as knowledge. The return on the investment, however, is very high when companies are able to save high-margin dollars at high volume with improved decision making. One participant noted that companies that look at the widest range of scenarios make the best business decisions. Currently, however, most businesses do not make the investment necessary to receive the best cost/benefit ratio. Current weak points in data-to-decision processes include lack of centralization of knowledge; lack of identification of the location of data; lack of identification of the keepers of data and knowledge, such as retired employees; corporate ownership of data as exemplified by firewalls; the need for resource discovery; the lack of analysis of the quality of decisions based on the data-information-knowledge processes; and, finally, the need to lower the cost of converting data to information.

Other challenges identified by the group are listed below:

DATA

- Data compatibility
- Data mining
- Normalizing data definitions, especially when the same terms are used for different meanings
- No incentive, time or reword for expert data creators to publish and/or explain their data to others
- Manage the dynamics of change from the input
- Ability to rapidly absorb & transfer data to knowledge
- Data resource publication and discovery
- Lack of standards
- Lack of skill or interest in collaborating data in a manner that it is valuable information to someone other than oneself
- Same data referring to different information
- Change management – new data formats break traditional client-server / DOOP models
- EDI versus the Internet for data transmission
- Closed systems and non-accessible data
- Model Mismatch – Data to Information to Knowledge
- Interpersonal communication facilitation
- Data location – How do data keepers communicate?
- Finding and storing the data – Where is it? We don't collect it in one big place
- Normalization of data

INFORMATION

- Lowering the cost of developing code that tests articulated scenarios
- Existence of data that is required to derive meaningful insights
- Cost of integrating information from different applications
- Information integration
- Cost to convert data to information – Procurement base of large history of information; Details for components to ease quoting future buy
- Ease of access to information
- How much can be automated to lower the cost of data cleansing, normalization and integration
- Integrate information from diverse sources with different terms and schemas
- Cost of data to information
- How to get information in near-real-time without regular polling of the applications where they live
- Data normalization and integration from any number of sources
- Easy, adaptable way to specify information extraction criteria to machines
- Managing overlap around different domains
- Differing taxonomies to represent similar concepts and/or items
- How to integrate feedback!

KNOWLEDGE

- Ontology management
- Developing Joint ontologies (schemas)
- Apply knowledge to the RIGHT data to connect it into information
- Lower the cost of developing code that tests articulated scenarios
- Identification of the high-value (high risk) problems to solve?
- Identify how a human being would solve a given problem and try to capture that process
- Identify the most important (valuable) criteria and solutions – Speed? Quality? Cost?
- Moving away from the relational data base (Sounds like "data" but allows easier associations)
- Interpretation of knowledge
- Publishing standards, taxonomies and access methods
- Centralizing repositories
- Privacy – How to flexibly share only the right level of data, information and knowledge among partners
- Modeling – How to gather data in order to answer a question one does not know

TAXONOMIES, ONTOLOGIES AND THE SEMANTIC WEB

In identifying solutions to many of today's challenges in the data-to-decision process, participants spoke a great deal about taxonomies, ontologies and the emergence of the semantic Web. Webster's dictionary describes ontology as "a branch of metaphysics relating to the nature and relations of being." Symposium participants overlaid that definition on top of the study of data and information and describe ontologies as the entire set of relationships within an enterprise. The ontology is the entire structure of relationships around data. Mapping tools, then, are used within the ontology to map one item's relationship to another. Taxonomies are defined by the participants as the vocabulary used inside of the ontology. Language and terms need to be common among users for the data to pass through the various relationships of the ontology and to mean the same thing to all parties. For instance, an engine part may be described by the manufacturer with a number, by the retailer by a different number and by the customer as a water pump. Allegedly, the semantic Web will be able to recognize all these terms as the same item and find it for all parties. In order to develop common vocabularies, participants saw the need for the build up of central repositories and huge "data dictionaries" to lower the cost of communication and integration and to manage and keep reliable.

Participants brought up several questions about the solutions they proposed. How precise do the vocabularies have to be? How can organic growth be created inside of a vocabulary? They suggested that applications be created that can deal with standardized ways of describing attributes and relationships among autonomic entities. They suggested building a dictionary focused on a few key problems. Other solutions included developing a case-based chooser with a meta layer that is presented to the user; coming up with an explicit way to describe a taxonomy, such as XML schema; developing a tool that can go out on a given network and find a data source; and, finally, coming up with a router that can pull data from a network and provide it to a combination of applications.

Next Steps / Further Topics

Participants felt that the next steps in analyzing the data-to-decision process would be to examine:

- Where and when real-time information is critical
- Some real life examples
- The role of intelligent agents
- The cost of decision making
- Business models for sharing data
- Information-integration architecture
- How to get knowledge published and how to empower the employee to do so
- Human aspects such as change management and culture
- Evolution of the problematic, who actually needs it and what the business value is