Intelligent Data

Edmund W. Schuster
Laboratory for Manufacturing and Productivity - Data Center Program
MIT

At Smart World 2004, Sunil Gupta of SAP paraphrased Samuel Taylor Coleridge by saying "data, data everywhere but not a byte to use."

Each year, the amount of data grows by as much as 40 – 60 % for many organizations. In 2004 alone, shipments of data storage devices equaled four times the space needed to store every word ever spoken during the entire course of human history.

The explosion of storage device sales has gone largely unnoticed because the selling price has decreased sharply during the past few years. While the increase in unit sales has meant larger revenues for companies like EMC, the amount of the increase has not been great enough to draw widespread attention. Since the growth rate appears to be exponential, businesses are likely to face a mounting data organization problem in the years to come. Compounding matters, a large amount of the increase is from unstructured data that are extremely hard to organize using current computer-based approaches. Unstructured data includes images, text, emails, and engineering designs.   In all of these cases, object representation requires more than a serial number that can be stored in a database.

For example, a serial number does not adequately describe data such as an Image or a section of text. It is only through using words in a machine understandable way that descriptions of these objects can become useful for search, organizational, and analytical purposes. Amidst the ever-increasing amounts of data, companies continue to struggle with making better connections between different types of data. Emerging technologies such as the EPCglobal Network and RFID technology, along with sensor networks and the use of "loyalty cards" are certain to generate even more data.

By one estimate, there will be 300 million RFID readers active in supply chains within ten years. In another prediction, the number of deployed sensors will dwarf the number of personal computers by a thousand fold in 2010.   Both of these technologies will boost the amount of raw data available for companies to use in decision-making.

Dealing with the increasing volumes of structured and unstructured data will require new standards and information architectures to improve integration and communication between hardware, software, and business entities. This becomes important as companies seek to overcome the barriers that limit the seamless transfer of data, internal and external to the firm.

Our work at the MIT Laboratory for Manufacturing and Productivity – Data Center Program focuses on new ways to connect data together within firms and across the Internet. The M Language, which is under development, provides the ability to label objects and abstractions with an exact semantic that is machine understandable. This is of great importance in many

ERP systems and is the base for creating intelligence within these systems. Our approach handles context, an important element of creating intelligence, in an innovative way.

Working with a core group of companies that include MorganFranklin Corporation, LG CNS (Korea), Siemens, Raytheon, ReadyTouch, and Ecostat, a number of practical applications of the M Language are actively being pursued in the aerospace and agriculture industries. Release of a key component of the M Language, a technologically advanced dictionary capable of handling context in a machine understandable way, occurred during January 2007. Since the idea of the M Language is to create an open system, the "M Dictionary" will be free to use much like the current case of Wikipedia.

Our next release will be April 15, and include some additions to the M Dictionary along with a prototype interoperable modeling system. An announcement will be posted at http://datacenter.mit.edu.

Data affects all companies. It is an area that few organizations can ignore. Creating intelligent data is a new development that will improve productivity and change the way we think about manufacturing systems and supply chains.

Edmund W. Schuster is co-author of GLOBAL RFID: The Value of the EPCglobal Network for Supply Chain Management, published by Springer Verlag (New York and Berlin).