

Existing unstructured and other nontraditional data will help your company enrich business intelligence, improve analytics capabilities and increase competitiveness.

# Data's Next BIG Adventure

{by Cheryl D. Krivda}

**T**he recipe for optimization in business intelligence (BI) is simple: powerful tools, high-quality operational data and streamlined data-integration processes. Yet this strategy effectively addresses only today's analytics requirements. To maintain a competitive edge in the use of decision-support applications, executives must also keep one eye on the future—watching for new technologies and seeking innovative deployment approaches.

An important innovation coming to the BI field will be the use of unstructured and other nontraditional data in enterprise warehousing. Industry analysts estimate that as much as 85% of an enterprise's data is unstructured and

therefore typically excluded from the data warehouse. Viewed from that perspective, it could be said that most companies are making critical decisions using only 15% of their overall data. Adding unstructured and nontraditional data as a source for BI could greatly enrich the quality of enterprise analytics, improve decision making and help companies discover previously unrecognized insights.

Which types of unstructured or nontraditional data are good candidates for enhancing enterprise analytics? Critical information able to offer such improvements can be harvested from such sources as language data, location-aware applications, sensory measuring devices and multimedia images.

## Talk to me

Language data includes information derived from human communications, such as text documents or voice conversations. Text data resides in such diverse sources as e-mail

### More data, greater insight

How might unstructured data enrich analytics?

Consider a manufacturer that uses a CRM application—fed by point-of-sale (POS) data and customer service records—to track complaints and product problems. This data resource would be more complete if the company could include external product reviews or text from Internet blogs that discuss its goods. Analyzing data from these external sources might reveal a product flaw that escaped quality assurance's discovery or a poorly positioned switch, for example. This insight could help the manufacturer proactively correct a problem before it compromises customer satisfaction.

messages, contracts, documents, sales notes and external market intelligence from both professional evaluations and consumer communications such as blogs.

Manually managing large collections of text data is time consuming and complex. To address this, a few vendors have introduced text-mining applications, which look for predictable language structures (such as sentence patterns) within the data and extract meaning from the text. These chunks of information are stored in the database for later mining or analysis; they also can be stored in the data warehouse, combined with traditional warehouse data and used to enrich the results. Until recently, text-mining products required arduous manual extraction of text as well as intense customization efforts for each application. A few software suppliers, such as

Teradata partner Attensity, have products that can automate and simplify the process of collecting and tagging unstructured data from text sources for later mining and analysis. Common uses for these products include warranty and claims support, opinion monitoring, fraud detection and government intelligence.

Compressed voice files also can be converted into searchable text. Analysts estimate that more than 225 million calls are recorded daily—creating a huge untapped store of voice data that provides a window into customer complaints, competitor mentions, notices of churn, instances of agent rudeness and up-sell and cross-sell opportunities. Speech-analytics applications consider the context and content of conversations, including who is talking, when certain words are spoken in reference to other words and when there is silence during the call. Some applications can also note a caller's rate of speech, volume or emphasis and words that the speaker stresses during the conversation. Metadata about the call, such as customer sentiment and satisfaction, can also be recorded.

Airlines are beginning to use these applications to improve BI about customer interactions. A customer requesting a Monday flight to Phoenix may book his travel on Thursday if the airline offers no other option. The call report may not reflect his preference, especially if the transaction concluded with a sale. But speech-analytics capabilities can measure not only the words recorded during the call but also the pitch and tone of the caller's voice. By analyzing this information, the airline can rate true customer satisfaction and identify additional service opportunities.

Understanding what employees, partners and customers are saying—and what it means to the business—can help enterprises reduce contact-center costs, increase customer retention and satisfaction and enhance agent performance.

However, only a few vendors such as CallMiner offer speech-analytics capabilities today, and the available applications are primarily designed for quality control or quality assurance. Yet industry watchers predict that the speech-analytics market will expand into a multi-billion-dollar business within just two years.

### Everything in its place

A second category of nontraditional data is location-aware information. Location data can be gathered through various technologies such as Wi-Fi for wireless networks, Bluetooth for personal area networks, radio frequency identification (RFID) for logistics tracking and positioning technologies such as the Global Positioning System (GPS), as well as radar, lidar and sonar. Geospatial applications map location data, providing users with a visual representation of the information. Most location data is collected and used locally. Some organizations, realizing this data can be stored in the warehouse, also use it to enrich traditional data warehousing applications.

Each of these technologies offers widespread, proven use in critical business applications. Yet using location-aware data as a source for the enterprise data warehouse (EDW) is just beginning. For example, RFID is widely used by businesses to automatically track the location of boxes or pallets and to timestamp their arrival at various locations. That time and location data could be used to better understand where transport and logistical problems occur, why certain products are out of stock or how inventory processes could be streamlined. Today, however, such applications are custom developed and thus not widely available.

Still, some industries are finding ways to use location-aware information to drive new efficiencies and savings. Forrester Research cites European insurance companies' new pay-as-you-go automobile insurance to consumers. Under this system, a "black box" is installed under the hood of the car; using GPS technology, the box transmits information about



Illustration by Headcase Design



the vehicle's position as well as the time and direction it was driven. Insurers analyze the data to determine the driver's risk, which is reflected in the customer's premium. This system gives customers greater control over their insurance costs and gives insurers improved price transparency and customer loyalty.

### Picture this

Sensory data includes information about such qualities as heat, moisture, temperature and rotational speed. Technologies that measure these characteristics are common. In the oil and gas industry, drill bits are equipped with sensors that detect the bit's rotational speed and temperature. An operator can determine whether the bit is drilling into oil, rock or water. By collecting this data and using it to create maps, companies can identify promising oil fields, determine how deep to drill and specify cost benefits for each field.

Sensory data is already being used in claims and warranty applications. Certain computer components can notify the vendor if problems

arise. For example, a malfunctioning disk drive can send a signal to the customer service group to indicate if it has overheated or its power supply has failed. Vendors can automatically dispatch a service person to replace the drive—often before the customer notices the problem.

Next up: Companies will feed this sensory data into their data warehouses for analysis. For example, telecom vendors that incorporate network node sensor data into their warehouses can better identify which wireless customers were affected by a node or cell-tower outage. By contacting these critical customers proactively, they can reduce the churn rate within a highly profitable customer segment.

Data derived from multimedia—including photographs, video, movies or other images that enterprises may own or store—is another promising unstructured data type. Data from these images can be used in many applications, including fraud and abuse, risk management, claims and warranty, customer relationship management (CRM) and intelligence. High-profile intelligence uses include face-recognition software to identify potential terrorists in public places such as airports.

Yet the technology promises quantifiable benefits to business, too.

Linking multimedia data to operational records in the data warehouse may one day support real-time decision making.

Warehoused images could help retailers provide instant, one-to-one service to high-value customers and eliminate fraud.

Using information from the data warehouse, a sales representative could greet a loyal customer by name and offer new merchandise in her size or favorite color. These images could also be used

to facilitate rapid repairs. By comparing an image of a faulty part with data in the warehouse, mechanics could have a replacement part sent to an airliner waiting on the

tarmac and provide the technician with real-time instructions on making the repair.

### Time to market

Using these new types of data for BI is not yet mainstream. Extracting this data from its original source, successfully structuring it within an EDW and using it to conduct meaningful analytic activities is still a costly, time-consuming effort that requires extensive customization.

To date, very few off-the-shelf solutions extract and convert unstructured data into analysis-ready information. However, several vendors have launched tools to help enterprises collect unstructured data, and others are preparing new products. Look for the release of tools to collect language data within the next six to 12 months, and new voice-analytics applications should follow shortly thereafter. Tools for sensory and multimedia data will likely be last to market, in two to three years.

Despite the newness of these technologies, decision makers should know that incorporating nontraditional types of data in the warehouse offers opportunities to differentiate a company from its competitors. As other companies have shown in their deployment of leading-edge solutions, early adopters may incur a higher implementation expense yet are often rewarded with significant market-share advantages. Later adopters may deploy these technologies at a lower point of entry but achieve only a “me, too” solution that delivers less dramatic results.

Regardless of when you begin to include nontraditional and unstructured data into the warehouse, the value of the strategy is clear: Incorporating this data enhances the richness of your data resources and the power of your analytics activities. To support these efforts, technology innovators are working to provide the solutions necessary to take advantage of these new opportunities. **T**

*Cheryl D. Krivda writes about the intersection of high technology and business practices for publications and corporations worldwide.*

## The power of tomorrow: Future advancements

- › Today, companies possess enormous volumes of unstructured data that cannot be easily imported into the warehouse or used to support BI activities.
- › New technologies will soon be available to make this data searchable. Be prepared: Consider how you can use unstructured data in your EDW to enrich analytics.
- › Tools to help you bring language data, such as text and voice, into the warehouse will be first to market. Later, expect technologies that address sensory and multimedia data.