

XBRL <<<<<<<

By Kristine Brands, CMA

XBRL and Big Data

>>> A lot of buzz has been circulating about Big Data recently. What is the hype all about, and what does it mean for management accountants? This month we'll look at Big Data, its challenges, strategies for harnessing it, and how eXtensible Business Reporting Language (XBRL) plays into the equation.

What Is Big Data?

Big Data is a term that was coined a few years ago to represent the massive increase in data and information being generated by organizations. International Data Corporation (IDC), a global information research and media firm, predicts that the amount of data will increase 44 times between 2009 and 2020 to 35.2 zettabytes, or 1.8 trillion gigabytes. Big Data includes traditional data that's generated within an organization and that can be stored in an organization's database or data warehouse, such as financial and administrative information, customer relationship management (CRM), supply chain and operations, and information technology (IT) management. It also includes other varieties of data, such as video, social media, e-mail, and sensor data (i.e., data generated by

following Web clicks on a social media site), data that has mushroomed in the Internet Age.

With analytical tools, data can be sliced and diced to identify patterns and trends, perform business intelligence, and make predictions to provide information for decision making. While adoption of enterprise resource planning (ERP) systems like SAP and Oracle has shown companies that they can share data and information throughout the organization to run their businesses, the availability of Big Data takes data sets and their analytical potential to a higher and much more complex level. For example, a few years ago traditional sales forecasting models may have been based only on a company's sales history and input from the sales force. Big Data dramatically expands information resources that can feed the forecasting process by allowing companies to include other information, such as clicks on their product websites, competitive research, and economic forecasts. The Securities & Exchange Commission (SEC) has embraced Big Data analysis by contracting a system to monitor real-time trading activity on U.S. stock exchanges

to look at suspicious activity that might indicate fraud. The system collects all order and trade activity to the 1/1,000th of a second and performs sophisticated analytics. The Trading and Markets and the Risk, Strategy, and Financial Innovation divisions of the SEC use the analysis generated by this system.

Challenges

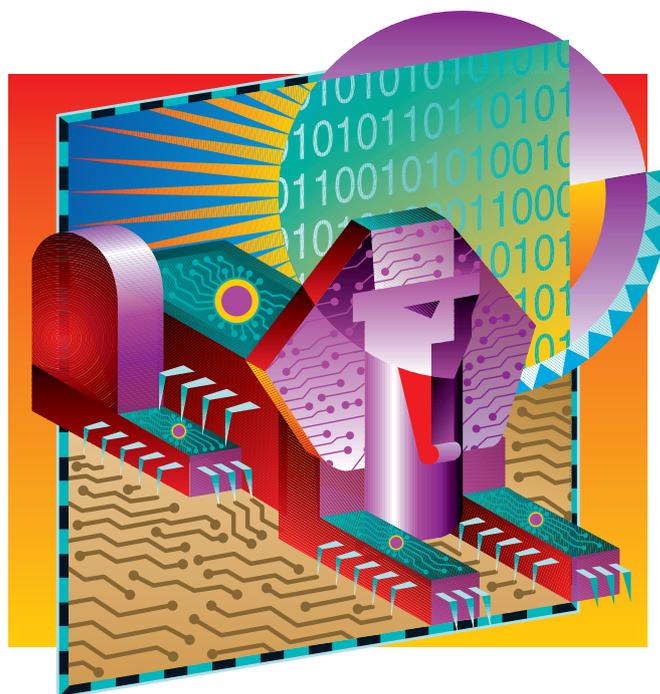
There are several issues surrounding Big Data: the volume of data, a company's infrastructure to manage it, lack of qualified Big Data analysts, categorizing and managing the information, and data integrity risk. Organizations are being swamped with information and may not know how to manage and analyze it. The amount of data collected and gathered is too big to handle and analyze using traditional database analytical tools because organizations' IT infrastructures may be constrained by server and storage media size. Also, analytical tool resources may be inadequate to leverage value from the information collected. If organizations don't have information management strategies, they will miss data analysis opportunities. And according to McKinsey & Company, Inc., there's

also a lack of qualified personnel who are trained in Big Data analysis.

Another challenge is making sense of Big Data because most of it is unstructured, which makes it difficult to organize, categorize, and analyze. Unstructured data means that the information contains no identifiable structure or identifier—it isn't formatted.

As a result, computer software can't read the data. This is in contrast to structured data like XBRL that matches a data element to an XBRL tag that is computer readable. For example, the U.S. GAAP Taxonomy's tag for Accounts Receivable Gross Current is `AccountsReceivableGrossCurrent`. The issue is how to sift through, filter, and analyze information to find the data needed in the data haystack if the data sources include e-mail, videos, and sensors that are unstructured.

Data integrity risk is another issue. If the data gathered is drawn from many different sources, how do you validate data accuracy? Failure to do so brings to



mind the saying "garbage in, garbage out." Expensive Big Data models and analysis are useless if the underlying data is unreliable.

Strategies for Harnessing Big Data

The sheer volume of Big Data places its collection, handling, and analysis beyond the scope of many organizations' IT departments. The cloud provides a way for organizations to outsource analytics to an environment that has the bandwidth to handle monster data sets. And new technologies are being developed

to perform Big Data analysis, such as Apache Hadoop (High-availability distributed object-oriented platform), that can manage such an analysis. Other technologies are being developed to analyze unstructured data and are expected to answer this need.

Because XBRL is a structured data language that provides a context and format to data, it's well positioned for Big Data analysis. Its major use today is

government regulatory reporting, such as the SEC's 2009 XBRL Reporting Mandate. For XBRL to play a major role in the Big Data analytics equation, however, broader adoption of XBRL for business information is necessary to be able to leverage the rich sources of business information that are currently unstructured. **SF**

Kristine Brands, CMA, is an assistant professor at Regis University in Colorado Springs, Colo., and is a member of IMA's Global Board of Directors and the Pikes Peak Chapter. You can reach her at kbrands@yahoo.com.