2015 STUDENT EDITION

Business Intelligence Journal

THE LEADING PUBLICATION FOR BI, DATA WAREHOUSING, AND ANALYTICS PROFESSIONALS

tdwi
Advancing all things data.

Love Data?

Go beyond business intelligence with

TDWI Student Membership

Study Smarter

Boost your GPA

Get a World-Class Data Education

ADVANCE IN CLASS AND YOUR CAREER WITH TDWI STUDENT MEMBERSHIP

TDWI Student Membership is your all-access pass to all things data. As a student member, you'll get the information, resources, connections, and answers you need to stay informed, solve problems, and understand emerging technologies and methods driving innovation. You have access to the same benefits as our regular members at a fraction of the cost!

Study smarter: Learn best practices from the best minds in the industry covering concepts in business intelligence, data science, data warehousing, analytics, and more.

Boost your GPA: You'll have unlimited access to research, articles, and publications covering all things data. Apply your knowledge and experience to coursework for immediate impact.

Get a world-class education at an affordable price: Learn online, onsite, or on your own with exclusive student member discounts on education, events, and exams.

Join thousands of Members who use their benefits to advance their knowledge of all things data.

Use promo code **STUDENT** for special pricing available only from this issue of the *Business Intelligence Journal* Student Edition.

tdwi.org/student-membership

tdwi

Advancing all things data.

2015 STUDENT EDITION

Business Intelligence Journal

Business Intelligence Journal

STUDENT EDITION 2015

tdwi.org

Advertising Sales: Scott Geissler, sgeissler@tdwi.org, 248.658.6365.

List Rentals: 1105 Media, Inc., offers numerous e-mail, postal, and telemarketing lists targeting business intelligence and data warehousing professionals, as well as other high-tech markets. For more information, please contact our list manager, Merit Direct, at 914.368.1000 or www.meritdirect.com.

Reprints: For single article reprints (in minimum quantities of 250–500), e-prints, plaques, and posters, contact: PARS International, phone: 212.221.9595, e-mail: 1105reprints@parsintl.com, www.magreprints.com/QuickQuote.asp.

[ISSN 1547-2825]

Reaching the Staff
Staff may be reached via e-mail, telephone, fax, or mail.

E-mail: To e-mail any member of the staff, please use the following form: FirstinitialLastname@1105media.com

Renton office (weekdays, 8:30 a.m.–5:00 p.m. PT)
Telephone 425.277.9126; Fax 425.687.2842
555 S Renton Village Place, Ste. 700
Renton, WA 98057-3295

Corporate office (weekdays, 8:30 a.m.–5:30 p.m. PT)
Telephone 818.814.5200; Fax 818.734.1522
9201 Oakdale Avenue, Suite 101, Chatsworth, CA 91311

***Business Intelligence Journal* (article submission inquiries)**
Marie Gipson
E-mail: journal@tdwi.org
tdwi.org/journalsubmissions

TDWI Premium Membership (inquiries and changes of address)
E-mail: membership@tdwi.org
tdwi.org/Premium-Membership
425.226.3053
Fax: 425.687.2842

From the Editor

This special edition of TDWI's *Business Intelligence Journal* offers practical career advice you'll need to succeed in the world of business intelligence, data warehousing, and business analytics. From technical skills to people skills, we examine what you'll need in the field and how BI is used in the real world to improve our lives.

Industry is excited about BI and analytics, but how do you know if you're a good fit for the field? *Journal* senior editor Hugh J. Watson explains the five types of analytics users and what skills each type needs. Are you curious? Do you have a basic understanding of statistics, good business sense, and a good feel for possible relationships and solutions? You may want to consider becoming a data scientist, a position that's currently tough for enterprises to fill. In a separate article, Watson examines 10 important insights about business analytics you'll want to consider as you weigh career options.

In our BI Experts' Perspective column, three industry experts discuss why BI is important and what career opportunities are available. This column reveals valuable answers to the questions you'll likely be asked at a job interview.

Because you probably won't learn everything you need to know in the classroom, developing and honing your skills is a lifelong pursuit. Watson examines the "soft" skills you'll find of greatest value in both the short and long term. Many of these skills—such as "people skills"—you'll learn by doing your job, through mentoring, or from coaching by a trusted colleague or manager. You may need coaching to develop your presentation skills (Please don't read me the text on your PowerPoint slides!) or your communication skills. After all, communicating the results of your analysis is key for your audience (and your employer) to understand the trends you've uncovered and make the best business decisions.

Some skills you've already acquired are in demand. Justin Hay takes a look at how a "younger generation" of business users (that's you) raised on games as an integral part of their lives can be engaged in business applications and learn to appreciate data governance though gamification.

The most valuable employees bring with them multiple disciplines. Larissa Moss takes a look at the "lost art" of data modeling, what she calls a "mature and proven technique that spans logical data modeling, metadata management, data governance, data stewardship, data-quality improvement, and master data management.

What can BI do for you? Coy Yonce examines how BI is being put to use at Apple, Blizzard Entertainment (the *World of Warcraft* folks), and healthcare organizations around the world.

Consider joining TDWI as a student member. Details and discounts are available at http://tdwi.org/pages/membership/student-membership.aspx.

James E. Powell

Should You Pursue a Career in BI/Analytics?

Hugh J. Watson is a Professor of MIS and holds a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia. He is senior editor of the *Business Intelligence Journal*.
hwatson@uga.edu

Hugh J. Watson

Business intelligence (BI), big data, and analytics are hot topics in both the popular and business press.[1] Articles in publications including the *New York Times*, *Wall Street Journal*, and *Financial Times*, and books such as *Super Crunchers* (Ayres, 2007), *Competing on Analytics* (Davenport and Harris, 2007), and *Analytics at Work* (Davenport, Harris, and Morison, 2010) have spread the word about the potential value of analyzing data to improve employee productivity, enhance customer service and satisfaction, optimize the supply chain, provide a competitive advantage, and ultimately improve the bottom line.

There is particular interest in big data, which is characterized by high volume, variety, and velocity. It streams into organizations from sensing devices on machines; websites; social media; RFID chips; GPS systems; and voice, image, and video files. The challenge is to store, analyze, and take action based on the data, and this requires new storage technologies (such as Hadoop), analysis tools (for example, R), and people who know how to work with big data (including data scientists).

Big data is creating IT jobs. Gartner (2012) predicts that by 2015 the need to support big data will create 4.4 million information technology (IT) jobs globally, with 1.9 million of these in the U.S. For every IT job created, up to an additional three jobs may be generated outside of IT.

Big data is also directly creating jobs outside IT. Organizations need people who can analyze and use big

[1] This article is adapted from three previous articles (Watson, 2013; Watson, 2014; and Watson, 2015).

data. A 2011 study by the McKinsey Global Institute predicts that by 2018 the U.S. alone will face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions (Manyika, et al, 2011). This shortage is significant because inadequate staffing and skills are the leading barriers to the use of big data analytics (Russom, 2011).

Is Analytics for You?

Because of the high demand for people skilled in analytics, you may be thinking that it's a field you should enter. It might well be, but it is important to understand what the field is all about, who uses analytics, and the skills analysts require. Even if you do not plan a career in analytics—a specialized career in analytics is not for everyone—you are likely to use analytics in your work.

You also need to honestly reflect on your aptitudes and interests. Let me share a personal experience. I majored in electrical engineering as an undergraduate and worked in the aerospace industry but never really liked the field. I majored in engineering because my parents encouraged it and I was good in math and science. I ignored the fact that I wasn't good at it, didn't have patience for it, and didn't enjoy fixing things. I could do engineering work but I didn't enjoy it. Ultimately, I went to graduate school in business to take a different career path.

Analytics Variety

The analytics term is used broadly. It is helpful to understand the various kinds of analytics because each serves a different purpose, which affects what today's workers use and need to know.

The objective of *descriptive analytics* is to describe *what has occurred*. Reporting, online analytical processing (OLAP), dashboards/scorecards, and data visualization are all examples of descriptive analytics.

Predictive analytics focuses on *what will occur* in the future. The algorithms and methods for predictive analytics include regression analysis, factor analysis, and neural networks; applications include demand forecasting, customer segmentation analysis, and fraud detection.

Prescriptive analytics investigates *what should occur* and is used to optimize system performance. Revenue management, which strives to optimize the revenue from perishable goods (such as hotel rooms and airline seats), is a good example. Through a combination of forecasts, predictions of consumer behavior, price sensitivity analysis, competitive information, inventory data, and mathematical programming, the price of a good or service is set dynamically to optimize revenues.

Most organizations progress from descriptive to predictive to prescriptive analytics. First, organizations monitor what is taking place. With that mastered, they turn to predicting what's ahead. Finally, they want to shape the future.

Skills Analytics Users Need

There is no such thing as a "typical" user of analytics; see Figure 1.

Casual end users are at one end of the continuum. They access analytics-related information through reporting, OLAP, and dashboards/scorecards. They have solid knowledge of the business and need to understand what data is available, how to access and manipulate it, and how to use the data to perform their jobs. They don't need to be technical experts.

Power end users are next on the continuum. They are often experts with Excel and can create information using the company's BI tool (such as MicroStrategy). Power users often answer questions for casual users, help them with analyses, and can build new reports for themselves and their colleagues. They have strong business skills and sufficient technical skill to access data, perform analyses, and create reports, dashboards, and scorecards.

If you plan to become a casual user—and especially a power user—don't underestimate the importance of analytics skills. A case study I conducted with First American Corporation (FAC), a regional bank headquartered in Nashville, Tennessee, illustrates this point (Watson, Wixom, and Goodhue, 2002). FAC made it clear that the good jobs of the future will require analytical skills.

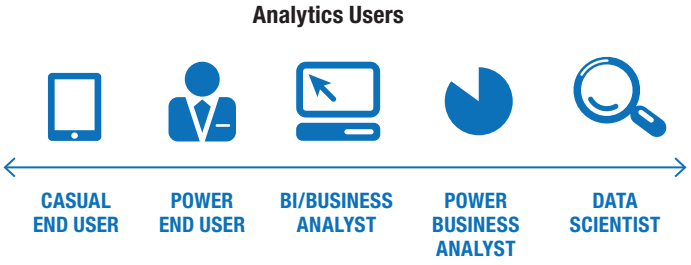

Figure 1: A continuum of analytics users.

FAC attracted my interest because it had been on the brink of failure, but a new management team turned the bank around by employing a customer intimacy strategy that relied heavily on analytics. It used analytics to understand its customers' needs and preferences; determine the profitability of customers, products, and services; and redesign its distribution channels (including branches and ATMs). Analytics and fact-based decision making replaced intuition and many previous banking practices.

I interviewed the CEO, who drove the changes. I asked him about the impacts to personnel. He said the most dramatic change was in marketing. The size of the staff hadn't changed, but none of the 12 team members was in the same job. The previous staff thought marketing was "giving out balloons and suckers along the teller line and running focus groups." Marketing was now analytical. New staff with the necessary analytical skills and experience had to be hired to do the work. The previous marketing people either left the bank or took other positions.

In the middle of the continuum you'll find **two categories of analysts**. *BI analysts* are part of the BI or analytics team and work throughout the organization. They typically understand the organization's data well and usually work on enterprisewide applications such as scorecard systems. *Business analysts*, on the other hand, are located and work in business units (say, marketing or manufacturing) and focus on analytic applications such as customer segmentation analysis in marketing or optimizing supply chain processes in manufacturing. They typically have strong business knowledge and can work with specific analytics tools.

Power business analysts are next on the continuum. They know the business well and possess strong analytics skills. Because of the high demand for and limited supply of data scientists, some of these analysts have seized the opportunity to do data scientist-type work. They operate in a very specific domain, such as finance.

Finally, the job of a ***data scientist*** is to discover patterns and relationships in data that no one else has seen or wondered about and turn these discoveries into actionable information that creates organizational value. To do this requires a rich mixture of data understanding, analytical skills, and business knowledge.

Where Do You Fit In and What Skills Do You Need?

To be proficient with analytics, you must be an analytical thinker and enjoy problem solving. Beyond that, the skills you'll require depend on where your job is on the analytics user continuum. The movement from casual user to data scientist requires ever-increasing aptitude and training in analytics.

Users

Casual users access analytics-related information and use descriptive analytics tools and applications in their work. What skills do they need? I'd argue that they

must (1) understand how data is stored in relational databases, (2) be able to access and analyze data using a variety of analysis tools, and (3) have data-gathering and data analysis experience. You can develop these skills in statistics, database, and BI/analytics courses.

Analysts

Analysts use tools and applications to understand business conditions and drive business processes. These users access and analyze data and generate information for themselves and others. Analysts should be analytical and inquisitive. Many are business school graduates with MBAs, but others have degrees in statistics, mathematics, engineering, or other fields where they gained analytical skills. If they are business school graduates, MIS, marketing, and finance are common majors.

Many universities are gearing up to produce business analysts (and data scientists) though degree programs, MBA concentrations, and certificate programs. These offerings are commonly in business, engineering, and statistics and the instructional delivery varies from on campus to online. One of the first and best-known programs is the Master of Science in Analytics at North Carolina State University. SAS has been an important contributor to the program; it is offered through the Institute for Advanced Analytics and has its own facility on campus. Deloitte Consulting has partnered with the Kelly School of Business at Indiana University to offer a certificate in business analytics for Deloitte's professionals. Northwestern University offers an online Master of Science in predictive analytics through its School of Continuing Studies.

Data Scientists

Much has been written about data scientists. A *Harvard Business Review* article calls it "the sexiest job of the 21st century" (Davenport and Patil, 2012). Despite this attention, confusion persists about who these people are, what skills they have, and what they do.

Data scientists use "rocket science" algorithms (such as neural networks) and interactive exploration tools (for example, SAS Enterprise Miner) to uncover non-obvious patterns in data. Some data scientists are also proficient in prescriptive analytics, such as mathematical programming. They often have advanced training in multivariate statistics, artificial intelligence, machine learning, mathematical programming, and simulation. Data scientists need to understand the different types of data and how they can be stored (RDBMS vs. Hadoop, for instance), write code (in Java, Python, or R, among other languages), access data (in SQL or Hive, for example), analyze it (using regression analysis, social networks, and similar techniques), and communicate findings to management in business terms. Data scientists are typically curious, like to solve difficult problems, and have advanced degrees in fields such as analytics, statistics, computer science, management science/operations research, physics, and mathematics. Ideally, data scientists should have the characteristics and skills described in the following list, although it is unlikely that any one person will be strong in all areas:

- **Curiosity:** A desire to understand relationships in data and solve problems
- **Intuition:** Good "business sense" and a feel for possible relationships and solutions
- **Data gathering:** The ability to access and integrate data from different sources
- **Experimental design:** The ability to design experiments (e.g., control and experimental groups) to test suspected relationships
- **Statistics:** An understanding of basic (e.g., expected value) and advanced (e.g., logistical regression analysis) statistical methods
- **Analytical modeling:** The ability to use analytical methodologies (e.g., CRISP DM), algorithms (e.g., neural networks), and tools (e.g., R, SAS Enterprise Miner)
- **Communication:** Clear explanations of analytics results using business terminology

Organizations don't need many data scientists, but they are useful for the most challenging problems. Because

their strong suit is often data and modeling, data scientists may need to be paired with business users and analysts to provide sufficient business knowledge to the team.

Conclusion

There are many great career opportunities in BI, big data, and analytics, but having a career in any of these areas is not for everyone. You'll need analytic aptitude, the right training, and hard study. The requirements for success increase as you move across the analytics user continuum.

To prepare for a career in analytics, take courses in statistics, databases, business intelligence, and analytics. Find an internship where you'll perform analytics work in order to develop your business and analytics skills and experience. Read about analytics beyond what is required in your courses. To advance in analytics, think about and plan to earn a graduate degree in an analytics-related field.

References

Ayres, Ian [2007]. *Super Crunchers*, New York: Bantam Books.

Davenport, Thomas H., and Jeanne G. Harris [2007]. *Competing on Analytics: The New Science of Winning*, Boston: Harvard Business School Press.

Davenport, Thomas H., Jeanne G. Harris, and Robert Morison [2010]. *Analytics at Work: Smarter Decisions, Better Results*, Harvard Business School Press.

Davenport, T. H., and D. J. Patil [2012]. "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, October, pp. 2–8. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

Gartner [2012]. "Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015," Gartner Press Release, October 22. http://www.gartner.com/newsroom/id/2207915

Manyika, James, et al [2011]. "Big Data: The Next Frontier of Innovation, Competition, and Productivity," McKinsey Global Institute, May. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Russom, Philip [2011]. *Big Data Analytics*, TDWI Best Practices Report, Q4. http://tdwi.org/bpreports

Watson, Hugh J., Barbara H. Wixom, and Dale L. Goodhue [2002]. "The Effects of Technology-Enabled Business Strategy at First American Corporation," *Organizational Dynamics*, Winter, Vol. 31, Issue 3, pp. 313–323.

Watson, Hugh J. [2013]. "The Business Case for Analytics," *BizEd*, May/June, pp. 49–54.

____ [2014]. "Tutorial: Big Data Analytics: Concepts, Technologies, and Applications," *Communications of the Association for Information Systems*, Volume 34, Article 65, pp. 1247–1268.

____ [2015]. "Data Lakes, Data Labs, and Sandboxes," *Business Intelligence Journal*, Vol. 20, No. 1, pp. 4–7.

Soft Skills for Professional Success

Hugh J. Watson

Hugh J. Watson is a Professor of MIS and holds a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia. He is senior editor of the *Business Intelligence Journal*.
hwatson@uga.edu

Introduction

Business intelligence managers and professionals must be technically competent. Although the skills vary, there are at least some technical requirements for every BI position. It may be possible to be successful based solely on superior technical skills, but the more likely case is that you also need to master some soft skills, especially if you want to progress up the organizational ladder.

We strive to turn out strong MIS (management information systems) graduates at the University of Georgia, including the students who take my BI classes. We make sure that they have a good technical background with courses that cover, for example, data modeling and SQL, business process management, Java, systems analysis and design, and Web-based application development. We also work on their soft skills. We have project management and information system (IS) leadership courses. Most of our courses involve group projects and presentations. We invite many speakers from the business community who discuss the realities of IS work and the workplace.

Our primary mechanism for interacting with the business community is the MIS Advisory Board (www.terry.uga.edu/misadvisoryboard). Board members meet in the fall (in Atlanta) and in the spring (in Athens) to discuss our curriculum, assess emerging trends and their potential impact, and explore ways of connecting our students with companies that want to hire them. Board members and their companies also speak to classes and to our student MIS organization, participate in research projects, and offer internships and scholarships.

This year we had the board members explore the skills our students should have in order to be professionally successful in both the short and long term. At the fall board meeting, we used the nominal group technique to generate and

prioritize a list of skills.[1] At the spring meeting, we had breakout groups and discussions about the skills and how to impart them to our students. Table 1 shows the list of skills we identified. I find it interesting that all of them are soft skills. The board members assumed that students would have the requisite technical skills.

Let's consider these skills and how they might apply to you. I will relate some of the more interesting observations the board made as well as some of my own experiences. I also asked several current or past BI directors—Jim Gallo, Donna Najafi, Celia Fuller, and Jim Hill—to offer comments and give examples that illustrate the importance of these skills.

Rank	Skill
1	People skills
2	Find mentors
3	Networking
4	Build a brand
5	Long-term career perspective
6	Value non-technical courses
7	Understand data gathering and analysis
8	Presentation skills
9	Separate personal and professional lives
10	Work/life balance
11	Prioritizing and executing

Table 1. Skills needed for professional success.

Skill #1: People Skills

Every BI job involves interactions with others, even if it is just your teammates. Most involve considerable interaction with people throughout the company. How well you interact with others influences your career success. As one board member said, "If you want a job, have technical skills. If you want a career, have people skills." You are evaluated based on what you do as well as how you do it.

Donna Najafi tells a story about a BI analyst who has excellent people skills. The analyst was given a set of specifications for a new application that the business manager thought was complete and ready for coding, but actually was too vague. Complicating the situation was the manager's previous negative experience with IT, which caused him to not want to spend any more time than necessary on the project. Fortunately, the analyst was very good at working with people and was able to convince the manager that with just a little more input she would be able to provide an application that truly met the business's needs. If the analyst had not been able to develop the relationship, rework would inevitably have been necessary because the IT team would have had to guess at the requirements.

Some people have better people skills than others. Although some are innate, people skills can be learned through observation and practice. What are some of the things you see others doing that either enhance or detract from their relationships with others? What practical advice can you take away from books on the topic? Even small things such as sending a birthday card or recognizing a teammate's accomplishments can help.

Skill #2: Find Mentors

The ideal mentor is someone outside your team (to avoid potential conflicts of interest), has more seniority than you have, knows the company, takes an interest in you, and is glad to help. When you have a mentor, let that person know that you are listening and taking their advice. When you follow their advice, tell your mentor what you did and that you appreciate their help. If you do this, your mentor will only want to help you more.

Mentors can also come from outside your company. Let's say that you aspire to be a BI director. A BI director at another company might be a great mentor, but how can you establish a relationship? The first step is to locate a good candidate, and this is where conferences like those offered by TDWI can help. Potential mentors may be speaking, sitting at a table with you, going through the exhibit hall, or having a drink during a social event.

Strike up a conversation, and if it goes well, ask if you can contact them for career advice. Most people, even very

[1] You can learn about the nominal group technique at http://en.wikipedia.org/wiki/Nominal_group_technique.

successful and busy ones, appreciate the implied compliment and enjoy helping others.

Skill #3: Networking

Although networking sites such as LinkedIn are useful, you can't beat a more personal approach, because this is how strong relationships are built. Seize any opportunities to meet people and let them know about you and your skills. In your own company, this can lead to being requested more often for the best projects. In addition, keep in touch with people who leave your company and find employment elsewhere. Externally, TDWI, user group conferences, and other professional meetings are potentially very useful. You never know when you might meet someone who may later contact you about a job opening. Celia Fuller has an interesting observation about networking.

She says that seizing networking opportunities is important, especially in terms of "internal" networking with the influencers and customers you work with. Work essentially gets done through relationships. Building those relationships enables you to build your program and accomplish your goals, with a longer-term effect of promoting your success and career. Because everyone is so busy, it's difficult to carve out the time to schedule lunch, drop by, or make calls. Set goals to do this with key people on a periodic basis, not necessarily with a specific "ask" in mind, but to build each relationship so that when the time comes, it's there.

Skill #4: Build a Brand

Brands are important for companies and products as well as for individuals. Your brand reflects who you are and how others perceive you. What is your brand? Is it favorable? Do you get BI projects done on time? Do you understand the business? Can you do many things? Are you easy to work with? Just as companies are protective of their brands, you should be, too.

An example of someone with a great brand is Cindi Howson. Cindi frequently teaches at TDWI conferences. She is my "go-to" person when I want to know about BI tools. She is knowledgeable, friendly, helpful, and willing to share information. That is her brand.

How do you develop a brand? Start by deciding what you want your brand to be. Next, build it through your actions and behaviors. Then communicate what your brand is. Being good at your job is a necessary but not sufficient condition if you want to be recognized. You need to make sure that others are aware of how good you are—but be careful. You do not want to be perceived as being too self promoting.

Skill #5: Long-Term Career Perspective

Although some career opportunities may be unforeseen, it is good to have a long-term plan. Do you want to work for yourself or for a company? Do you want to be a BI director, CIO, or move into a line of business? What kind of organizational culture or colleagues are important to you? The answers to these and similar questions can help you decide what career you want and how to realize it.

Just as there are road maps for companies implementing BI, develop a road map for your career. Identify the positions, responsibilities, salaries, and locations that you want to achieve by various deadlines. When you are done, share your road map with others (such as your manager) and ask for feedback. Ask if your plan is realistic. Are there other opportunities? In what areas do you need to improve?

Celia Fuller suggests that rather than thinking about the next step up the chain, consider career opportunities that are "sideways," such as from a business role to IT, or from one department to another. By broadening your experiences and perspectives, you can build a solid foundation for later upward progression.

Skill #6: Value Non-Technical Courses

Some BI roles are almost purely technical. Others, such as decision support analyst, demand a combination of technical and business skills. The more your job requires you to work with others throughout your organization or manage people and business units, the more important business skills become. You can develop these skills

through reading or by completing business classes. Consider earning your MBA if you really want to progress.

At our board meeting, one member said that she took courses in philosophy, art, literature, and languages when she was in school. She wasn't sure of the value of these courses until her job with a global company required her to travel extensively in Europe and she found she could identify with people on a cultural level.

Skill #7: Understand Data Gathering and Analysis

At its core, data gathering and analysis is what BI work is all about. What separates the exceptional BI analyst from others is whether the person understands what data is needed and how it should be analyzed without being given considerable direction. This skill requires a combination of aptitude, education, and experience.

One of the best examples of this occurred many years ago when I worked with George Houdeshel at Lockheed- Georgia on the company's executive information system. George was an engineer by training, had worked throughout the company, and knew all of the firm's senior executives well. He also had an uncanny understanding of what information people needed and how to best analyze and display it.

Another example is Continental Airlines. I was impressed by how quickly they could roll out new BI applications. One reason was that they had a rock-solid data infrastructure to draw upon. Another reason was the BI analysts who worked with the business units on new applications. Many of them had originally been on the business side before joining the BI staff. Because of their previous work experience, they really understood the business's needs and requirements.

Skill #8: Presentation Skills

We all have to make presentations at various times, and I enjoy watching people who are really good and trying to learn what makes them special. The interesting thing is how different they tend to be. Each presenter has a distinctive style.

One common denominator, however, is that outstanding speakers have practice and experience. I observed this one day while watching noted Harvard professor F. Warren McFarlan, who walked around the stage and spoke mostly to the floor—not what I expected, nor a presentation technique commonly taught today! However, his choice of words, phrases, and examples were outstanding, and he seemed to be speaking extemporaneously. I was impressed. Later that day, he gave the presentation again, and I attended again. To my amazement, the presentation was identical. What had seemed so unrehearsed was close to a stage performance. I was impressed again, but this time with the work he had obviously put into his talk.

Just as there are road maps for companies implementing BI, develop a road map for your career.

In developing strong presentation skills, you can learn from the different kinds of presentations people make and the different media that exist. For example, presentations can be made with and without slides, in varying lengths, to different audiences (e.g., technical versus business). Some communications are face-to-face, while others are electronic—especially as the number of virtual teammates increases. There are many ways to learn about making presentations—you can read books, attend classes, and even participate in Toastmasters. Remember the importance of practice and garnering honest feedback on your talks. Improving your presentation skills can reap rewards, such as when your next BI project has to be presented for approval.

Jim Gallo tells an interesting story about the first time he made a presentation on the merits of creating an enterprise data warehouse. In his presentation, he made references to logical data models, EDW, metadata, and similar words and phrases. The result was blank stares. The next time around, he asked about the need for a reporting database and a glossary of terms. This time

there was great support for the project. Since then, he has been careful to use terms that business people relate to.

Skill #9: Separate Personal and Professional Lives

To some extent, this is a generational issue, but not exclusively. We all know older workers who say too much about their personal lives. Younger workers who grew up with Facebook and Twitter sometimes share too much personal information because, in part, they always have.

When you add in the long hours that people work, the opportunities for saying things you shouldn't are abundant. When this becomes a problem, colleagues and mentors should step forward and counsel others about the importance of being careful about what to disclose about personal lives.

Skill #10: Work/Life Balance

In the early days of PCs, I resisted having one at home because of the potential temptation of working all the time. For professors, there is always something that is worth reading or worth writing about in your next article or book. That resistance didn't last long, however, and now I'm like nearly every other white-collar professional—connected 24/7.

The best solution for balance is to carve out some time for the important things in your life and become disconnected (as much as possible). One of our recent graduates is working for a consulting firm and is putting in 60-hour weeks. He said that he assessed what is really important to him and decided that it was taking time to play sports. He joined a Saturday soccer league, and those afternoons playing soccer are "his time."

Skill #11: Prioritizing and Executing

BI managers and professionals are always adapting to changing priorities and requirements. In fact, that is why agile BI has become so popular. We know that as the business's needs or requirements change, the work priorities and execution plan must change as well. Some people are better at adjusting "on the fly" than others. It is up to BI managers to use methodologies that accommodate change as rapidly as today's business environment demands and to hold subordinates accountable for being able to quickly reprioritize and execute.

Jim Hill describes how his BI team uses an agile development methodology that helps the team prioritize and execute work. Weekly iteration planning meetings are held; all business areas are represented. When competing business interests are vying for the same scarce resources, the team discusses the business impact and urgency of each request and comes to an agreement on which requests should go first.

Frequently, an urgent business request will arise in the middle of a planned iteration. Because the tasks are already prioritized, the new, urgent task can be inserted into the current work period and the lowest priority task(s) can be pushed to the next iteration. When deciding which tasks to push, the priority order is considered, as well as the team skills required and the scope of the new task.

Being able to weigh competing priorities, manage conflict, and negotiate with business clients are all soft skills that are important when managing BI work.

Closing Thought

If you are like me, you are better at some of these skills than others. The key is to build on your strengths and work to correct your weaknesses. If you do, you are likely to have a highly successful professional career. ■

BI Experts' Perspective

Business Intelligence as a Career Choice

Dave Schrader, Ph.D., just retired as director of big data marketing for Teradata. His popular "BSI: Teradata" video series on YouTube shows 12 different cases in which data scientists "solve" hard business problems using a mix of big data and traditional BI technologies. He is a board member of the Teradata University Network, which provides teaching tools and cases about BI for faculty and students (www.teradatauniversitynetwork.com). His Ph.D. in computer science is from Purdue and he worked at Teradata for 22 years. drdaveschrader@gmail.com

Ronald S. Swift is managing partner at International Management Consulting (IMC) and "executive-in-residence" at the Terry College of Business at the University of Georgia. ronswift@uga.edu

Coy Yonce is the product owner for software solutions from EVtechnologies, a services and software partner with SAP. Coy also advises organizations on implementing and utilizing analytics to help them understand the effectiveness of their analytics solutions. coy@evtechnologies.com

Dave Schrader, Ron Swift, and Coy Yonce

Q **Susan Stephenson is the BI director for a national auto parts company. She has 10 years of BI experience, with the last three as director. Susan keeps up to date by reading, watching online seminars, and attending BI conferences.**

After church last week, Susan was chatting with Craig Mercer, who teaches computers and information systems at the local college. At the end of their conversation, Craig asked Susan if she would be willing to be the speaker at the school's fall management information systems (MIS) banquet. About 100 students and faculty will attend, and she will speak after the social hour and dinner and before scholarships are awarded. Her talk should be about 15-20 minutes long.

Craig further explained that the current curriculum requires students to complete course work in programming languages, databases, systems analysis and design, telecommunications and networking, project management, and computing architecture, but there is little on BI. That is why he would really like Susan to speak.

When she asked what Craig would like her to cover, he responded, "What BI is, why it is important, what the career opportunities are, how to prepare for and start a BI career, and what technical and 'soft' skills a graduate needs to be successful." He also said that students want know what's currently "hot" in BI and what is likely to be important in the future.

If you were Susan, what would you tell the students?

A **DAVE SCHRADER**

Susan should begin by explaining the differences between traditional BI and the hot new areas. Traditional BI is about the use of traditional data by back-office business analysts—with the help of IT—to track product sales, inventory, costs, customers, promotions, and campaigns. This often includes dashboards and scoreboards as well as forecasting of demand and markets.

Some of this work can be done in Excel, but much of it uses tools such as MicroStrategy, Cognos, Business Objects, or perhaps Tableau

Software to access a data warehouse provided by Teradata, IBM, Oracle, or Microsoft. Current state-of-the-art projects using traditional technologies include putting dashboards and alerts on mobile devices (so people can work more hours!) as well as placing data and BI tools in the cloud. Of course, there are ever-increasing amounts of data to collect each year, and data is arriving at a higher velocity (sometimes "near real time," meaning seconds, or even "real time," meaning milliseconds, between data creation and ingestion).

What's hotter than hot at the moment is the entire area of "big data," which is a bit of a misnomer. Although big data partly includes new data types such as voice, text, and sensors ("the Internet of things"), it's more about the ability to use algorithms to transform raw, non-traditional data types into useful additional insights—finding the signal in the noise, because many of these data inputs contain low-density information. For example, these systems might transform a raw voice file into "what did the person at the call center say?" (voice to text) or "how irritated was the person?" (sentiment score). It might mean linking the symptoms a doctor types into patient notes (text) with other traditional data to find out if the doctor is over-prescribing or allowing off-label uses.

The most interesting BI trend is the need for new and different discovery tools, as well as the people who are able to drive them. The new paradigm for discovery is "grab and go" for fast insights, often using dirty data that's not completely modeled and may not even have known schemas. This is exploratory work, often fast-fail. Only after you find interesting insights worth saving do you go back and perform the rigorous work associated with capturing, modeling, and storing information in a data warehouse for wider use within the enterprise.

The most interesting BI trend is the need for new and different discovery tools, as well as the people who are able to drive them.

A new kind of data analyst—the data scientist—is in extremely high demand to do this discovery work. In fact, data scientists can command starting salaries of $90,000 with stock options—if they have the right skills. The role requires a blend of curiosity, statistics expertise, computer science, business domain knowledge, and communication skills. Almost no one has all five of these characteristics, but students in MIS who want to increase their marketability can acquire some of these skills, partly at school but also by watching online Webinars, reading books and articles, and finding faculty who understand this trend.

Specific tips include:

1. If you are an MIS major, choose a minor that provides domain expertise and take any classes that have "Computational" in the title. For example, "Computational Chemistry" or "Computational Biology" would provide background for a data scientist role with a pharmaceutical company doing genome discovery work.

2. Make sure you are well grounded in some of the hot areas within computer science such as Web analytics, visualization, or wearable devices/sensors.

3. Read any of the new books on big data and big data analytics. My personal favorite is *Taming the Big Data Tidal Wave* (Wiley, 2012) by my friend Bill Franks, Teradata's chief analytics officer.

4. Read anything—anything at all—by Tom Davenport. His landmark book *Competing on Analytics* (Harvard Business Review Press, 2007) is a classic. See also his white paper "The Rise of Data Discovery" (available on www.asterdata.com).

5. To see some examples of big data in action, take a look at my "BSI: Teradata" series, in which business scenario investigators (BSI) use data and BI to solve cases such as "The Case of the Dropped Mobile Calls" or "The Case of the Tainted Lasagna."

These are extremely popular with faculty and students learning big data and BI and will show you the new discovery tools in action.

Anyone with the skills to translate needs and results between business users and IT groups is guaranteed a long and valuable career.

6. Finally, listen to Webinars or download reports from TDWI. TDWI research analysts Philip Russom, David Stodder, and Fern Halper publish excellent surveys of what companies are doing and what technologies those companies are adopting. Matching your learning efforts to the areas that companies are investing in will give you a tremendous edge in finding a great BI job.

   All of these will help build up hard skills, but soft skills are equally important. You must be curious, like those investigators on the TV show *CSI*, who let the evidence and clues guide their investigations. You must also learn to communicate effectively—the buzzword in the industry at the moment is "storyteller." Often I think of it as not just storytelling but translating as well. Anyone with the skills to translate needs and results between business users and IT groups is guaranteed a long and valuable career.

Those are my suggestions. I hope they help Susan build a great presentation!

RON SWIFT

Our BI director, Susan, has quite a challenge in truly informing and giving excellent advice to 100 graduating MIS students. What might be most important is to keep their attention through actual (real) examples from her career or others' experiences to enlighten and also entertain the audience.

Speaking to 100 varied minds is not an easy task. During my 40-year career, I've faced this daunting objective dozens of times, mostly with experienced long-term IT, BI, or business executives who were attending a conference or asking me to advise them on how to make big decisions about their future information technology investments and practices. These students will be making big decisions about their future employers and the technological expertise to be gained and utilized throughout their lives. This is the shared challenge that Susan must address.

In preparing for a speech about business intelligence, we might also include definitions and good business examples of data warehousing, data integration, Internet traffic analysis, business analysis, and "big data" skills. These BI subjects (and skills) intersect in the appropriately defined IT and business organizational constructs, as well as in the skill sets to achieve a successful career in BI. Primarily, firms seek skills as well as business knowledge (accounting, budgeting, planning, marketing, services, operations, and educating a varied group of users).

There is an adage about making a great speech: "Tell them what you are going to them, tell them, and then tell them what you told them." With varied audiences, in a non-technical subject area, it is wise to add flourishes and great examples to the "info-tainment" aspects of the educational and informative speech. Primarily, I would make points using cases and examples of real experiences, including those of my peers and vendors.

Here are the key messages I would leave with this group:

1. Communication is the foundation of success. It helps you sell ideas, make your points, contribute to a team or project, and position yourself for additional interactions and opportunities. (Don't forget to become a great listener—don't speak over others' thoughts.) Becoming a great communicator will take you farther and faster than you imagined, and you will also be called upon to stand in for or replace others without such

skills (including high-ranking positions in your firm).

Part of learning to communicate is to practice, practice, practice. This goes for any speech, proposal, or presentation of ideas; if necessary, practice in front of your family, listening for questions and clarifications. An IBM class manager told me in 1970 that I was the worst presenter he'd ever seen among thousands of students passing through his classes. Ten years later, I was presenting to thousands in conferences throughout the world! Practice and learn.

2. Never stop educating and learning. This is the hallmark of a highly successful professional (especially in a technology field). The ever-increasing body of knowledge provides a challenge but also an opportunity to move on to more interesting jobs and higher salaries. For example, as a college graduate/trainee, I started as an IBM systems engineer and eventually became a nationwide marketing manager for data warehousing and business intelligence, before retiring early at age 47 and moving on to consulting and then a vice presidency in a similar field (marketing technology and ideas) to executives and technical people worldwide for NCR Corporation. Then I retired again at 62. Re-educating and learning from others constantly throughout your career will provide unexpected pathways and opportunities.

   2B. "Know what you don't know," and constantly seek people and resources where you can find information you need or at least pass your questions on. Many successful high-tech and analytical managers have benefited from this approach.

3. Networking is essential for succeeding in our business world. It can not only open doors to new jobs, but also build a volume of references, friends, supporters, mentors, and others who will be part of your joint success. Susan should tell the audience that their network starts right there in the very room where they are sitting, in all of their classes, and with all of their professors. Advise the students to build strong relationships with professors because they hear of jobs that may fit their skills and personalities. A professor can be a mentor for life.

   Students should expand and maintain their network relationships, especially with all the new personal communication devices that are available now, and keep in touch throughout their career. They may need to choose between hundreds of Facebook "friends" versus networking with peers, especially those who are positive in their thinking and actions. Your network can get you jobs you never considered, and college alumni are with you forever if you keep up your participation. Do not lose these "friends." It is said that Harvard and Stanford are not colleges, but clubs for life!

4. Analytics skills are of the highest value today in many companies, and they can be useful in areas you might never think of when majoring in information technology (i.e., MIS). The Web offers a myriad of new opportunities and is the beginning of the major wave of the future, which is currently called "big data."

   In fact, big data has been around for 30 years, depending on your perspective and the available technology to support it. What is different today: the unstructured data and the high volume and velocity of the data becoming available for analysis to bring value to the organization. Knowledge of sophisticated data management (meaning integration and analysis) of such volumes is key for most enterprises, many of which are not utilizing the available software and services to achieve the highest value. This is an area that will continue to blossom and provide massive numbers of jobs to construct and utilize the architectures and resources required to analyze and bring value to the data. This is still

an emerging area, but good examples exist from many BI vendors and would help round out Susan's message.

5. Positioning creates opportunities. One of the most practical but also the wisest of hints I received early in my career was: "Look smart, be smart, feel smart, speak smart, and listen to smart people around you." Using this as a guideline, develop a positive position with your peers and your upper management. Seek opportunities to interact with those who can advance your knowledge and experiences. Dress for your industry and your company (or department), but don't position yourself as a geek, even if you are one. You create your own brand.

   Make sure others desire to be around you. As a well-dressed professional, you may even be noticed on the elevator by someone who may be inclined to bring you onto their team. Positioning also means accepting challenges that may not be fully achievable but contribute to a team that has ambitious goals. Your contributions will be noticed, your attitude will be appreciated, and your teamwork will be a stepping stone to other teams.

   The right positioning will help you move within your organization. I have only had five formal interviews in my entire career (after the first one at the job I accepted at IBM). Four of these interviews were internal and moved me into new positions where I was probably pre-sold by others who advised the team leader or the manager of my skills, experience, personality, teamwork, and contributions. These conversations were ultimately more of a check on my personality and career desires than interviews for the specific jobs.

   The other interview was with a VP of data warehousing marketing at AT&T (to become NCR), who spent the better part of three hours showing me his plans and his slides about what they planned to do. I spoke for maybe 15 minutes during this time, mostly answering non-technical questions about industry contacts, relationships, and experience. I guess I was "pre-sold" to him!

6. Identify what you like to do. You have a myriad of questions on your mind about BI and a career, but consider this: you will probably have five to eight major jobs in your lifetime. Do you like sitting at a screen or working by yourself all day? You could be working in that mode for a long time. Do you like people, interactions, varied experiences, and new challenges? You will probably be offered many unexpected job opportunities to grow your skills, your network, and your supporters. Think about successful people you know—beyond those single inventors of specialized technologies. I'll bet they all have communication skills, a great network, good peer recognition, create value for their audiences, and give food for thought. Do what you think you will enjoy, and seek out new joys!

7. There are many great opportunities in different industries: healthcare, pharmaceuticals, finance, transportation, logistics, supply/inventory, drones, marketing, services, and so on. Business intelligence provides both the foundation and the innovation behind these industries' power for financial success. If you can find a way to share in that, you will enjoy your career and secure a great retirement.

My additional suggestion for Susan and her audience is to read books by or about Bill Gates and Steve Jobs. These are important testimonies to the value of selling creativity and innovation in a difficult industry.

Remember, the gift of knowledge is one thing; the gift of being able to communicate and convince others of your ideas or your products, services, or value brings a whole new world of opportunities. Good luck!

COY YONCE

Susan has a unique opportunity to inspire an incoming group of eager BI professionals. At the same

time, she needs to ensure that she sets proper expectations about why organizations implement BI and how they go about creating successful deployments of a BI suite. Susan's success in properly exciting this group of young adults about BI will depend upon how she defines what BI is and why it is important. To this end, Susan's first task is to start at the beginning: Why do organizations leverage BI?

Business intelligence provides organizations with information to make informed decisions about their day-to-day business and long-term strategy. Susan would serve the students well by providing real-world examples for how BI helps organizations to be better at what they do. She should talk about how all of the bits of information that companies such as Best Buy and Amazon gather about their customers helps them to make better decisions about how to market to us, how to stock more of what we like, and how to determine prices to which we will respond positively. Even talking about how Facebook, Twitter, Google, Netflix, and Hulu use data to make decisions will go a long way toward getting these young adults excited about the usefulness of BI.

In addition, Susan should spend time explaining that BI is not a tool. It is the act of tying all business processes to actionable data and then understanding how to use that data to make better decisions. This will ensure that the students do not think that BI is just software. The software just makes it easier to implement BI. This will allow for an easier transition into the next part of the conversation: a career in BI.

Susan's first task is to start at the beginning: Why do organizations leverage BI?

As business intelligence is tied to goals that ensure the long-term success of an organization, department, or project, Susan should speak about how all individuals within an organization are the consumers or creators of BI. Everyone who works for a company is responsible for ensuring that company's success. They all play a role both in using information and creating information for others to use. Susan should then speak about the more traditional roles of those who seek out, install, maintain, and deploy BI solutions, but she should note that this is really only a fraction of the story when it comes to a career involving BI.

Although getting the software installed and running are important, it's even more important to ensure that the organization understands how to leverage data toward the goal of becoming more agile and responsive. Susan should talk to the class about the role of business analysts as they help executives, upper management, and departments decide how to use data for activities such as increasing public safety, starting new employee benefit programs, acquiring other companies, or shutting down a manufacturing plant.

Susan should highlight how important it is to ensure not only that accurate data is being used, but also that it is presented in a way that is easy for the consumers of that information to understand. Talking about the role of a data scientist and the study of information design will be an important part of Susan's discussion.

Rather than jumping into talking about preparing for a career in BI, Susan should focus her initial message on understanding how students can be good stewards of the organizations they are running or for which they work. She should then discuss how they can leverage their own strengths and passions toward the end goal of helping those organizations. If an individual is interested in programming, then their role in relation to BI could be working for a software company that is creating BI or helping the organization augment a chosen BI suite. If they are more interested in marketing, then perhaps their best role will be as a business analyst within the marketing department or working as the face of IT as they implement a BI solution. Trying to stick to the more traditional roles related to BI or condensing the message down to a list of roles would do a disservice to the importance of BI and will likely turn off the younger audience.

Susan should close by talking about the exciting things that are happening in BI, but she should do so within the context of how those exciting things are being used. Susan should also be wary of using buzzwords. Big data is a hot topic but not necessarily a new concept. There has always been lots of data available for analysis and consumption. What is new is the wide availability of technologies to process large amounts of data quickly.

Talking about the role of a data scientist and the study of information design will be an important part of Susan's discussion.

More important are the examples of how these technologies and data are being used. Susan should consider talking about Project Artemis, which is being run by IBM and UOIT (http://hir.uoit.ca/cms/?q=node/24). The project's goal is to capture and process large amounts of information about premature babies and then make the results available to physicians and nurses in real time.

Susan could also talk about how augmented reality technologies are being paired with business analytics by SAP. This combination could allow organizational leaders to track inventory, access machinery maintenance records, track sales of a product, or access patient's health records, all by pointing a mobile device at the product, machine, or person.

The end result of this approach to her speech will be an engaged and interested audience of students who are excited about the possibilities of BI. ■

Data Governance Gamification

Justin Hay

Abstract

This article discusses the application of gamification mechanics to data governance implementations. It outlines the goals of data governance, particularly in the context of data warehousing, and introduces the concept of gamification. It then explores five key motivators and shows how each would be addressed by a gamified governance program.

Ten gamification mechanics are discussed with notes on how each relates to data governance. A hypothetical gamified governance program is described in story form, including notes on its practical application. The article concludes by advocating a gamification approach as part of a cultural trend to engage a younger generation of business users who have been raised on games as an integral part of their lives.

Justin Hay is vice president of solutions with Stream Integration. He is also the author of the blog "EDW Strategic Positions" (justinhay.wordpress.com). justin.hay@streamintegration.com

Introduction

> When governments, businesses, and organizations of all sizes and types embrace game thinking and mechanics, they are better able to engage their audiences, cut through the noise, drive innovation, and ultimately increase their revenue.
> —*The Gamification Revolution*

Building a data warehouse is a large and challenging undertaking for any organization. One of the biggest challenges involves ensuring the accuracy of the information that is presented to end users. This requires data governance, including:

- Governance of the **business metadata**, including the names of terms being used, ensuring they are accepted across the enterprise, and that they conform to a naming standard (including work order, class words, and standardized abbreviations)

- Governance of the **definitions** applied to those terms, such that they are meaningful, unambiguous, and well-formed

- Governance of the **data architecture**, so that the information integrity remains sound and users can access it with relative ease and confidence

- Governance of the **data content**, ensuring that the information conforms to business rules and meets standards of completeness and correctness

Data governance can be a dry and time-consuming affair, frequently necessitating lengthy committee meetings and careful maintenance of lines of communication between business users and IT staff. Typically, the list of data elements to be processed with names, definitions, and business rules is longer than a program can deal with in advance of a data warehouse development effort. As a result, governance may be conducted in parallel, creating inevitable rework as late-arriving governance changes are enacted.

Gamifying data governance is about working to motivate employees to actively engage in the process and employing ideas from games to do so.

This article proposes an alternative to the traditional committee approach to data governance based on the gamification movement as articulated in such books as *The Gamification Revolution* (Zichermann and Linder, 2013), *Loyalty 3.0* (Paharia, 2013), and *Nudge* (Thaler and Sunstein, 2009). Gamification involves mechanics that help motivate and incentivize participants to perform a set of tasks. This article draws from the ideas of these books with regard to gamifying business processes in general, and applies them specifically to the activities involved in data governance.

Applying concepts from games to achieve business goals is an ancient idea. Contests have been used extensively as a marketing tool for decades, eliciting an outpouring of creative public participation for everything from breakfast cereal to the design of national monuments. More recently, technology has opened up possibilities for customer loyalty programs, where people gain points for purchases and receive special offers and perks when certain levels are achieved. Educational applications are also using gaming conventions to help motivate users. An excellent example of this is Duolingo, which helps users learn a language while encouraging them with points, achievement levels, and rankings against friends.

Here is a breakdown of a potential gamified data governance program:

- **Overall concept.** The responsibility for data governance is taken on by the business user community as a whole. The process is gamified to set group targets and monitor the progress of the program through measurable sets of activities and milestones.

- **Goals.** The data governance initiative strives to improve the quality of business data by establishing policies and rules to be applied to the data and measured and monitored on an ongoing basis. The program will also work to improve the collective knowledge and understanding of that information by business users across the enterprise by involving the community in creating, assessing, and utilizing the business metadata.

Motivators

Gamifying data governance is not about making a game out of the governance process for the entertainment of employees. It is about working to motivate employees to actively engage in the process and employing ideas from games to do so. In *Loyalty 3.0*, Rajat Paharia identifies five key motivators:

Autonomy. For broad acceptance, it is essential that data governance not be imposed upon the business user

community. As a gamified process, the responsibilities are shared across the user base, with individual responsibility to create content and the ability to give feedback responses actively influencing the outcomes.

Mastery. There is a certain degree of skill involved here, particularly in the tasks related to lexicography, that is, forming the definitions of terms. Making these concise and meaningful to a range of users can be difficult and requires considerable communication skills. Similarly, crafting policies and rules to ensure the integrity of data also requires discipline and attention to detail. Getting it right will involve the mastery of the processes involved.

Purpose. The purpose is built in. Data governance is already a necessity; gamification is a means of driving the process forward. The ultimate goal is to improve the business, optimizing the data assets at hand to increase return on investment. Gamifying the program also arms management with levers to target specific subject areas, such as customer information and marketing-campaign-related data. The program can be tuned to meet both strategic and tactical goals as required.

Data governance is already a necessity; gamification is a means of driving the process forward.

Progress. Gamification mechanics enable the program's progress to be measured. Each participant can see the extent of his or her individual contribution, as well as the entire community's progress toward the ultimate or immediate goal. Progress can be seen in different terms depending on management objectives such as short-term sprints, specific sets of terminology, alignment with data warehouse releases, or targeted business units. This stands in stark contrast to the never-ending list of business terms that the governance council addresses on a weekly basis in many programs, where the distance from the end goal can be difficult to determine.

Social interaction. In the context of data governance, this aspect relates to business users working together toward a common goal within business units, across lines of business, and between business and IT. Technical and business understanding of the data is involved, and an understanding of the data life cycle—from data entry to business intelligence—will need to be communicated. Everyone must be engaged and participating for the information to be entered correctly and trusted when it is used. The gamified data governance program encourages communication between users, employing both cooperative and competitive gaming mechanics.

Gamification Mechanics

Gamification mechanics are the methods by which gameplay concepts are applied to other spheres of activity. *Loyalty 3.0* identifies 10 different mechanics, which we'll discuss in relation to a gamified data governance program. Each of these mechanics combines a number of the five key motivators just listed.

Fast feedback. Users receive immediate response from their actions. In our governance program, users gain points by contributing data definitions to the business glossary, reviewing and/or correcting existing definitions, and stewarding accurate data with defined policies and business rules to ensure data integrity and usability. Conversely, users are penalized for poorly written definitions, non-standard names, and inaccuracies in data for which they have stewardship responsibility. More than just points, the process is interactive across the program; names and definitions are assessed and receive an active response from other users.

Transparency. Transparency is about making the process clear to everyone. Users can view graphs and reports to see their level of achievement, which activities have had the greatest impact, and how their contributions to the program compare to the wider community.

Goals. With the motivator of purpose in mind, the goals can be both immediate and long term. Users are invited to set personal goals as well as respond to individual and group challenges. Goals can include setting a certain number of definitions, mapping a set number of sources

to targets, completing business rules, or achieving metrics on data quality.

Badges. Badges are essentially a public sign of achievement. Users are awarded them for reaching levels of data quality, for their active participation in the data governance program, and for specific achievements. These badges can be displayed on Intranet sites, posted on the walls of cubicles or lunch rooms, and embedded in e-mail signatures. Some of the badges are advertised and the criteria to win them clearly articulated; for others, their existence is made public, but the means to achieve them remains hidden; still others are kept secret and awarded as a surprise.

Users can view graphs and reports to see their level of achievement, which activities have had the greatest impact, and how their contributions to the program compare to the wider community.

Leveling up. Both individually and as teams, users are taken through levels of participation and achievement based on scores and badges. Although climbing levels should initially be quite simple (e.g., level 1 to 2 requires the user to add five new data definitions), each subsequent level will be increasingly difficult to achieve (e.g., level 2 to 3 might require 10 new definitions). Management defines the objectives of the program, so the levels can correspond to these goals. A level can be defined as being a domain of information (for example, customer information), the completion of a set of policies (business rules have been applied to all vendor address information), or achieving a level of data quality (99.9 percent of customers have an identified representative).

Onboarding. Business users will join and leave the program over time. New users must be able to join as easily as possible. The onboarding process will introduce users to the tools, conventions, and processes involved, as well as the data governance program applications: enterprise business glossary, data profiling tools, and dashboards to monitor data integrity. Users will need to become conversant with naming and definition conventions and how business rules are applied to information. This is done in the context of the gamified program, incrementally showing users how to gain points and win badges.

Competition. Users can see the leaderboard of the most productive people in the company or by line of business (or any other group) as well as their particular placement in the ranking of the program and the ranking of all other participants. Individual and group scores show users their level of contribution to the program. In this way, the program utilizes both competitive and cooperative drivers.

Collaboration. The whole data governance program works toward the same collective goal. The collaborative nature of the work can be underlined by identifying end-to-end data governance needs by project. Just as testing of development work must be conducted by a third party, the crafting of definitions and rules must be checked and confirmed through reviews and testing. It will take a multi-disciplinary effort to ensure that the program works.

Community. The idea of community in this context is about the visibility of everyone's work. Too often, work is pushed off from one group of business users to another or from business users to the IT department. Here, everyone sees the work that everyone else is doing. It becomes clear to everyone, management and users alike, who is pulling their weight and who is dragging the process down. If someone needs help, the community can pitch in.

Points. Points are awarded and deducted; they show individual and group status and can be redeemed for rewards. This provides the fast feedback mentioned earlier and supplies the measurable result that is key to determining the progress and effectiveness of the program.

A Gamified Data Governance Story

A story about a hypothetical gamified data governance program will best illustrate how this approach might be put into action. (Note: The concepts of ideators, collaborators, and connectors were drawn from *The Gamification Revolution*.)

NE Corp. had a problem with their business users' understanding of business terminology. Over the past five years, NE had launched a number of initiatives to define business terms within an enterprise glossary, but there had been little engagement from users, and the definitions produced were either poorly written or not accepted by the general community.

Mistakes and miscalculations were costing the organization millions of dollars annually. Terms were ambiguously defined, leading to the wrong data being used for reporting. Policies regarding what constituted complete and accurate information were absent, and rules to measure and continually monitor the quality of data were not in place.

Participants in the governance initiatives cited several causes for their failure. The cleanup task was daunting, with thousands of terms to be defined, policies and data quality rules to be established, and links to where terms appeared in databases to be identified. The connections between the data stewards who "owned" the domain of information, the business users who produced and consumed it on a daily basis, and the information technology staff who managed it were weak or nonexistent.

Management knew they needed to try something innovative to fully engage users in the process across the spectrum of stakeholders. Enter Gamified Data Governance (GDG): a gamified approach to data governance.

With GDG, the governance of information was taken on by a wide set of information producers and consumers.

The program had three sets of participants:

Ideators. This group consisted of data stewards responsible for drafting the definitions, policies, and rules that governed the data. Ideators could see lists of complete and incomplete terms. Their focus was on lexicography and the data dictionary aspects of the business metadata.

Collaborators. These business users worked with the metadata generated by the ideators. Their role, within the context of the program, was to review and score the work the ideators produced. They also provided feedback on the connectors' work when mistakes were found. Collaborators saw lists of reviewed and un-reviewed terms based on terms with completed definitions. Their role was to ensure that names and definitions were accurate and meaningful. They provided feedback on the ideators' work.

Connectors. These IT personnel were responsible for linking data assets to the terms. Connectors saw lists of terms with numbers of attached data assets by source system. The data assets were part of operational systems as well as the data warehouse, MDM system, and data marts, among other sources.

The gamification mechanics of GDG involved participants earning points through one of the three sets of activities. Ideators earned points by defining terms, policies, and rules; collaborators earned points by reviewing and rating the terms' definitions; and connectors earned points for each data asset mapped to a term.

Levels of achievement were awarded based on points earned. For example, ideators received 1 point per definition. Level 1 was reached when a participant amassed 5 points, Level 2 required 15 points, Level 3 was set at 30 points, and Level 4 required 50 points.

Badges were awarded to individual participants based on the completion of certain units of work. These units were based on a subject area domain of terms (e.g., customer, account, organization).

Trophies were awarded for cross-group achievements, such as a set of customer information having a complete set of reviewed term definitions with source and target data assets attached.

Individuals worked within defined team groupings. Leaderboards were displayed at both the individual and team level.

The GDG program appeared as a Web page to participants, displaying personal and group scores as well as achievement levels, badges, trophies, and leaderboards. The scores were drawn from calls to the business glossary metadata. Each activity, such as the addition of a term, the definition of that term, or the inclusion of a business rule, triggered the game engine to add a point to the user's running score. The GDG program was customizable and permitted a given participant to perform more than one role.

The program was run in six-week sprints, each of which had a defined scope of terms to be included. Scopes coincided initially with the releases of the MDM program and later with iterations of the data warehouse, the GDG program being integral to the quality of both products. The limited timelines helped to focus the energies of participants and give a sense of closure to each sprint, even though the overall work entailed a much longer time period.

The success of the GDG program was measurable. Engagement from the participants was high, and with peer review included, the quality of the ideators' work quickly achieved and maintained a high standard. The attention and collective engagement focused on the work led to a higher than anticipated adoption of the business glossary for all stakeholders, and there was a marked improvement in data quality adherence to policies and rules.

Summary

The authors of *The Gamification Revolution* point out that the millennial generation demands this type of gamified engagement in the workplace and that business will ignore the gamification of its processes at its peril. *Loyalty 3.0* calls out gamification as a trend spanning all industries, including healthcare, banking, retail, telecommunications, education, and government, specifically relating to the loyalty of employees.

There is evidence of a widespread cultural shift toward gamification. Look at the gamified mechanics of electronic readers, the collection of "likes" conferring status on Facebook, and the loyalty programs of cafés, cinemas, and grocery stores. This shift is having an impact on industries of all types and presenting new challenges to harness the torrent of gamified data.

The cultural movement toward gamification also offers fresh opportunities to engage and motivate employees, many of whom come from a generation raised on the interactivity of new media.

Business must govern data. Information is piling up and must be addressed with discipline and rigor. Traditional methods and organizational structure may not garner the level of participation required to be effective. Gamification does not trivialize this work, nor will it necessarily transform the sometimes tedious task of governing data into entertainment. However, people perform better when they are motivated, and gamification mechanics can provide that motivation. Gamification offers a way to ensure that the business wins. ■

References and Related Reading

Burke, Marsha, and Troy Hiltbrand [2011]. "How Gamification Will Change Business Intelligence," *Business Intelligence Journal*, Volume 16, Number 2.

Paharia, Rajat [2013]. *Loyalty 3.0: How to Revolutionize Customer and Employee Engagement with Big Data and Gamification*, McGraw-Hill.

Thaler, Richard H., and Cass R. Sunstein [2009]. *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Penguin Books.

Zichermann, Gabe, and Joselin Linder [2013]. *The Gamification Revolution: How Leaders Leverage Game Mechanics To Crush The Competition*, McGraw-Hill.

Enterprise Data Modeling: Lost Art or Essential Science?

Larissa T. Moss

Larissa Moss is president of Method Focus Inc., which specializes in improving the quality of business information systems. methodfocus@earthlink.net.

Abstract

Enterprise data modeling (EDM)[1] got a bad reputation in the late 1980s, and for good reason: it took too long to model the entire enterprise, and the model had little or no effect on how systems were built. It was perceived as a wasted effort, and the practice was abandoned by most companies. However, in the late 1990s it became clear that we had thrown the baby (sound data administration principles) out with the bath water (tedious and ineffective approaches). As a result, new disciplines started to emerge, such as metadata administration, data governance, data stewardship, data-quality improvement, master data management, enterprise information management, the information center of excellence, the integration competency center, and so on.

These new disciplines all strive to achieve the same goal as EDM: to create and maintain an inventory of data assets that are unique, accurate, reusable, and traceable. This article explains the origin of data modeling and the significance of EDM; compares top-down, bottom-up, and middle-out EDM approaches; describes related data administration principles; and shows how data quality and master data management (MDM) relate to EDM.

Origin of Data Modeling

In the early days of data processing, systems development was focused on automating manual business processes such as order processing or payroll. Thus, all the early modeling techniques in the 1970s concentrated on process diagrams (data flow diagrams, functional decomposition charts, structure charts) popularized by Gane-Sarson, Ed Yourdon, Tom Demarco, Larry Constantine, and others. In 1976, just as process-oriented

[1] I will use the abbreviation EDM for both enterprise data modeling and the enterprise data model.

modeling techniques were becoming standard in system development methodologies, Dr. Peter Chen broke the pattern of process-oriented system development with his invention of the entity/relationship (E/R) model, which soon became known as the logical data model.

E/R modeling was revolutionary in that—for the first time—data, not processes, were at the center of both business analysis and system design. The implications were enormous: data could now become a reusable commodity, which meant that every unique data element was identified and inventoried once and only once. Business analysts and data modeling practitioners realized they could finally create a business model of their organization that would logically portray a non-redundant, single version of the truth of their enterprise in terms of its data resources. Companies created data administration (DA; also known as DRM, IRM, or EIM)[2] departments to manage their business data as a corporate asset, just as they managed their financial assets, fixed assets, real estate, or human resources.

At the same time, Dr. Edgar F. Codd realized that data could also be physically organized and stored independent of any process, so that many processes could share the same data. In his seminal article "A Relational Model of Data for Large Shared Data Banks," Codd presented the relational model. In 1985, a few years after the first relational DBMS (DB2) was born, Codd published the famous 12 rules of relational databases (actually 13 rules when counting Rule Zero). The power of today's RDBMSs is still inextricably tied to the concepts of Peter Chen's original E/R model.

Significance of Enterprise Data Modeling

The greatest benefits of logical data modeling are achieved from building the 360-degree view of a business (i.e., the EDM). The difficulty in building this view is that the current data chaos in most organizations is so immense that it may take significant time and effort to rationalize the existing data into an integrated, nonredundant EDM.

Many people confuse data integration with data consolidation. Consolidating data simply means gathering data elements that identify or describe the same business object (e.g., customer data or product data) from multiple source files or databases and storing them in one table or in a set of dependent tables. Integrating data goes much further. In addition to consolidating data, integration enforces data uniqueness—the building blocks of the single version of the truth that enable you to reuse the same data without the need to duplicate it. Data integration requires several actions during enterprise data modeling. An organization must:

1. Examine the definition, semantic intent, and content of each logical entity to find potential duplicates of business objects that would otherwise not be discovered because the objects are known under different names in the systems.

2. Ensure that each entity instance has only one unique business identifier, which, in turn, is never reassigned to a new entity instance even after the old instance has expired.

3. Use the six normalization rules to put "one fact in one place;" that is, one attribute (data element) in one (and only one) owning entity. This means that an attribute can be assigned to only one entity as either an identifier or as a descriptive attribute of that (and no other) entity. This modeling activity ensures that each attribute is captured once and only once, and that it remains unique within the data universe of the organization. Hence, the single version of the truth (no synonyms, no homonyms).

4. Capture the business actions (or business transactions) that connect the business objects in the real world. These business actions are shown as data relationships among the entities. It is paramount to capture them from a logical business perspective (not from a reporting pattern or data access perspective) because these relationships are the basis for all

[2] Data resource management, information resource management, enterprise information management

potential access patterns, known and unknown, now and in the future.

Top-Down, Bottom-Up, and Middle-Out Approaches to EDM

The top-down EDM technique facilitates fact-finding discussions with participants from various business areas who own, originate, or use the data. The participants identify the major business objects, the data relationships among them, the unique business identifiers of those objects, the most significant and obvious attributes of those objects, and the main business rules for those attributes. This data modeling technique produces a relatively valid EDM in a rather short time. However, the EDM will be incomplete because it is too time-consuming to identify and model all the complex business rules during these types of modeling sessions. (Forcing business people to attend a never-ending number of modeling sessions is how EDM earned a bad reputation in the 1980s.)

Bottom-up EDM involves the painstaking task of normalizing existing process-oriented (denormalized) data structures into "best-guess" logical data models and merging these models into an integrated, non-redundant EDM. Using the bottom-up data modeling technique produces a relatively complete EDM. However, an EDM based on "best-guess" logical data models cannot be trusted until it is meticulously reviewed and validated by all the business people who own, originate, or use the data. In addition, many business rules cannot easily be extracted from existing data structures and would therefore be missing in the EDM, rendering the EDM incomplete at best, and incorrect at worst.

The term "middle-out" logical data modeling describes the approach of building the EDM iteratively using both techniques, starting either top-down or bottom-up and then continuously alternating between the two approaches. The advantage of this technique is that it can be used by data administrators or enterprise information architects to build an EDM for their data governance or enterprise architecture programs as well as by BI and DW teams to reverse-engineer source files and databases for their integrated ETL process, thus contributing to the EDM.

Data Administration Principles

An EDM is not merely a simple high-level pictorial representation (E/R diagram) of an organization's data resources. Its ultimate value comes from applying stringent data administration principles during the logical data modeling process. For example, most people recognize that creating business metadata, such as data definitions, is very important, but even more important is applying a formalized structure to the data naming process. Using "favorite" names or blindly copying informal names from source systems is not allowed.

Forcing business people to attend a never-ending number of modeling sessions is how EDM earned a bad reputation in the 1980s.

Without formal data names, data elements cannot readily be recognized, or they may be misidentified and therefore misused. An unrecognized data element may be unknowingly duplicated as a synonym, or a data element name may be reused as a homonym for another data element that is described much more accurately by that name.

There are numerous data-naming conventions, the most popular being the "prime words/class words/qualifiers" convention. It prescribes that every data element must have one prime word, one or more qualifiers, and end in one class word (qualifiers can apply to both prime and class words). Class words are predetermined and documented on a published list (e.g., date, text, name, code, number, identifier, amount, count, etc.). Furthermore, every data element must be fully qualified (to avoid homonyms and to avoid limitations on naming future data elements), and it must be fully spelled out. An example of a standardized business name for a data element is "checking account monthly average balance."

The main component (prime word) is "account," which is further qualified by the word "checking" to indicate the type of account. The class word indicating the type of data value contained in this data element is "balance," which is further qualified by the words "monthly" and "average" to indicate the type of balance. In his book *Data Resource Quality*, Michael Brackett describes other data-naming taxonomy components, such as data site, data subject, data occurrence role, data characteristic, and data version.

Another DA principle is the formalized process for creating business-data definitions. The definition should be short, precise, and meaningful (a short paragraph). It must thoroughly describe the data element name (i.e., what it is) and, optionally, it may contain an example. A data definition should never contain information about the source or use of the data element. Data definitions should be reviewed regularly to ensure that they remain current and correct, and that they are understandable and agreeable to all business people.

One of the logical data modeling rules is "one fact in one place." This rule refers to the DA principle that every data element must be unique (it must have one and only one semantic meaning), and, therefore, every data element (attribute) belongs to one and only one entity[3]. In other words, detail overload is not allowed. You cannot have a general "type code" attribute where the values A, B, and C correspond to "type of customer" (A= corporate, B= partnership, C= individual) and the values D, E, and F correspond to "type of product" (D= mortgage loan, E= consumer loan, F= construction loan).

A related DA principle states that all data elements must have a specified business data domain, which refers to data values that are allowed with respect to the data name (specifically the class word), data definition, and its business rules. Data domains can be expressed as a list of values, a range of values, a set of characters, or pattern masks.

[3] This rule does not apply to foreign keys because they are not part of logical data modeling. Foreign keys are physical manifestations of logical data relationships.

Data normalization is another adopted DA principle. It is fundamental to all other DA principles because, by following the formal normalization rules, you can ensure that all unique descriptive attributes of an entity are fully dependent on the unique business identifier of that entity.

This is the premise for modeling the single version of the truth or 360-degree view of one's business.

Data-Quality Improvement with EDM

Four categories of data-quality rules are applied during logical data modeling:

Rules about business objects (entities): Entities are subject to three data-quality rules: uniqueness, cardinality, and optionality. Uniqueness specifies that every entity has a unique business identifier. Cardinality refers to the degree of a relationship; that is, the number of times one entity can be related to another. The "correct" cardinality in every situation depends completely on the definition of your entities and the business rules governing those entities. Optionality is a type of cardinality, but instead of specifying the maximum number of times two entities can be related, it identifies the minimum number of times they can be related. There are only two options: either two entities must be related at least once (mandatory relationship) or they don't have to be related (optional relationship).

Rules about data elements (attributes): Attributes are subject to two data-quality rules: data inheritance and data domains. Data inheritance rules apply to supertype and subtype entities. Data domain rules refer to a set of allowable values, as discussed earlier

Rules about the dependencies between entities or attributes: This category includes seven data dependency rules: three for entity relationships and four for attributes. The three entity relationship dependency rules are:

- The existence of a data relationship depends on the state (condition) of another entity that participates in the relationship.

- The existence of one data relationship mandates that another data relationship also exist.
- The existence of one data relationship prohibits the existence of another data relationship.

The four attribute dependency rules are:

- The value of one business attribute depends on the state (condition) of the entity in which the attributes exist.
- The correct value of one attribute depends on the values of two or more other attributes.
- The allowable value of one attribute is constrained by the value of one or more other attributes in the same entity or in a different but related entity.
- The existence of one attribute value prohibits the existence of another attribute value in the same entity or in a different but related entity.

Rules about data validity: These rules govern the quality of data values (data domains):

- The data-completeness rule comes in four flavors: entity completeness, relationship completeness, attribute completeness, and domain completeness.
- The data-correctness rule requires that all data values for an attribute must be correct and representative of the attribute's name, definition, domain, and business rules.
- The data-accuracy rule states that all data values for an attribute must be accurate in terms of the attribute's dependency rules and its state in the real world.
- The data-precision rule specifies that all data values for an attribute must be as precise as required by the attribute's business requirements, business rules, intended meaning, intended usage, and precision in the real world.
- Five data-uniqueness rules address entities and attributes.
- Two rules enforce data consistency.

Since the activities for creating an EDM include the validation of the data-quality rules applied during logical data modeling to the existing operational data in the source files and databases, EDM contributes directly to improving data quality.

Master Data Management Using EDM

The combination of MDM and master metadata management is currently hyped as a new technique to ensure that an organization's data about its core business objects[4] is unique, consistent, reliable, and traceable. Many definitions exist for MDM:

- "A set of disciplines, applications, and technologies for harmonizing and managing the system of record and system of entry for the data and metadata associated with the key business entities of an organization." (Colin White, Claudia Imhoff)
- "MDM comprises the business applications, methods, and tools that implement the policies, procedures, and infrastructure to support the capture, integration, and subsequent shared use of accurate, timely, consistent, and complete master data." (David Loshin)
- "MDM is a set of corporate disciplines that ensures that corporate reference data, such as charts of accounts and customer information, is presented in a clear and consistent manner across the enterprise." (Al Moreno and Greg Mancuso)
- "The organization, management, and distribution of corporately adjudicated, high-quality information with widespread use in the organization for business gain." (William McKnight)

[4] Core business objects are the kernel entities on an EDM

MDM is certainly much more than EDM because of its technical implementation with various products and tools. However, it does have much in common with EDM. The most striking commonality is summed up by the title of Philip Russom's report published by TDWI in October 2006: Master Data Management: Consensus-Driven Definitions for Cross-Application Consistency. As described in previous sections, one of the DA principles applied during EDM is creating precise, consensus-driven definitions, along with consensus-driven names, domains, and business rules.

Almost all MDM articles mention that MDM needs data governance. Some definitions of data governance include:

- "The execution and enforcement of authority over the management of data assets and the performance of data functions." (Robert Seiner)
- "The process by which you manage the quality, consistency, usability, security, and availability of your organization's data." (Jane Griffin)
- "A process and structure for formally managing information as a resource; ensures the appropriate people representing business processes, data, and technology are involved in the decisions that affect them; includes an escalation and decision path for identifying and resolving issues, implementing changes, and communicating resulting action." (Danette McGilvray)

These definitions, especially the last one, also describe EDM.

Conclusion

EDM is a mature and proven technique that spans logical data modeling, business metadata management, data governance, data stewardship, data-quality improvement, and MDM. The use of Chen's E/R modeling technique, Codd's six normalization rules, and the extensive DA principles of the 1980s still make EDM the most effective technique for achieving the single version of the truth or the 360-degree view of a business, at least logically. Data management groups, such as DG, DQ, MDA, EIM, ICOE, and ICC[5], as well as project teams of physical implementation initiatives such as MDM, CDI, PIM, CRM, EDW, and ODS[6], would benefit greatly from leveraging the science of EDM. ■

REFERENCES:

Adelman, Sid, Larissa T. Moss, and Majid Abai. *Data Strategy*. Addison Wesley, 2005.

Brackett, Michael H. *Data Resource Quality*. Addison Wesley, 2000.

Codd, Dr. Edgar F. "A Relational Model of Data for Large Shared Data Banks." *Communications of the ACM* (Volume 13/Number 6).

Duncan-Hepp, Karolyn, and David L. Wells. "Rule Based Data Cleansing." *The Journal of Data Warehousing*, 1999.

Moss, Larissa, and Steve Hoberman. *The Importance of Data Modeling as a Foundation for Business Insight*, White paper, Teradata, 2005.

[5] Data governance, data quality, metadata administration, enterprise information management, information/integration center of excellence, integration competency center

[6] Master data management, customer data integration, product information management, customer relationship management, enterprise data warehouse, operational data store

Business Analytics Insight: Hype or Here to Stay?

Hugh J. Watson is a Professor of MIS and holds a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia. He is senior editor of the *Business Intelligence Journal*.
hwatson@uga.edu

Hugh J. Watson

We all know that analytics is a hot topic. It would be hard to miss the large number of books, articles, research reports, Webinars, and survey findings that suggest its importance. Despite the recent attention, I feel that analytics is not fully understood. There are many incorrect, imprecise, and incomplete understandings. I'd like to share 10 important and interesting insights about business analytics.[1]

#1: Analytics Has a Longer History than Many People Think

Even though the term analytics has become popular in the past few years, its roots date back many years. In the late 1960s and early 1970s, the first decision support applications appeared. They were most often called decision support systems (DSS), and over the years, this term was used to describe specific applications or as the name of a field or discipline (as it is still often used in academia). Over time, other decision support applications emerged, such as executive information systems (EIS) and online analytical processing (OLAP), which became part of the decision support applications portfolio.

In the late 1980s and early 1990s, data warehousing (DW), and soon afterward, business intelligence (BI), entered the decision support lexicon. BI, in particular, has been used as an umbrella term to describe the technologies (e.g., data cleansing tools), processes (e.g., data governance), and applications (e.g., dashboards/scorecards, OLAP) for supporting decision making. Today, the word analytics is often used as an umbrella term.

[1] I'd like to thank Doug Laney at Deloitte for his comments and perspectives on analytics.

Some people think of analytics as the data analysis component of BI, and that BI is a larger environment that includes everything needed to support analytics, such as a data warehouse. However, even when the focus is on mathematical and statistical modeling, this hearkens back to operations research and management science (OR/MS), which predate even DSS.

Why all the name changes? It isn't that the previous names were bad; more likely, vendors, consultants, writers, and others who offer decision support products and services saw the opportunity to have people and companies take a fresh look at their offerings by promoting them as new and different.

#2: "Analytics" Means Different Things to Different People

In the broadest use of the term, analytics involves the analysis of data using statistical and mathematical techniques. This includes performing multidimensional analyses (i.e., OLAP), calculating performance metrics for inclusion on dashboards/scorecards, conducting market segmentation analysis for designing CRM campaigns, and using mathematical programming for revenue management.

These are clearly very different kinds of analytics, and the differences have important implications for where they are used, who performs them, the skills that are required, and the technologies that are involved. For example, the skills needed to use and understand sophisticated data mining/predictive analytics techniques are different and greater than for performing OLAP. To avoid confusion, be clear about what kind of analytics you are discussing.

A useful distinction can be made among descriptive, predictive, and optimization analytics. As the names suggest, the different kinds of analytics describe a situation, make inferences about it, or suggest an optimal solution.

The following list categorizes the variety of analytics.

Optimization analytics

- Mathematical programming (e.g., linear, integer), simulation

Predictive analytics

- Decision trees, CART, genetic algorithms, neural networks

Descriptive analytics

- Data visualization
- Dashboards/scorecards
- Drillable/OLAP reports
- Published reports
- SQL queries

There are other classification systems that are appealing and useful. For example, Deloitte distinguishes among analytics that provide (1) an understanding of what happened (hindsight), (2) an understanding of why it happened (insight), and (3) an understanding of what will happen (foresight).

#3: Analytics Is Becoming a Competitive Requirement

I've used the term "BI-based organizations" to describe those firms where BI is more than a "nice-to-have" and is now a requirement for competing in the marketplace (Watson, 2010). The number of these firms is growing rapidly. Some of the earliest, most successful firms used analytics as an enabler to support new business strategies.

For example, when the gambling laws changed to allow gaming on Indian reservations and riverboats, Harrah's Entertainment adopted a growth strategy that included building new casinos and buying existing properties. The

strategy also included creating a Harrah's brand and incentives to encourage customers to play at Harrah's wherever they traveled. Key to this strategy was understanding customers well and extending offers that would entice them to gamble at a Harrah's casino.

Harrah's became an industry leader by using analytics in such areas as customer profitability and expected lifetime value analysis, customer segmentation analysis in campaign design, control and experimental group testing to better understand which offers work best with different market segments, and revenue management analytics to optimize hotel room revenue.

While most companies don't share Harrah's high profile in their use of analytics, many do depend on analytics to compete. They base decisions on data rather than intuition. Their understanding of customers' wants and needs are based on analytics. They run experiments to determine what works best, such as examining alternative Web site designs. Analytics are used to analyze current business processes and to design better ones, and analytics are integrated into work flows to monitor and increase their efficiency and effectiveness.

#4: Analytics Is Overhyped but Is Here to Stay

You can't miss all the resources (e.g., articles, books) that describe the wonderful things companies are doing with analytics. Just wait. It won't be long until there are publications that discuss the failures in using analytics. This hype-and-failure publishing cycle always occurs with new technologies and applications.

Is analytics overhyped? Probably so, but it is too important to go away. Analytics' ultimate success, however, Busi ness Analytics depends on how well companies make the required organizational (e.g., focus on fact-based decision making) and technological (e.g., a scalable, flexible BI infrastructure) changes, and the business value these changes create.

#5: Big Data Is Changing the Scope and Technologies for Analytics

We are in the era of big data. In addition to the usual structured data from operational systems, organizations are capturing and storing less-structured data from their Web sites, call centers, e-mail, documents, social media, and elsewhere. There are more data sources, and the data is arriving at a higher velocity. This vast amount of data contains a wealth of potentially useful information but creates challenges for capturing, storing, and analyzing it.

Is analytics overhyped? Probably so, but it is too important to go away.

If BI directors fail to plan for and integrate big data into their BI strategy, governance, architecture, technologies, processes, and activities, they risk facing a vacuum filled by the business units, resulting in a new generation of analytic silos.

Sentiment analysis illustrates the usefulness of being able to analyze big data in general and data from social media in particular. Consider a company that has introduced a new product and wants to know the public's reaction to it. Sales figures are useful and important, but what the public is saying about the new product is also valuable.

Using data from social media and employing text mining techniques, a company can learn whether they have a winner or loser on their hands or what problems need to be resolved quickly.

New technologies are emerging that can be used to store and analyze large amounts of poorly structured data. For example, Hadoop and MapReduce are open source technologies that distribute data and analyses across many computers, breaking the analysis into many parallel workloads that produce results faster. Companies such as Yahoo!, Twitter, and eBay that must work with huge amounts of data were some of the early adopters. In addition to these technologies, BI directors must become familiar with the technologies and processes for capturing new kinds of data and the specialized tools for analyzing it.

#6: BI Platforms Are Changing

BI platform discussions once focused on which vendor's offering was the best (e.g., Teradata, IBM, Oracle) and whether the Inmon (hub-and-spoke) or Kimball (datamart-bus) architecture should be used. Now there are many other considerations. For example, data warehouse appliances from vendors such as Netezza (now an IBM company), Greenplum (now part of EMC), DATAllegro (now part of Microsoft), and Teradata can be used to off-load specific analytic applications or even to host an entire data warehouse.

These appliances strive to integrate hardware, operating systems, databases, and analytic applications to provide optimized, powerful, and cost-effective solutions. Some appliances (e.g., Aster Data) employ a columnar database to increase query processing speed. The cloud is now an option for BI and analytics. It can be used for targeted work, such as providing an environment for developing a new analytical application, or for hosting some or all warehouse data.

Some analytics are moving inside the warehouse database. For example, SAS is working with other vendors such as Teradata and Oracle to provide in-database data mining and predictive analytics, which eliminates the need for ETL to a separate server to perform the analytics. Some companies are finding it best to provide a separate BI platform for power users and applications that have heavy processing requirements. This approach, possibly using an appliance, helps take the workload off the main data warehouse accessed by casual users and less computationally intensive applications.

Also focusing on improving processing speed are inmemory BI tools (e.g., QlikView from QlikTech) that store large amounts of data drawn either from source systems or a centralized repository.

As you can see, there are many new things to think about when planning your BI platform.

#7: Analytics Are Used in New Places

Most people don't think of human resources (HR) as being an area ripe for analytics. In fact, in some companies, HR is an important user. This makes sense when you consider the money invested in personnel and the likelihood that when analytics is used elsewhere in a company, it will filter its way into HR. Consider several interesting examples:

Some companies receive a large number of job applications that must be screened. This is especially true of large and "destination" companies (e.g., ESPN, Google).

Predictive analytics can be used to identify the attributes that are best for predicting job performance and thus score applicants for positions. With this approach, every application is considered, but only the "best" applicants are interviewed.

Other companies use analytics for workforce planning and optimization. For example, simulation models can assess the supply and demand for workers with specific skills (e.g., electrical engineers) over time to have the right number at the right time. As conditions change, the models can be rerun to update hiring and retention plans.

#8: Analytics Requires a Diverse Set of Skills

Analysts who perform analytics must have a diverse set of skills: the ability to work with large data sets as well as an understanding of analysis methods, domain knowledge, and communications skills. Few people are strong in all of these areas. An analyst may not possess all of these skills, but someone on the team must. Organizations must also be prepared to develop internal, analyticsoriented training programs to grow necessary skills.

#9: There Is a Shortage of People with Analytics Skills

Years ago, analytics courses and degree programs were more common in business schools than they are today. In addition to a course in statistics, students typically had to take a course in quantitative methods, where topics such as linear programming and simulation were taught. There were departments of Management Science and Decision Sciences that taught and researched analytics.

When I teach a graduate course in BI today, it is only the international students, typically from India, who are familiar with topics such as linear programming.

Beginning in the 1980s, many Management Science and Decision Sciences departments were disbanded. The end result is that most students coming out of business schools today have limited training in analytics.

Because of the need for people, some schools are ramping up to meet the demand (Wixom, et al, 2010). Deloitte's graduate-level certificate program with Indiana University is one. North Carolina State University, St. Joseph's University, and the University of Denver offer graduate degree programs in BI or analytics. A growing number of MBA programs are offering a concentration in BI. Departments such as marketing include analytics in some of their courses.

Overall, however, the pool of business school graduates with significant analytics training is small. Companies need to include graduates in statistics, mathematics, computer science, engineering, actuarial sciences, econometrics, and some of the social sciences in their search for people with advanced analytics skills. Be careful, though. Knowledge of analytics without any business training, experience, and perspective is of limited value to companies. Organizations also need to look internally to find people with a strong knowledge of the business who can work with data, have a demonstrated ability to use Excel, Access, and OLAP tools, and have inquisitive minds.

#10: Advanced Analytics Are Packaged Better

It used to be that only "rocket scientists" (e.g., those with graduate degrees in statistics) could use advanced analytics. This is changing as analytics are being integrated into packaged applications.

For example, consider software that designs marketing campaigns by determining which customers should receive an offer. In the typical package, the software assists in selecting the sample data—people who have and have not responded to a similar past offer. Using this data, the software builds alternative predictive models and helps select the best one. (This is where the predictive analytics is performed.) Next, the best model is used to score the candidates in the target population.

The final decision about who will receive the offer depends on the probability that each person will respond, the profit to be gained from those who do, and the cost of those who don't. The software considers all of these factors and recommends the optimal candidates to include in the campaign. The analyst still needs to know about campaign design, the handling of data, and the use of analytics, but most of the "heavy lifting" is performed by the software. ■

References

Watson, Hugh J. [2010]. "BI-based Organizations," *Business Intelligence Journal*, Volume 15, Number 2, pp. 4–6.

Wixom, Barbara H., Hugh J. Watson, Olivera Marjanovic, and Thilini Ariyachandra [2010]. "Educating the Next-Generation BI Workforce," *Business Intelligence Journal*, Volume 15, Number 3, pp. 26–31.

How Your Favorite Companies and Causes Use BI to Succeed

Coy Yonce

Coy Yonce is the product owner for software solutions from EV Technologies, a services and software partner with SAP.
coy@evtechnologies.com

Business intelligence (BI) promotes efficiency and productivity while providing the information that organizations need to achieve their goals.

BI is used by Apple to track the types of songs, artists, or content that are most popular in iTunes. Blizzard Entertainment can use BI to understand which areas of their massively multiplayer online role-playing game (MMORPG) World of Warcraft are visited most frequently. BI can help healthcare workers determine how changes in health practices can reduce infant mortality rates.

The goal is always the same: leveraging data to create positive, productive actions. When an enterprise adopts business intelligence, it is not necessarily concerned with the choice of specific tools as much as with how the technology, when implemented correctly, can help the organization become more efficient, more profitable, and/or meet other goals.

It All Starts with Well-Defined Goals

Let's look more deeply at the three examples.

In the case of iTunes, Apple's goal might be to optimize revenue from their digital sales and distribution platform. The company offers music, book, game, TV show, and movie downloads, as well as other content types such as podcasts and music streams. More than 800 million customers downloaded more than 35 billion tracks in 2014. To optimize revenue, Apple must understand who is downloading this content, the destination of those downloads (PCs, mobile devices), and which artists are most and least popular, among other things. It must track

a host of metrics that are critical to delivering a positive experience for satisfying and expanding the customer base. These metrics must be tracked against defined goals continuously to determine whether goals are being met as well as where processes can be made more efficient.

Blizzard Entertainment's user base for World of Warcraft grew to 7.5 million subscribers in 2014. This could be the start of a recovery after their decline from 12 million subscribers (at its peak in 2010) to a low of 6.8 million users in Q2 2014. Such a large decline in subscribers is no doubt constantly analyzed at Blizzard. The company needs to understand where the declines are occurring (which regions or countries), whether such declines occur following the introduction of specific product features, and whether declines were triggered by changes to character classes or as a result of competitors' game releases. Blizzard also must distinguish between these causes and the slight increase in subscriber volume that typically occurs after releasing a new expansion pack.

The goal of healthcare workers studying infant mortality rates is to monitor the number of deaths of infants (those no older than 12 months) per 1,000 live births in a particular year. For example, in 2014, the U.S. was rated 169th out of 224 countries; its infant mortality rate was 6.14. In addition to analyzing number of deaths per country per year, healthcare workers want to understand what demographic factors might account for the rate differences—either among countries or within a particular country. Health officials will also want to investigate if higher rates occur in neonatal or postneonatal age groups, are related to specific care routines after a baby is taken home, or if more deaths occur in specific hospitals or birth centers.

In all of these cases, organizations must have the necessary data and people with the right skills to analyze the data. They must ensure there are well-defined organizational goals that will drive their use of BI. They must also ensure that they have the ability to drive change in either how the organization tracks metrics or directly in the process based on performance against defined metrics.

For example, Apple might want to increase the iTunes revenue for music-focused media by 15 percent, but they might seek a 50 percent increase for movie-focused revenue to profit from new content. Blizzard Entertainment may want to target a slight increase in subscriber rates for World of Warcraft while they target a significant initial subscriber rate for their new games HearthStone and Overwatch. The goal for healthcare workers is always lower numbers of infant deaths, but their goals could be made more specific by targeting specific countries, regions, states, or cities.

Each of the goals must have an associated time frame, such as a small increase (say, 5 percent) in revenue per quarter or an increase of 2 percent in clickthroughs on a enterprise's website each month for six consecutive months.

The Next Step is Structure

Once goals are defined, the next step is to outline how to collect and analyze data—both historical and current—in order to make changes to business processes, change marketing programs, or identify other tactics to achieve those goals. In many cases, you'll need to integrate data from multiple sources. In all cases, this means defining measurable metrics to determine whether your organization is on track to reach its goals on time.

For example, if Apple's goal is to increase movie revenue by 50 percent in 2015 in order to cover the cost of adding new content, they will need to include data such as:

- Total number of movies downloaded
- Total amount of revenue generated by movie downloads
- Number of movies downloaded by genre, country, region, state, city, age group, and other demographics
- Amount of revenue generated by movie downloads across these dimensions
- Number of movies downloaded and amount of revenue generated from sales versus rentals

- Amount paid for the rights to sell or rent each movie
- Download and revenue rates for competitors in the streaming media market (e.g., Amazon, Google, Netflix)
- Overall increases in the number of users purchasing or renting digital movies versus discs

This data likely originates from at least five sources: billing data, iTunes content data, supplier data, an independent analysis of competitor data, and an independent analysis of overall market data. With a business intelligence tool that can combine all of these sources—some of which are not owned by Apple—the company can examine current conditions, monitor changes, and respond quickly.

The same is true for Blizzard Entertainment as they seek to secure a large number of new subscribers for Overwatch with only a small increase in the number of subscribers for World of Warcraft during 2015. If they are betting on Overwatch as their strategy for significantly increasing revenue, they will need data to help them analyze and monitor subscriber rates over time. This data could consist of items such as:

- Overall subscriber rates for World of Warcraft
- Overall subscriber rates for HearthStone
- Overall subscriber rates for Overwatch
- Subscriber rates for each of these games broken out by dimensions such as country, state, province, region, age, sex, and other demographic data
- Subscriber rates for competitor games (e.g., League of Legends, Star Wars, Rift, EverQuest, WildStar, etc.)
- Overall interest in the gaming community for MMORPGs
- Overall interest in the gaming community for MMORPGs of specific genres and by specific demographics within that community
- Overall interest from the gaming community in continuing to play games on PCs, consoles, and mobile devices

Again, most of this data will come from disparate sources that must be combined to create an overall picture of the current environment.

For infant mortality rates, the data integration problem is even larger. There are 224 possible countries to monitor for this data, *and* there are a significant number of hospitals, birth centers, and clinics that are set up for delivering babies. In the U.S. alone, there are nearly 6,000 registered hospitals, a figure that does not include other types of birth centers and clinics, nor births that occur at home. Health officials in the U.S. have to consolidate data—sometimes missing details—from all locations where births can take place. Furthermore, they must roll this data up from the hospital level to the city, state, and regional levels. Analysts must break the rates down by all possible demographics of interest, such as the age of the parents, the number of family members nearby, whether smoking takes place in the home, and other factors. These are all potentially critical points of data.

With all of this data and a clearly defined goal, an enterprise must determine the metrics to use in detailing those goals, setting specific action plans, and monitoring program progress. Clearly defined metrics will limit the number of possible data sources and the size of the resulting data set that must be analyzed.

Finally, Tell a Story

After goals are identified, metrics defined, and a structure built to support analysis of the required data, the next step is to gather people to your cause by telling stories with the information you have. This is where the design and reporting tools from your chosen business intelligence platform become important. It is critical to tell an *accurate* story, but you must also tell a *compelling* one.

In the case of iTunes, suppose Apple is combining data from multiple sources to select relevant metrics and monitor progress toward their quarterly and yearly revenue goals from movie downloads. They will track how downloads and revenue are increasing each day, week, month, and quarter. They will slice revenue and download numbers across different dimensions.

For example, Apple would maintain a chart showing which countries are initiating the most and least number of downloads. They would also produce a chart showing which age groups are downloading the most content, whether males or females are downloading the most, and the number of downloads for each genre of movie so they could adjust their business strategy and tactics.

In addition, Apple could use metrics to illustrate download, revenue, and profit trends across each studio from which they source films, showing them which suppliers are providing the most profitable movies. Finally, by incorporating data about competitors and the general market of digital content streaming, Apple could project how the overall market is doing and how their competitors compare to their own revenue and download numbers.

Data tables and charts are good, but they do not tell stories. The tools provide the visuals on top of these different sources of data, but the real power of business intelligence comes when you wrap a narrative around the results.

The power of business intelligence is realized when the people who use BI tools to determine what the data is telling them tell that story to others and help them take action accordingly. That's what really changes the way organizations compete in their industries, become more efficient, and become more productive.

By leveraging a business intelligence platform, organizations such as Apple, Blizzard Entertainment, and worldwide government health officials are better able to monitor their progress toward their defined goals. If they are not progressing as quickly as they would like, they also have the data behind the metrics to help them drill down into the specifics—to determine which countries, age groups, competitors, or others are keeping them from achieving those goals. They can shift their focus with confidence because they have real data to back up their change in strategy. ■

TDWI EDUCATION OPPORTUNITIES 2015

In-Depth Education and Training on Big Data, Business Intelligence, Data Warehousing, Analytics, and More

Upcoming Education Events for 2015

TDWI offers the most comprehensive education and training on data-related topics, including business intelligence, big data, data warehousing, visualization and advanced analytics, and more.

CONFERENCES

TDWI Chicago // Hilton Hotel // May 3–8
TDWI Boston // Sheraton Hotel and Hynes Convention Center // July 26–31
TDWI San Diego // Manchester Grand Hyatt // September 20–25
TDWI Orlando // Royal Pacific Resort // December 6–11

SUMMITS

Solution Summit in San Diego // Loews Coronado Bay // May 31–June 2
Solution Summit in Austin // Omni Barton Creek Resort // September 13–15
Executive Summit in San Diego // Manchester Grand Hyatt // September 21–23
Solution Summit in Scottsdale // Firesky Resort // November 1–3

SEMINARS

Seminar in Washington, D.C. // Big Data // April 20–22
Seminar in London // Agile BI // Radisson Blu Portman // April 20–22
Seminar in Chicago // Advanced Analytics // August 3–6
Seminar in Minneapolis // BI Essentials // August 10–13
Seminar in New York // Dimensional Modeling // October 5–8
Seminar in San Jose // Business Analytics // November 9–12

For registration, academic credit, and scholarship information, visit **tdwi.org/education**

Advancing all things data.

Platinum Sponsors

Sponsors

TDWI Partners

These solution providers have joined TDWI as special Partners and share TDWI's strong commitment to quality and content in education and knowledge transfer for business intelligence, data warehousing, and analytics.