

The Foundations of Analytics: Visualization, Interactivity and Utility

THE TEN PRINCIPLES OF ENTERPRISE ANALYTICS

“Studies show that successful organizations consistently get remarkable performance from ordinary people by applying one simple concept: learning by doing. Providing analytical tools that are comprehensible and adopted across a wide range of skills and training, organizations position themselves to find lasting competitive advantage by leveraging the informational assets they already have.”



About the Author

Neil Raden is the founder of Hired Brains, Inc., <http://www.hiredbrains.com>. Hired Brains provides consulting, systems integration and implementation services in Business Intelligence, data warehousing and performance management for clients worldwide. Hired Brains Research provides consulting, market research, product marketing and advisory services to the Business Intelligence, data warehousing and information integration industry. Based in Santa Barbara, CA, Raden is an active consultant and widely published author and speaker on data warehousing, business intelligence and information technology strategy. He welcomes your comments at nraden@hiredbrains.com.

The Foundations of Analytics: Visualization, Interactivity and Utility

The Ten Principles of Enterprise Analytics

Table of Contents

About the Author	ii
Table of Contents	iii
Executive Summary.....	1
What is “Analytics?”	1
The Problem: The Current State of Affairs.....	2
Data Warehousing/Business Intelligence: Not the Solution.....	2
Good Analytics Software.....	3
The Need for Interactive, Visual Analytics	4
Data, Data Everywhere	4
Reports Are Not Enough – The Need for Speed	4
Sample Topics for Analytics	5
Data Warehouse Methodology versus Analytics	6
Limitations of Data Warehouses and OLAP.....	6
Recalibrating “Analytics”	7
Definition of Analytics.....	7
Operational vs. Analytical	7
OLAP	7
Limitations of OLAP	7
Statistical Tools	8
Interactive Visualization Is the Answer	8
The Ten Principles of Enterprise Analytics.....	9
Conclusion	12

Executive Summary

It is a cliché, but nonetheless true, that the volumes of data pouring into organizations are growing at a frenzied pace. The pressure to find new and creative ways to exploit this stream of raw facts for deeper insight is extreme. It stands to reason that there is a pressing need to expand the population of people who can quickly review, analyze and act upon data without assistance from Information Technology (IT) or reliance on a small corps of “power users,” but the common perception is that “analytics” is reserved for a small group of highly trained and quantitatively oriented professionals. However, studies show that successful organizations consistently get remarkable performance from ordinary people¹ by applying one simple concept: learning by doing. Providing analytical tools that are comprehensible and adopted across a wide range of skills and training, organizations position themselves to find lasting competitive advantage by leveraging the informational assets they already have.

The focus of this paper is to define the term “analytics,” contrast it to Business Intelligence and to provide a framework to understand the requirements of analytics, which are defined in the final section, “The Ten Principles for Enterprise Analytics.” Analytics is suitable for large numbers of people, not just specialists, and there is a pressing need for broader and more comprehensive proliferation of true analytical tools.

1

The proper term for interacting with information at the speed of business, analyzing, and discovering and following through with the appropriate action, is “analytics.”

What is “Analytics?”

The proper term for interacting with information at the speed of business, analyzing, and discovering and following through with the appropriate action, is “analytics.” The term “analytics” is badly misapplied by the Business Intelligence community, which uses it widely, because their offerings fail to meet too many critical criteria to be effective as analytics. In particular, analytics requires: [visualization + interactivity + utility](#).

[Visualization](#) goes beyond simple charting. Useful visualization adds to the analysis, it doesn't just display the results of it. When visualization is [interactive](#), meaning a person can conduct analysis by interacting with the visualization itself, not the program behind it, an immense amount of business [utility](#) is unlocked. This interaction has to happen on both a mechanical level (linked visualizations allowing the user to simultaneously explore many variables) as well as a functional level (iterating through the data to find other products that resemble a sell-through curve or customers that exhibit similar behavior to define a “segment” for further analysis).

Analytics is wrongly perceived to be too technical for most people to master. The key to proliferating the use of these capabilities is to encourage people to discover the benefits through not only training, but through practice. Each person has his or her own particular questions and theories, which are generally not addressed through reporting and canned analysis such as dashboards. Unlocking these observations for examination and discussion is essential. This typically starts by

¹ Pfeffer, Jeffrey and Sutton, Robert I., *The Knowing-Doing Gap* (Boston, Harvard Business School Press, 2000)

incubating the process and allowing it to spread organically as the utility of it becomes widely known. Sharing is key, and it has to include the ability to not only publish the results of the last step, but to provide encapsulated and interactive “guides” for others to replay, examine and experiment with the thread of analysis. Reports and dashboards lack this quality; they do not allow people to learn by doing.

The Problem: The Current State of Affairs

Despite using the word “analytics” freely, most “solutions” in the market today fall far short of providing support for real analytics. Business Intelligence tools have barely made a dent², hovering between five and fifteen percent adoption rates, depending on which study you believe. Data mining and statistical software exist at the margins, adding value but not widely used. This tends to push more and more of the effort to the ubiquitous tool at hand, Excel spreadsheets, which exacerbates a growing problem called Shadow IT³, distinguished by poor data quality, misappropriation of critical skills, erosion of data warehousing ROI and soaring maintenance costs. In addition, inefficient processes for handling data and analysis lead to a knowing/doing gap⁴ – subject matter experts spend too much time manipulating data, make too many manual steps in the analysis process and spend too little time implementing decisions. Current Business Intelligence solutions accommodate this gap by providing information without visualization, interactivity or the ability to share findings and close the loop.

Data Warehousing/Business Intelligence: Not the Solution

Data warehousing, that is, extracting, transforming, integrating, loading and storing data to provide fodder for digital dashboards and other reporting efforts does little more than rearrange the problem. Until the information can be examined, understood and acted upon, it is, truly, just a “warehousing” operation. The solution to the problem is to provide more adaptive tools that are accepted and used by a wider audience across the business processes that most impact differentiation and competitiveness. The BI industry uses the term “analytics” casually, but real functionality for analytics is found mostly in the specialized tools and industry verticals, not the most popular BI tools.

Analytics has not been prominent in data warehousing and BI, despite its obvious benefits, for a number of reasons, including:

- Data warehousing methodologies are data centric, not analytical
- The deployment of BI has historically been driven by replacement of reporting systems
- The perception is that visualization is bound to statistics and too difficult for people to use

² http://www.intelligententerprise.com/info_centers/data_warehousing/showArticle.jhtml?articleID=19502113

³ Shadow IT is defined loosely as IT work performed without the knowledge or control of the formal IT organization. The bulk of this effort is personal productivity tools like spreadsheets and databases.

⁴ Pfeffer, The Knowing-Doing Gap

Increasing the return on information invested across the population of decision-makers in your organization has to follow a formula for analytics that is not supported by today's data warehousing push to BI reports – visualization + interactivity + utility.

Good Analytics Software

The solution to the problem is to provide analytical software that responds to, provides an intuitively effective interface, is powerful enough to meet the requirements, guides less skilled people and is accepted by a wide range of workers⁵. Good analytics software should adjust itself to the level of skill and familiarity of each person using it. The simple test should be, if a question can be expressed in words or writing, it should be just as straightforward for a good analytics software tool to pose it as well. In contrast, most BI tools require the use of, and understanding of, terms like table, cube, dimension, attribute, join, key, view and aggregate. This is why BI software is mostly shelfware with little to no impact on organizations – the pretty interfaces give way to these difficult concepts as soon as the demo is over. Though people who succeed in using these tools may come to find these concepts easy in everyday usage, the much larger majority of people who try the tools and abandon them do not.

Increasing the return on information invested across the population of decision-makers in your organization has to follow a formula for analytics that is not supported by today's data warehousing push to BI reports – visualization + interactivity + utility:

1. Visualization – The most effective and unambiguous way to understand data.
2. Interactivity – Knowledge of the business user – BI software can't keep up with the natural iterative nature of the mind. The typical "user" is not a user at all; he/she is usually smart, educated and knowledgeable. Software products should serve them, not vice-versa.
3. Utility – The analytical process has to be organized beginning with the business problem and business user in mind while remaining open to new and changing data-static data models feeding information to static user displays won't work. Usage patterns in analytical software follow a "life cycle" that is characterized by different levels of knowing and doing over time for all participants. Unless an analytical software solution can accommodate all of these user variations, no utility will be gained. The ability to produce "guides" or "guided analysis" is crucial in helping people to learn and progress through the various stages from reviewing others' work to creating their own.

⁵ Peter Drucker said, "No institution can possibly survive if it needs geniuses or supermen to manage it. It must be organized in such a way as to be able to get along ... composed of average human beings."

The Need for Interactive, Visual Analytics

Two forces are at work that drive the need for analytics: data volumes are increasing geometrically without corresponding increases in staff or tool efficiency, and the pressure to rapidly respond to competitive and regulatory factors is forcing organizations to improve their ability to manage, analyze and act on the information at their disposal.

Data, Data Everywhere...

Software is more complex than it used to be and generates and stores vastly more information than prior generations of application software. Commercial concerns capture detailed data from their supply chains, fulfillment chains and third-party syndicated data about customers and products. There are also emergent technologies that generate more data. When chemical and paper plants implemented computerized process control systems, capturing and analyzing the telemetry – like data added a whole new set of possibilities for fine-tuning. When laser scanners became widespread in grocery stores, the volume of point-of-sale data skyrocketed. More recently, e-metrics, the analysis of clickstream data from websites, created the need to mine vast amounts of data in order to understand the behavior of customers, perform dynamic pricing and customization, and a host of other data-intensive functions. Today, RFID (radio frequency identification) devices are entering the business environment and will eventually dwarf the amount of data we manage now. Pharmaceuticals have automated and industrialized laboratory research introducing technology for high throughput experimentation, dramatically widening the search for the next blockbuster drug.

In addition to just the sheer volume of data that has to be digested, competitive and regulatory pressures to deal with the problem are building. In the past, it was enough for organizations to prepare external reports on a quarterly basis. Today, daily operations are subject to examination and scrutiny. Product safety and quality concerns are top of mind. Decisions about switching channels or suppliers are made on the spot. Monitoring operations for minute improvements and understanding the behaviors of customers, suppliers and competitors are crucial. With this (potential) wealth of information available, those that delay their efforts to exploit it will find themselves slamming on the brakes rather than the accelerator.

Reports Are Not Enough – The Need for Speed

All of this information is useless if it can't be acted upon. With the volume of information accumulating at the speed of business, our existing tools and approaches can no longer keep pace. When computing and storage hardware was scarce and expensive, best practices for the capture and reporting of data were formed around this notion of scarcity. Expectations were kept fairly low. However, with the cost of computing, networking and storage falling every year, and capacities increasing, it is possible and even logical to think more expansively, to envision acting on large amounts of data that is only hours old. As a result, many approaches, methodologies and best practices, and the software paradigms that support them, are becoming obsolete. Everyone is looking at the same data and companies are



Data warehousing methodologies today are still designed around the concept of data delivery or data presentation, terms that resolve, in practice, to reporting.

in search of “next” best practices to recreate competitive differentiation.

Though data warehouses can be quite large, the BI tools that are deployed to leverage them typically address a small subset of aggregated data. Data mining and statistical tools extract slices of data and operate on it in isolation. The richness and depth of the data is lost unless it can be analyzed in its totality, in a relatively short period of time.

Data warehousing methodologies today are still designed around the concept of data delivery or data presentation, terms that resolve, in practice, to reporting. Newer innovations, like Enterprise Information Integration (EII), bypass the data warehouse structure, but still present data to a more graphical kind of report, the dashboard.

Dashboards are very popular because of their real-time nature and at least offer limited capabilities to business users to customize their display, but their acceptance is mostly a consequence of the lack of comfort with BI, which is limited to viewing and creating reports. Perfecting modes of reporting is useful to a point, but it still does not solve the problem of linking analysis with action and getting more people involved in the process.

A report, even a useful, attractive one, even a report that can be animated through drilldown or template rotation or filtering, is still a report. And information, no matter how well integrated, cleansed and artfully modeled, is inadequate if the latency in the process is greater than the business need. Today, business people have to deal with problems and opportunities, large and small, in real time. The tools they use to assist them have to operate in real time, interactively, providing continuous discovery, what-if capabilities, collaboration and the ability to fully explore all

relevant attributes of the data to take informed action with confidence. Real business issues that require real analytical tools abound but are not being sufficiently addressed with current practices, for example:

SAMPLE TOPICS FOR ANALYTICS

Routing and Rescheduling

Customer Segmentation

Promotional Lift/Cannibalization

Homeland Security and Intelligence

Regulatory Compliance

Labor Costs

Production Yields

Defect Analysis

Capacity Planning and Optimization

Risk Management/Portfolio Analysis

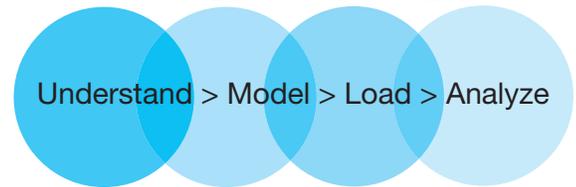
Sales Performance

Pipeline Analysis

Most of these subjects require access to vast amounts of detailed data, often in real-time or with very low latency. The relationships between the elements of each of these models transcend simple parent-child roll-ups, and being able to visualize and explore different scenarios is critical. Capacity planning is a good example. Variables from the supply side, the demand side and the effects of peaks and valleys combine to form a picture that cannot be imagined from static reports. Promotional lift is a process of almost endless “what-if” analysis and the output of each set of assumptions has to be analyzed and compared. Routing and rescheduling models can employ heuristics, brute force or even predictive models like neural nets, but each one generates vast amounts of outcomes that have to be sifted and compared. Only a powerful analytics tool that supports interactive visualization can handle problems like these.

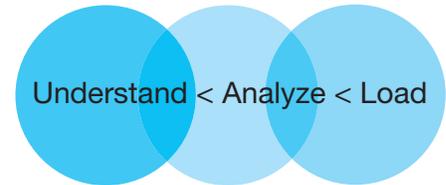
In addition to the power needed to solve these problems, an analytics tool

DATA WAREHOUSE PROCESS



has to be easy enough to use so that less sophisticated or quantitatively oriented people can use it to review the situations and the outcomes in a collaborative fashion. A facility for encapsulating an analytical investigation shares two purposes. First, the entire process that leads to a particular finding can be recorded and saved so that others can follow the whole thread, not just the final report. The analyst can record and save the model, data and manipulations, and secure them in a repository for others to view, comment on, change or refine. This can serve to elevate analytics from a singular activity to group effort and it offers an immense opportunity for continuous improvement and sharing of best-known practices or best “next” methods. Second, more skilled people can prepare these “guided analyses” for others. The proportion of people who can learn to use and interact with a guide rather than start from scratch is large.

ANALYTICS PROCESS



Data Warehouse Methodology versus Analytics

Peter Drucker once said that computers have done a great deal to damage businesses by making managers even more inwardly-focused than they were. The notion that a company’s data has unmined value may be true, but it obscures that fact that a great deal of useful information exists beyond the company’s walls, as Drucker said. BI (and the data warehouses that underlie it) is designed around static data models and cannot accommodate a constant flow of new data sources. There is a pressing need to be able to easily flow new data (and new types of data – i.e., adding demographic data to a segmentation analysis) into a workspace and examine it to understand its content. This reverses the data warehousing concept, (see figure). Classic data warehousing attempts to understand the existing data flows first, creates a static data model, then populates and refreshes the structures repeatedly. BI tools continue to operate against these relatively static structures. In a useful analytical environment, subject matter experts, domain experts, people who understand the business or the operations, business users or business analytic users, evaluate new information in real-time, visually, and gain an understanding of it as they work. This is a very different paradigm. Business opportunities and threats happen in real time – business analytics has to maintain at least the same pace.

Limitations of Data Warehouses and OLAP

Dimensional models and OLAP operate on an atomic basis. In other words, the relationship between things is parent-child, and attributes exist only along a dimension (color, size, etc. for an item). It is easy to examine groupings and filtering of things, but it takes a lot of effort to go beyond that because neither SQL nor OLAP cubes are designed for that. They are designed to aggregate first, and then view the result. But in business (and in everything else) there are emergent properties that are pretty shifty and escape notice, but when examined as a group, exhibit very interesting behavior and potentially great competitive value. The trick is defining the group first, not by its hierarchical relationships, but by, in the case of people and products, demographic and behavioral characteristics. Market basket analysis, what people buy with other things, is a typical example. An interactive, visual analytic capability is a good way to discover these phenomena. Other examples are purchasing behaviors, promotional effectiveness, fraud and criminal networks, insider trading, money laundering and, hopefully, many things we have not even conceived yet.

Recalibrating “Analytics”

Before defining the requirements for analytics in “The Ten Rules for Analytics,” it helps to define the term more completely and compare it to other approaches that are incorrectly labeled analytical, such as OLAP, packaged analytical applications, statistical tools and data mining.

Definition of Analytics

The formal definition of analysis is, roughly, the study of something by looking at its parts and examining the relationships between them. Analytics is a more active term, meaning the method or process of analysis. However, when the terms analysis and analytics are applied to Information Technology, the meanings are often quite different. Before specifying a set of rules or characteristics of analytics, some clarity is needed.

Operational vs. Analytical

The world of IT is divided today between systems that support the actual operation of a business or enterprise, operational processing, and those that track, report and analyze it after-the-fact, analytical processing. The division is slightly arbitrary as it is not always clear what is operational and what is analytical, but applications supporting each are called, respectively, operational applications and analytical applications.

Analytical applications (and this includes analytical processing, which is described below) are usually capable of breaking things down to their parts, but the problem is that the decomposition alone is not what makes something analytical, it's the analytical process that ensues. That's where the term “analytic” as applied to today's BI applications is misused. Software, such as OLAP, is often positioned as analytical, but lacks many of the necessary features to support real analytical work.

OLAP

E.F. Codd first defined the term OLAP (On-line Analytical Processing) to describe what we now know as multidimensional models and the navigation around them. Though OLAP tools provide the ability to explode data into greater detail (drill down) or compare different slices of data (drill across), it is completely possible for someone to perform these manipulations without doing any analysis at all. The same is true for analytical applications, which describes a wide range of reporting, OLAP and dashboard applications designed for specific functions or specific industries or both.

Limitations of OLAP

OLAP tools deal with data depth by employing a combination of aggregation and dimensional navigation. Because there is too much data to display in a grid, it is compressed and rendered in aggregated form, from which an analyst may decompose parts of it via pre-defined hierarchical paths. Most OLAP tools are not capable of constraining queries or formulating new measures using the qualities or attributes of these hierarchical units. In other words, in most OLAP applications, the relationships that can be analyzed are parent-child, wasting the information embedded in the attributes of parents and children. To introduce more attributes expands the dimensional complexity of the model and produces an undesirable result called sparsity, which limits the ability of the models to scale to the depth (amount of

7

The formal definition of analysis is, roughly, the study of something by looking at its parts and examining the relationships between them. Analytics is a more active term, meaning the method or process of analysis.

data) and breadth (richness of the data) that is needed for true analytics. There are exceptions to this, notably OLAP tools that are based in a relational model, but there are drawbacks there as well, especially in performance and ease of use.

One barrier that cannot be breached with reports is the limited amount of data, displayed as characters, which can be scanned and comprehended on a single display. It is clear that a report cannot scale to millions of pieces of information. Visualization, however, can not only place millions of pieces of data on a page, it can expose relationships that can be discovered through reports only with great effort, if at all.

Similarly, grid displays of numbers are only useful to a point. It is fairly easy to spot a few outliers in numbers, such as very large, very small, negative or inappropriate values in a report, but when numbers are aggregated, these anomalies are often masked. Drilling down does not reveal them because the point of departure is usually something that appears interesting at the aggregated level, not something that appears normal. This is especially true where a large number of values are rolled up – the aggregated amount may be within reasonable boundaries, but it can mask extreme variations that cancel each other out. Because it is impossible to visually scan thousands or even millions of values, the only way to spot these anomalies is with statistics and visualization. OLAP tools provide little to no support for this.

To a large extent, today's analytical applications and OLAP tools are deployed for reporting purposes. The interactive, speed-of-thought surfing through data to analyze what is behind the numbers is the domain of a very

small number of “power users using complex tools.” In other words, analytical applications and on-line analytical processing tools are, for the most part, not analytical at all. For analytics to occur, a person has to consider the information in front of them, follow a thread of reasoning, iterate through possible conclusions, share their findings and, most importantly of all, act with confidence on their results.

Statistical Tools

Statistical analysis can generate useful information easily, but very few people are comfortable enough with the concepts to place much trust in them, or themselves. In fact, most people cannot explain the difference between a mean, a median and a sample mean. The misapplication of statistics is so likely, that only those with a solid background in the discipline use them routinely. This does not diminish the importance of statistics; they're just not for everyone. It's not a needed feature for the vast majority of applications. Deployed on their own, they cannot serve our wide audience.

Interactive, Visual Analytics Is the Answer

On the other hand, interactive, visual analytics offers advantages to people of all levels of skill and training. With interactive visualization, it is possible to scan and analyze great volumes of data and to navigate through the data. This is a key discriminating factor. One doesn't navigate the data in visualizations; one visually navigates the relationships between the data and draws inferences instantly. This matches our definition of analytics as an active process, not just a set of capabilities.

In the next section, we propose ten “principles” for enterprise analytics.

The Ten Principles of Enterprise Analytics

Given the strong evidence of “buyers’ remorse” for BI revealed in surveys⁶ and the requirements for analytics described in this paper, the following ten “rules” lay out what is needed for a complete analytics platform that can serve the needs of a high percentage of knowledge workers.

1. Visual + interactive

Visualization of data is the quickest way to understand information, and it’s easy to use. Analytics that are visual have better usability and are easier to use than BI. Making analytics visual makes it ten times more powerful. To serve a wide audience of business users, data has to be rendered unambiguously. It must be visual. Knowledge workers must be able to navigate through the data and all its attributes. It must be interactive.

2. Zero code

Tools designed for software engineers, database analysts, programmers, network engineers and even specialized functions like chemical engineers or computer animators require the use of programming languages. The art of programming requires a great deal of skill and experience to be effective, and should be managed for standards, reuse, quality and maintainability. Tools for analytics, on the other hand, are designed for business users who mostly were not hired for their programming skills. To be truly useful, analytical software should provide the full range of capabilities without the need for programming, scripting or any other kind of code development, with no exceptions.

3. Actually Easy to Use

Analytical software must be easy to use, but what exactly does that mean? A Google search on “ease of use” will return 10,600,000 hits. The Google interface is code-free, but needing to sift through millions of documents is far from easy. Learning the nuance in using Google takes time and effort to get it right. The GUI is only part of the answer. If you use a GUI and still have to decide what order to place the parameters to a function or whether to do an inner or an outer join, it’s not easy to use. The problem is that GUIs are sometimes a thousand times more difficult to use than the old command line prompts. It isn’t the code that’s the problem; it’s the conceptual model. When you’re not sure what the effect is when you click something, or if you don’t understand the effect of loosening a constraint and sliding a slide bar, it’s hard to use. To make it truly easy to use, it has to deliver benefits of value and it has to be understandable. The connection between the interface and the underlying conceptual model has to be obvious.

9

Fast is always a relative term, but you know it when you see it. Interactive visualization and data analysis require near-instantaneous response time to keep up with the analyst.

⁶ Gartner reports BI penetration of a modest 30% in firms, a number vastly overstated, according to Forrester, who reports 10% and Hired Brains, less than 5%

Because the success of analytics depends on its adoption and use by a wide cross-section of the user population, an analytics tool must be useful for people with extreme variations in skill, training and application without the intervention of IT. It should support investigators and authors, administrators and casual viewers. The software must also foster a learning and collaborative environment. In the same way a reporting tool provides a means to distribute reports, an analytics tool should have the means to store and distribute analyses, a thread of reasoning that can be replayed and evaluated by others.

4. Fast

Fast is always a relative term, but you know it when you see it. Interactive visualization and data analysis require near-instantaneous response time to keep up with the analyst. But speed is not limited to response time. Cycle time is also critical. It is not just the speed of the first question but of the many others to follow, all contributing to the total time to understanding. While data warehouses employ data models and refresh the static model repeatedly, a great deal of data that is the subject of analytics is not recurring, it is one-off and/or sourced externally. Analytics has to be fast enough to ingest this kind of information and make it available to the analyst in the least possible time.

5. Disposable/Persistent

Analytic opportunities are not uniform. Some are repeated consistently, some happen only once and others have a longer but still limited lifespan. For the latter two, it must be possible to conceive of and execute analysis quickly, and to dispose of the analysis when it is no longer needed. For those scenarios with a medium to long lifespan, a smooth mechanism must function so that it can persist and be reused as needed. What is missing from most BI tools today is a systematic discovery process, a way to continuously discover things without the need for a development cycle. Data has to be organized beginning with the business problem while remaining open to new and changing data – static data models feeding information to static user displays won't work.

6. Collaborative

No matter how useful or relevant a set of investigations or analyses is, unless they can be smoothly conveyed to others and discussed, modified and perfected, the effort has only momentary value. OLAP allowed numerate analysts to surf through mountains of data and draw useful conclusions, but it provided no facility to share the insight gained, only the final report. Analytics is an active process that continues even after the analyst completes a thread of analysis. Making not only the final result, but the steps to get there, available to others is the collaborative part of analytics that is missing in BI.

Most importantly of all, for analytics to be useful, people have to be able to express themselves – the selection and managing of data, the design of the model (relationships, calculations, hierarchies, etc.), and the appearance of the visualizations.

7. Conceptually Sound

Iterative manipulations by a user have to hold together because the underlying model of the software is based on a sound set of relationships, not just one-at-a-time graphics. Applying filters, constraints or other rules should perform accurately and consistently so the analyst can be confident in the answer and there is no need to double-check every result. Because most analytical inquiry involves a series of steps, this conceptual soundness must apply across all the steps in the current analysis. Conceptual soundness contributes to decision confidence. Likewise, there should be no need to understand (relational) data models. Analytical tools should present results to business users, not data models. All references to the source or storage modes of the data should be abstracted. Elements of the model should match the jargon of the business user, not the software industry.

8. Depth

All of the attributes of the data must be exposed, and not limited to small data sets or only a few dimensions. A spreadsheet model can accommodate about 16 million cells, with a hard limit of 64,000 rows and 256 columns. For many applications, this is far too small to be useful. Any tool can address huge amounts of data by utilizing a separate database server, but these solutions have proven to be too slow. CPU and memory resources are plentiful. Models that exploit this abundance will be most effective for interactive visualization.

9. Good Software Citizen

Like any other enterprise software, analytics software has to make provisions for all of the requirements to be a good participant in enterprise architecture such as honoring and using security, adhering to standards and being driven by a sensible architecture that minimizes risk due to maintenance, changes or error.

10. Expressive

Most importantly of all, for analytics to be useful, people have to be able to express themselves – the selection and managing of data, the design of the model (relationships, calculations, hierarchies, etc.), and the appearance of the visualizations. Moving from idea to visualization to action has to be simple and straightforward and the range of ideas and approaches cannot be unreasonably constrained by the technology.

Conclusion

People in business, education, research, government and all other kinds of organizations are being pressed to capitalize on the swarm of data that flows through their networks. So far, the technology industry has done a good job handling scalability and reliability in the capturing and storing of that information, but making sense of it all by more than a handful of experts is still a green field adventure. Data warehousing and Business Intelligence provide excellent tools and practices for refining, formatting and reporting that information, but the need to understand it in order to act is still largely unmet.

Analytics is a vital activity for a broad range of people, but the term has become associated with Business Intelligence, on the one hand, and statistics and data mining on the other. In point of fact, it is neither. Anyone in an organization who views quantifiable information, whether in reports, spreadsheets, grids, bands, graphical, dashboards or greenbar, has a need to understand what the underlying numbers mean, and that implies a discovery process. Reports may satisfy a need to standardize the information, but fail to address the fact that everyone has their own questions. Well-made business analytics powered by interactive visualization facilitate the discovery process in a painless way and open up the membership to the “analysts” club up and down the organization. Every day, people are exposed to active visualization, from their graphic equalizers to the animated weather and traffic maps in the morning. It is an accessible metaphor that has not been widely deployed because of the lingering mindset that it is too computer-intensive, too difficult to use, requires a statistical background and the temperament of an astrophysicist, none of which is true.

However, no matter how useful it is and uncomplicated to use, certain people will never actually take the driver's seat. For them, guided analysis is the answer, allowing them to ride in the passenger seat, to stretch the metaphor, and follow a thread of analysis performed by someone else.

The Ten Principles of Enterprise Analytics are a guide to understanding, selecting and deploying analytical applications that can have a powerful impact on the use and value of information in your organization. Use The Ten Principles of Enterprise Analytics to help you to understand and determine whether while you are looking at BI, are you really looking for analytics?



Spotfire® DecisionSite™

Spotfire, Inc.
212 Elm Street
Somerville, MA 02144 U.S.A.
Telephone +1.617.702.1600
Fax +1.617.702.1700
Toll-Free +1.800.245.4211

Spotfire AB
(European Headquarters)
Första Långgatan 26
SE-413 28 Göteborg, Sweden
Telephone +46.31.704.1500
Fax +46.31.704.1501

Spotfire Japan KK
(Japanese Headquarters)
Kinokuniya Bldg. 7F, 13-5,
Hatchobon 4-chrome
Chuo-Ku, Tokyo 104-0032 Japan
Telephone +81.3.5540.7321
Fax +81.3.3552.3166

www.spotfire.com