

# An Analytics Manifesto



By Neil Raden  
Hired Brains Research  
January 2006

# TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	1
MANIFESTO: NEW RULES .....	3
ANALYTICS IN PERSPECTIVE.....	3
THE ANALYTICS GAP .....	4
CLOSING THE GAP .....	5
OPERATIONAL VERSUS ANALYTICAL.....	5
LOW-LATENCY ANALYTICS.....	6
INVERTING THE PYRAMID.....	7
FACTORS AFFECTING ANALYTICS.....	9
SEMANTICS.....	10
PRINCIPLES .....	12
ANALYTICAL FUNCTIONALITY .....	13
APPLICATIONS OF SOA ANALYTICS .....	13
Types of Analytics .....	14
PREDICTIVE MODELS.....	14
OPERATIONAL ANALYTICS.....	14
INTERACTIVE ANALYTICS.....	15
VALIDATION FROM REAL COMPANIES.....	15
Boeing .....	15
Procter & Gamble.....	16
CONCLUSION.....	16
ABOUT THE AUTHOR.....	18
ENDNOTES.....	19

## EXECUTIVE SUMMARY

The size of the Data Warehousing/Business Intelligence industry pales in comparison to the operational software business. Total spending for software (globally) is expected to reach over \$100 billion in 2005<sup>1</sup> according to IDC, but BI's share of that total is less than 5%. The stated purpose of enterprise systems is to improve organizations' effectiveness through better utilization of information, especially for decision-making. Yet somehow, analytics have taken a back seat role in corporate IT for over 40 years. The forces of technology and business are about to change that.

Technology factors drove analysis and operations apart almost thirty years ago. Thanks to a different set of technology innovations available today, this gap is about to close. Software to support operations can no longer fulfill its requirements without embedded or cooperative analysis to not only measure business processes, but to drive them. At the other end of the strategy/operations continuum, analytics are performing a key role in not just the monitoring of strategy, but in the formation of it as well. Enterprise systems can no longer rely on analytics and analysts to salvage data after the fact, gain insight and communicate it back to operations. Analytics is a crucial part of the enterprise, from the capturing of strategy as metrics and Key Performance Indicators (KPI) to driving the enterprise in real-time.

After two decades of focus on business process, organizations around the world are turning their attention to solutions capable of enabling and enhancing agility, inclusiveness and alignment. A combination of factors, including new blueprints for enterprise architecture, the relentless (but welcome) progression of Moore's Law, the emergence of commercially viable semantic integration technologies and, not least of all, the overall maturity and success with process orientation are catalyzing this shift. For almost 20 years, the refinement of internal business processes and the search for greater levels of efficiency through process orientation, including enterprise software, has been a focus of business management. While managing business processes is not a solved problem, some current research points out that the development of process measurement standards will hasten the outsourcing of "commodity" processes<sup>2</sup>, freeing up resources to concentrate on higher value-added implementation. In addition, the rapid acceptance of open standards like Web Services, which makes Service-Oriented Architecture (SOA) possible, provides an appealing opportunity for re-architecting business process software as a set of cooperative services. The promise is increased alignment through vastly improved agility, reduced maintenance costs, which account for up to 75% of IT budgets overall<sup>3</sup>, and a reduction in integration expenses, which make up a large portion of IT development budgets.

The element missing from most road maps, however, is business analytics. This absence is an understandable consequence of the uneasy relationship between analytics and enterprise systems that has stretched over four decades. Operational and analytical disciplines exist in separate universes, and analytics has always been a secondary concern to corporate IT. Agile organizations, however, need analytical insight embedded into their operational processes to "close the loop" between analysis and action. The lag between actions taken, results tracked and conclusions drawn has been getting increasingly smaller, but the last step - connecting analysis to operations in a continuous cycle - is restricted by existing tools, methods and approaches derived from years of isolation of analytics from the operational flow. Organizations also need the insight and control offered by a branch of business analytics, Corporate Performance Management (CPM), which includes a range of disciplines and applications such as Budgeting, Planning, Forecasting, Regulatory Reporting and Scorecarding.

Many current analytical tools limit and simplify the models and scale of data that can be investigated because of their client-based architecture. Even in server-based BI, many of the tools scale poorly and do not leverage the data resources already available. In addition, organizations typically employ more than one analytical tool and skills are spread too thinly to have much impact. Collaboration across tools is difficult and duplication and redundancy and

error are the results. In an environment like SOA where analytical tools will be expected to cooperate with a standards-based architecture, these shortcomings will either have to be resolved or the tools will be replaced. The loosely-coupled<sup>4</sup> nature of SOA will gradually expose these shortcomings and favor those vendors that can provide optimal functionality for the new architecture. Current tools for analytics are hampered by a lack of common semantics, too many proprietary engines, layers of stand-alone functionality and poor data federation capabilities.

A combination of forces is working to tip the scales in favor of a new generation of analytics:

- Moore's Law: Analytical schemes today are still managing from scarcity. The continued rapid increase in capacity and decrease in relative cost of hardware and bandwidth allows for the faster and broader analysis of data.
- SOA: Open standards foster an environment where analytics can cooperate with operational processes without disrupting or interfering with the smooth operation of these systems and processes
- Maturity: Most organizations, and some of the people in them, have some experience with analytics now and are open to improvement.
- Business Process Commoditization: Visionaries like Davenport are getting the word out that a lot of the work in process optimization has been done.
- On Demand: Opening business up with the Internet creates an entirely new timetable for conducting business – on demand.
- Semantics: The rapidly expanding field of Semantic Technology will power analytics to new levels, fueled by open standards for ontologies such as RDF and OWL.
- Hegemony of a few vendors and influence of analysts: These can act as negative forces, but once enough momentum is gathered, they tip abruptly.
- The status quo is simply not good enough. BI has not reached as many people as it should, and has not been as effective as it could be.
- Vision of companies like SAP that understand that analytics is an integral part of almost every application.

The fundamental nature of data warehousing and Business Intelligence today is the bulk gathering of data and generic presentation of it to a wide audience, predominantly for reporting and extraction to spreadsheets. Benchmarks, best practices and vendor architectures are geared to this paradigm. In the rapidly emerging world of services-oriented architectures, fluid and agile business processes and on-demand responses, the old paradigms simply do not suffice. Analytics must be a central part of most business processes and the tools, techniques, products and best practices are just now being formulated.

In this paper, a new set of principles for analytics is described that not only powers this new architecture; it enables vastly smarter business processes. Analytics is finally poised to take its place at the table with the rest of the enterprise software portfolio.

## MANIFESTO: NEW RULES

The entire field of analytics has been, to date, driven by the marketing of BI vendors and their influential customers. As analytics emerges from its various niches and proves its centrality to operational and strategic roles, a clear set of rules is crucial:

- To achieve its mission, the use of analytics at all levels of skill and function, has to be pervasive in organizations.
- Analytics, even complicated analytics, does not have to be difficult and should not be the preserve of a cadre of statistical experts.
- Analytics is useful at achieving ROI at the operating level. At the same time, it plays an important role in setting and managing strategy with tools like CPM. Going forward, these roles must be aligned, consistent and seamless.
- Analytics cannot just inform; it has to be active, stitched into the fabric of work
- Embedded analytics needs to be composed of standard services with consistent function and metadata across the enterprise.
- Analysis in isolation is not effective; it must close the loop within business action.
- Commodity business processes will be outsourced, increasing the need for reliable measurement.
- The requirements for real-time will grow as the availability of reliable real-time tools increases.
- The effects of Moore's Law will gradually move thinking about analytics from "managing from scarcity" to capitalizing on its capabilities.
- SOA will move analytics from a departmental to an enterprise pursuit.
- Semantic technology promises to be the fuel behind increased use, utility and value in analytics. The application of industry standards to the building, discovery and merging of ontologies will vastly improve everyone's ability to inform themselves of the meanings of complex terms and their relationships
- Metadata must rise above proprietary standards; abstraction layers must be the preferred path for access to data.
- Analytical tools will eventually learn how people work, not vice-versa. In the meantime, it will still require customization and configuring to produce analytical tools that are relevant and understandable to a wide audience.
- In a SOA world, BI tools that bundle query, interface, output, metadata and calculation will have to gradually un-bundle their services.
- What makes BI difficult for people is the lack of understanding of the data, models and tools. Analytics has to be relevant to work that people do and make their work easier, not more difficult.

## ANALYTICS IN PERSPECTIVE

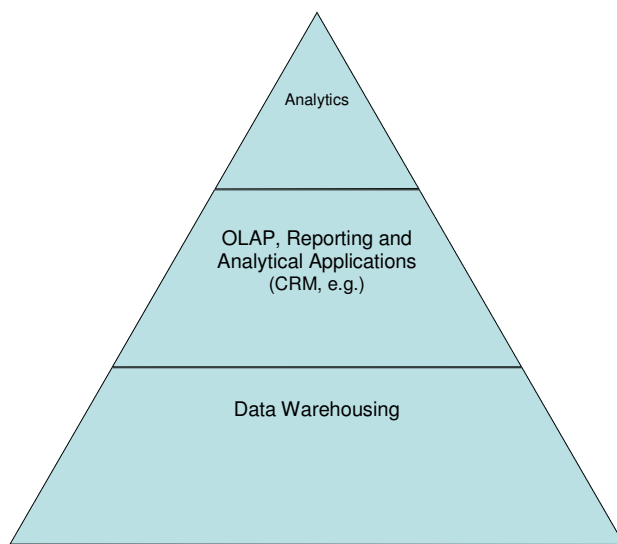
A new generation of analytical tools is needed that can be seamlessly embedded in business processes. Ideally, they must:

- Extend from viewing of results to interactive exploration of data and models;
- Be as tightly connected to the strategy of the organization as it is to the minute operations;
- Foster collaboration and group decision-making;
- Provide the ability to share knowledge models and techniques across the enterprise and beyond, through the easy assembly of components by business stakeholders.

Today, fewer than 20% of knowledge workers in organizations avail themselves of analytics.

Information Technology (IT) is undergoing a radical renewal at its most fundamental levels as a result of widely accepted standards made possible by the success of the Internet. Until recently, enterprise architecture was an inwardly-focused discipline, but the opening of business to the outside world caused a complete rethinking of the way business processes are supported by technology. Previously, applications were monoliths that had to be connected to other applications by custom interfaces or batch transfers. With the new architectures, reusable, loosely coupled services can be orchestrated and re-orchestrated as business needs demand. The longer-term results of this change will be greater agility, the ability to reach a large audience of partners efficiently and include them directly in your business process and to vastly reduce integration expense, which is a significant part of ongoing IT costs.

Most of the discussion about SOA centers on operational systems that run the day-to-day business processes of organizations. But for SOA to work, it is essential that business analytics are stitched into the fabric. Labeled over the years as decision support, analytics, Business



Intelligence, data mining/predictive modeling and Corporate Performance Management, analytics covers a broad range of tools and methodologies to inform, alert, assess, identify and investigate business trends and results. These tools have always operated in the background by extracting data from diverse systems (Data Warehousing), operating off-line and usually out of the loop. To the extent that analytics informed or guided users, implementation of the insights was always informal and not part of a systematic flow.

Business analytics needs a new set of principles to be effective in this new approach. Currently, the market is dominated by a handful of vendors who offer a diverse set of tools to accomplish a set of overlapping, but not always effective, after-the-fact tasks. In a typical

**Figure 1: Current Definition of Business Intelligence**

large organization, there may be a half-dozen different (competing) products in use, each with its own interface, connection requirements, security, metadata and learning curve. Different types of informational tasks, from reporting, to investigation to modeling to being alerted (to what?), use different “modules” within the same product suite. Overall, the adoption rate is very low: from 10 to 20% based on various studies published in the past year or two. Analytics at the tactical level is typically embedded as part of a process-oriented application, adding some degree of optimization, but only to each “silo.” Examples include some automation of marketing in CRM applications or optimization in supply chain tools. There is, however, little abstraction of the models or rules across multiple silos, making it difficult to leverage their capabilities across multiple data streams and other applications, except in rare instances, and at considerable cost.

## **THE ANALYTICS GAP**

Reporting/analysis tools for business people evolved in parallel with the operational disciplines, first through various host-based Decision Support Systems, then PCs and LANs and finally, Data Warehousing and Business Intelligence. There are entirely different industries for these two

spheres of computing, but forces are coming together to close this decades-old gap. The rationale for keeping them separate is failing the test of time. Software architectures have historically been managed from scarcity, and service levels for operations were maintained by jealously guarding the resources they inhabited. Today, hardware is abundant and far less expensive than the alternative - redundancy and inefficiency. Unfortunately, much of what passes for “best practices” in analytics today is still based in these assumptions. However, people are not lined up to use BI and are kept from it by IT. On the contrary, there is a general lack of interest, mostly caused by factors such as lack of relevance to their work and lack of understanding of the data and the (largely hidden) models behind it, but loosening up the resources bottleneck would be very useful.

Businesses have come to realize that analysis in isolation is not effective and that “closing the loop” by locating analysis close to, or even inside operations, can pay large dividends.

## **CLOSING THE GAP**

As operational software evolved from small, homemade programs, to a mixture of packaged and developed applications, and eventually enterprise solutions like SAP, the focus shifted from applications to business processes. Today, after 10-15 years of experience with process-oriented design, the similarity in many business processes, across organizations and even across industries, is creating a market for commodity business processes<sup>5</sup>. Tom Davenport recently wrote, “...outsourced processes will become a commodity...the move to process standards makes so much economic sense that it is probably inexorable.”<sup>6</sup>

The critical element of this move to commodity processes, and outsourcing them, is *standards of measurement*. For the first time, analytics will be central to managing processes in real-time, not just measuring them after the fact. This is a radical change in the software world. From a technology standpoint, this will be facilitated by SOA and Web Services, which provide the standards for embedding and coordinating analytics with all operational processes. It will also require much more capable tools than many BI vendors currently offer. The historical gap between operations and analytics is closing very quickly and some adjustments are needed.

## **OPERATIONAL VERSUS ANALYTICAL**

What really is the difference between operational and analytical processes? One revealing aspect is the data itself. Operational data is mostly atomic data and analytical data is mostly synthesized (from operational data). For example, a scanner, such as the one used in a grocery store checkout line, an employee ID card required to enter a secure building, or the ones used for package delivery, records some instantaneous information about the time, location and object of interest. This can be a person, a package or a box of Oreos. There is no question about the meaning of this data, it is logged in the appropriate tables and secured. Suppose employee entrance and exit from the headquarters building is tracked and logged. An operational system may need to know, “Is Roman Bukary in the building right now? If so, page him to come to the personnel office to pick up the keys to his new company BMW.”

Analytical data, on the other hand, may start with atomic data, but it is by examining many instances of it, or by merging and crunching it, that insight is drawn. The development of insight, however, can occur in many different ways. Traditionally, it was assumed that numerate analysts would find nuggets in the data by exploring, forming hypotheses and testing them with the data, looking for anomalies and patterns. Alternatively, insight could be derived through formal processes such as statistical analysis, data mining and predictive modeling. These results could be rendered into rules, parameterized models or other guidelines for other analysts. Extrapolating from the previous example, for example, let’s suppose a corporate HR department needs to know the mean time of arrival of people in the building, and the variances classified by day-of-week,

week-of-month and normalized by season. This would use the same atomic data, but sift and screen the results based on some simple statistical calculations. Or perhaps the company wants to know the average time spent in the building by employees, sorted by job class (obviously, it is not desirable for outside salespeople to spend much of their day inside, but it is desirable for Accounts Payable clerks to do so). Analytical work gets even more interesting when data sources are merged. Suppose HR wished to understand the commuting patterns of employees. The badge transactions could be merged with internal HR data giving the home address of personnel, and even use advanced techniques like graph/network analysis to weight the distances by routes and travel times by time-of-day. This could generate an evaluation matrix to truly understanding how much time people spend both in the building, and traveling to and from work.

Until now, applications, large or small, were designed and implemented with the expectation that they would perform reliably, but mostly within the boundaries of their explicit mission. These “silos” proved to be difficult over time as the development of “interfaces” between them to provide needed business functionality became overwhelmingly complicated, difficult to maintain and a barrier to innovation. Enterprise systems that took a horizontal view of business processes instead of the prevailing vertical functional view solved many of these problem, But some remained, especially the integration of external processes brought about by e-business. Originally a proposed solution to managing Web traffic efficiently and consistently, Web Services was quickly recognized as solution for systems architecture behind and beyond the firewall. SOA (potentially) provides the ability to decompose applications in services and to expose them to a wide audience. Which technology was employed to build them and which platform hosts them becomes irrelevant, provided they adhere to a few relatively simple standards. What is not known for certain is whether these early standards will hold up to the volumes and complexity resulting from an expanding market, or how they will evolve and maintain their independence from any one vendor or methodology.

What is certain is that, for perhaps the first time, it is possible to compose applications from components that are already available, and to configure applications to perform tasks that are more heavily weighted by the business requirements than they have been historically by available technology, tools, directions and prior decisions. This leaves analytics in an enviable spot. Instead of building a new call center approach, and relying on a separate set of analytical tools to evaluate the results, analytical services and insights can be embedded into the new models from services that are exposed, defined and operate through lightweight standards. The net effect is that, applications will no longer be identified as operational or analytical because the distinction will disappear. Will all applications be hybrids? Of course not. Many operational applications do not need analytical capabilities or can be adequately analyzed after the fact in much the same way we do today. Many analytical applications will continue to use data culled from multiple sources and retain the need to operate separately from them. What is clear is that the artificial gap between operational and analytical systems, a gap caused by previous generations of technology, is closing.

## **LOW-LATENCY ANALYTICS**

Both of the analyses mentioned above are “offline,” meaning, they happen outside the flow of the actual events. This is a typical use of analytics and it will endure, as there are an infinite number of questions that can be asked and they are limited only by the imagination of the analysts (and the quality of their tools). Offline analytics are useful for measuring things, understanding variances and solving problems, or asking questions to inform oneself and/or others. It is in offline mode that insight is developed. The area where analytics have been historically weak performers is lending themselves to decision processes that occur in the flow of business, especially in running business processes. This can take a number of forms, such as:

- People using real-time data: One alternative is to take the latency out of the data sources, but continue to use the same *interactive* tools. Instead of analyzing data that is a day or a week old, the interactive analysis is applied to data that is very fresh. The definition of “fresh” typically means less than a day, since it breaks through the barrier of overnight batch processing. For some applications, fresh can mean hourly, for others, only a few minutes or even seconds. If the requirement is for data that is more than a day old, by convention, this is not considered fresh or real-time, even if it is from the data consumer’s point of view.
- Analytical engines using real-time data: Instead of people deliberately interacting with real-time data, analytical procedures can operate in unattended mode on streaming data from running processes (as well as integrating the analysis with previously stored data). This is especially useful for capturing conditions that cause alerts to be issued, dashboards to be refreshed, or other actions initiated without human intervention. This is especially true in highly volatile situations, such as customer touch points, where a suggestion may be dispatched to a customer service rep’s screen based on the up-to-the-moment interaction, for example. The insight behind these decisions is primarily drawn from interactive data analysis, predictive models and other advanced analytical processes and represented in these analytical processes as models, parameters and rules.
- Operational processes employing analytical capability: This is an area where SOA has great promise because specialized analytical tools can be used across the enterprise, and beyond, without duplication of installations. An operational process can call on an analytical service to perform a specific function. For the time being, the composition of these “composite applications” will be done by people, in advance, but the dynamic configuration of applications is not far away.

Real-time analytics will play an important role in boosting an organization’s capabilities, but there is also potential to overdo it. A logistics cockpit for a manufacturer of consumer products needs up-to-date information, but not like the immediate information required by an air traffic controller. A call center runs more smoothly with predictive models keeping an eye out for changes versus patterns to alert management about increasing or decreasing call volume, but the urgency, though high, is not the same as watching for security breaches and intrusion against a network. Regardless of the freshness or latency of the data, embedded analytical capabilities that can be configured, deployed and re-configured for multiple uses are finally possible. What is missing is hands-on experience with SOA, good tools and methodologies, and a new set of analytical tools that are ready for SOA.

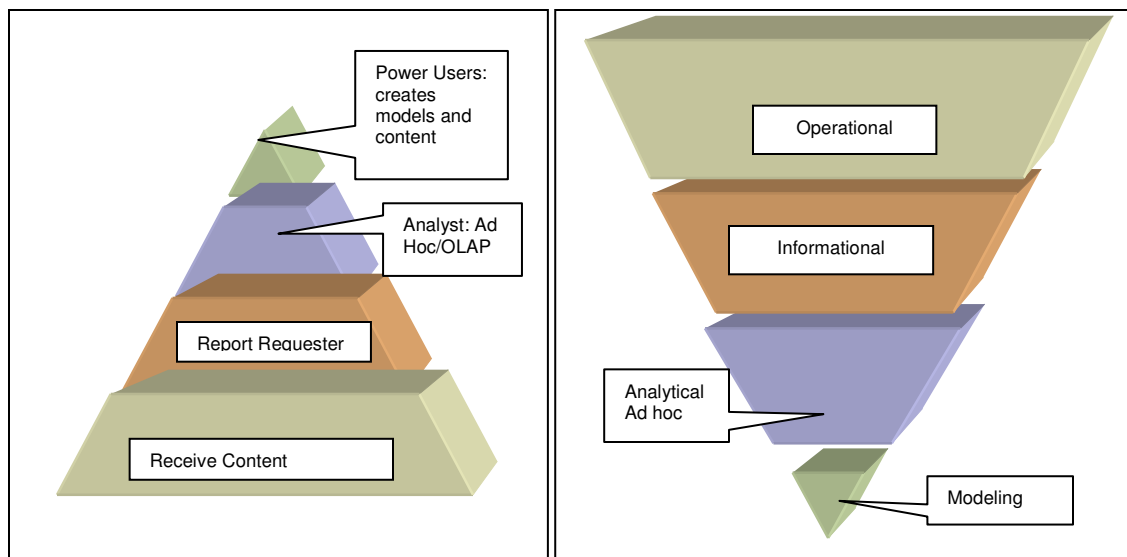
## ***INVERTING THE PYRAMID***

In the figures below, the pyramid on the left represents a typical BI implementation in a large organization today, with the levels representing the type of work performed. Power users, those who have the highest level of skill in using the analytical tools and the most access to resources represent a small but prominent role. The analysts, those who can operate the tools interactively and perform at a certain level of self-sufficiency, occupy the level just below the power users. The two bottom sections represent, by relative volume, the bulk of the other users, both those who perform some navigation and/or investigation, such as run-time selection of parameters, for example, and those who only view output from the system.

The inverted pyramid in Figure 2 represents the future state of analytics once SOA is widely implemented. Notice that the layers are now more descriptive of the purpose of the work, not the manner in which the technology used. Two new groups are represented: operational and informational analytics. Operational analytics run with the operational processes, rather than after-the-fact like most current BI, with analytics performing rules-based, real-time decisioning and intervention while measuring the performance of the processes themselves. Operational

analytics can take many forms, but essentially, it reacts to business events as they occur and either informs, such as alerts, or performs pre-built decision models. The size of this slice of the pyramid is meant to depict its relative proportion with respect to the other types. In this case, you can expect operational analytics to become the largest element, over time. Operational analytics employ the more powerful analytical tools such as statistical and predictive modeling, plus even more esoteric routines such as genomics modeling, for example. These tools are embedded in business processes that include elements of decisions, investigating, rendering and selecting. Process control applications, collaborative filtering, dynamic pricing and other forms of yield management are all examples.

Informational analytics combine, roughly, the two lower levels of the pyramid on the left. Analytics at this level informs through reports, alerts, syndicated feeds, dashboards and even mobile media. People who are involved in other levels, such as analysts, modelers or even programmers



**Figure 2: Inverting the Pyramid**

of embedded analytics are also participants in the informational level. Everyone needs to be informed and as analytics permeates organizations, a great deal of information needs to find its way to people who may be experts in their own domains, but are in the humble role of information recipients in others. The informational level, like the operational level, expands rapidly as analytics becomes entrenched into every corner of the business.

The analytical users remain, as their role is very important, in greater numbers but roughly the same proportion. This group includes the primary users of CPM, which is intimately involved with the formation of strategy and the monitoring of the results. Without these, there would be nothing for the other analytics to analyze. Their smaller size relative to the pyramid is no reflection on their importance. Performing analytical work and creating analytical models are very different skills, but they are not mutually exclusive. However, only some of the analysts are capable (or motivated or trained) for modeling.

Over time, as knowledge becomes easier to capture and represent, and applying knowledge becomes simpler, more people will build their own models and collaborate with others in modeling. For the time being, models will be discovered and built by a small group of highly trained technical professionals (statisticians, mathematicians, engineers or scientists, for example), not by IT departments.

There is a perception that analytics implies difficulty. Though there may be very sophisticated mathematical models behind some analytics, it is not necessary to understand the details of the theory to use them effectively. In most cases, the results of advanced analytics will be rendered into parametric models that can be inserted or called as needed without repeating the intensive investigations, such as data mining, that were used to derive them. Furthermore, powerful analytics will be applied indirectly, by alerting and informing managers when something approaches a boundary condition, such as a shortfall, a bottleneck or leading indicators that require attention. Offers of assistance or guidance will appear or “pop up” based on a particular person’s habits, choices and experience.

In the new analytics, people will interact with applications at their own level of skill and interest. It is possible to do this economically, because loosely coupled services communicating in standard protocols can be configured to solve a multitude of problems. Currently, these tools are duplicated across the enterprise, each one performing functions for its own intended area. It is not unusual to find a large organization with five or even 10 different software solutions for analytics. As a result, they are expensive, brittle and labor-intensive to maintain. In the new architecture, analytics are far easier and economical to deploy. For example:

- A professor needs to manage her research grants, and handle the staffing of the projects with post grads. This is a natural part of a professor’s job and analytics should make it easier to do it well. A reusable analytical service might employ advanced statistical or predictive modeling algorithms to arrange the lab schedules and keep track of the funds. It is not necessary to understand how these tools work to use them. Because the underlying tools can be applied across a wide range of applications, deploying and maintaining these services centrally can reduce redundancy and improve understanding.
- A VP of Sales in a growing company is constantly redefining sales territories as the staff expands, and new product and service offerings are added. Powerful analytical programs are standing by behind the scenes to assist in reconfiguring the map based on rules and principles established by, and modified by the sales department, not IT.
- Understanding demand effects by continuously analyzing sales through Point-of-Sale data in real time and adjusting replenishment is far more accurate than working from forecasts. In many cases, this is not possible with analytical functions embedded in a POS system because the data needed for the analysis spans more than one system. Analysis capabilities in current operational systems are unable to connect with multiple systems and data streams without costly customization (and subsequent maintenance).

## **FACTORS AFFECTING ANALYTICS**

Before e-business and the Internet, before business process reengineering, organizations were able to function in a timeframe that seems inordinately long by today’s standards. Few decisions needed to be made immediately, no one was in danger of losing their market to a competitor without a protracted fight, and the process for reviewing results and deciding on tactics was measured in weeks, not hours or minutes. It was in this environment that our current models for analytics and business intelligence were formed. It explains why data warehouses are still a largely batch process, why business intelligence software is still focused on “seats” or individuals instead of communities of work and whole enterprises. But today, the luxury of latency is gone. On demand is not just a clever TV commercial; it is the way business is being conducted.

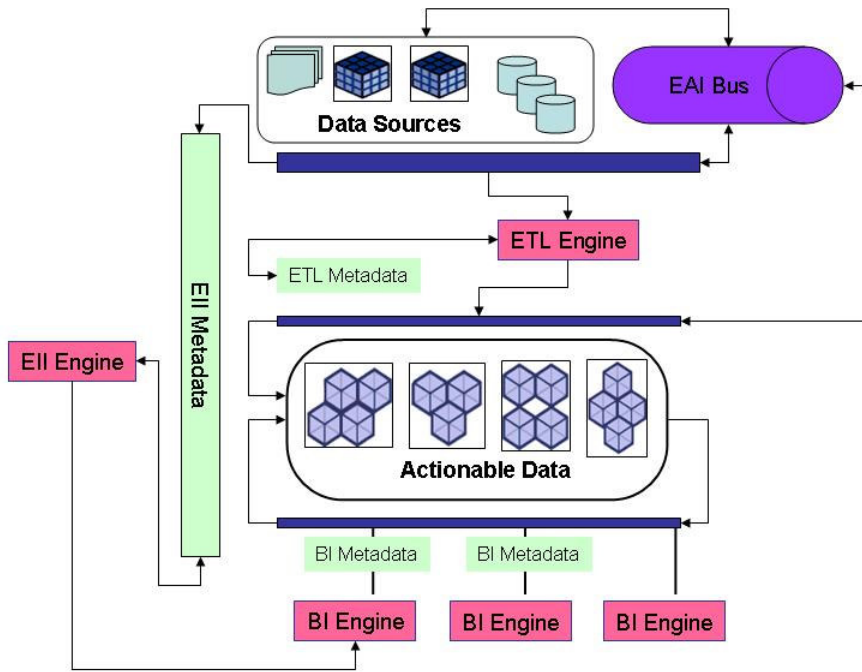
Business intelligence is still a departmental or even individual affair in most companies. The move to SOA will force IT managers to look at analytics as an enterprise asset. Because analytics will become an integral part of many operational systems, the current situation, with many different BI tools in place and skills spread too thinly to be useful, will become unwieldy. Standardizing on

those analytical services that are best suited to the propagating “composite” applications (operational and analytical), instead of each department’s selection, will have the effect of elevating the visibility of analytics. It does not mean that every organization will need to apply advanced analytics to their business processes; they will need to have faith in the output of analytical processes. There are various ways for this to happen, but one method that is gaining notice is to create a team of specialists who perform most of the advanced analytics on a centralized basis, and who have the confidence of the senior executives of firm. Their work, finding the underlying causes and relationships and predictors, can be reduced to models with a set of parameters that can be run repeatedly. FICO® credit scores are a good example of this, where Fair Isaac develops and continuously tunes a model that can be rendered as a series of variables that produces a credit score.

This premise, most recently advanced by Tom Davenport<sup>7</sup>, historically has some drawbacks, but they were likely caused by organizational separation of the modelers and those who used the models. Nevertheless, Davenport and others writing on this subject do at least highlight the need for these capabilities and focus on executive acceptance of analytics, which is crucial. In an SOA world, it is likely that advanced analytics will be palatable to a wide audience because it will be transparent.

## SEMANTICS

A major impediment to the use of BI tools is not the mastering of the tools themselves, rather this is caused by a lack of understanding of the data, what it means, and its relevance to the task at hand<sup>8</sup>. Metadata management is positioned as the solution to the problem, and it certainly is, but the current discipline of metadata management is not a complete answer because, in particular, existing BI tools provide incomplete and proprietary metadata. In figure 4, the location of different sources of metadata can be spotted in this *simplified and idealized* chart of information flow for analytics. A single BI tool may actually have as many as five different metadata structures, but gathering them into a single metadata repository without altering their form or function does not solve the problem.



**Figure 3: Fractured Metadata in Current BI Architecture**

Ideally, metadata should allow for different tools used in different locations (logical or physical) to be able exchange information about their semantics and operation. The lack of uniform metadata across applications makes it difficult to collaborate and hinders standardization. BI metadata comes in roughly three flavors:

- Production – Describes the movement and translation from one data source to another, establishes ownership and logs updates
- Catalog – Definition of tables and attributes
- Presentation – Additional data and calculation definitions, report layouts, preferences and roles in the reporting/analysis tool

A great deal of work has gone into defining metadata standards and attempting to open up the metadata, but to a lesser degree, competitive pressures between the vendors impede the effort. To a much greater degree, however, technology itself stands in the way of a useful metadata solution. Using relational databases and relational modeling techniques as a solution may be inadequate to capture the richness and nuance of the data, models and conditions in even the simplest business. Using extended relational modeling, such as UML, offers some marginal improvement, but all of these schemes effectively leave metadata in a passive state.

Semantic technology holds the promise of breaking through the metadata ice jam. Semantic technology is a broad field, but the rapidly growing commercialized part of it is called *ontology*. An ontology goes beyond definition to capture the semantic meaning of things and, as a result, creates a structure from which a machine can draw inference. In a metadata query, all of the knowledge needed to frame a question is in the query, composed by the query writer. In ontology,

the relationships that are captured are capable of revealing, through a process of introspection, more information than was consciously placed into it. In addition, ontologies are constructed in a language that is based on XML, the Resource Description Framework (RDF), or built with the Web Ontology Language (OWL), all of which are open standards that are managed by the same committee, the W3C that is responsible for Web Services.

Semantics will play a role in the orchestration of SOA in a multitude of ways. The discovery of services using UDDI and WSDL is very limited, but semantic extensions to these services make it possible to find a service with a conceptual search such as “Find a seasonal smoothing function that is used for consumer products.” The massive amounts of data in data warehouses no longer need to be limited to a single, isolated definition. Semantics can point out which elements are dependent on others or are predictors. Semantics can be extremely valuable even when they are incomplete. Semantics can be constructed incrementally by many people simultaneously and incorrect entries can be handled by the semantic engines. There is even a name for ontologies that are built by people over time – *folksonomies*.

Semantic technology promises to supercharge analytics by removing the major obstacle to using analytics by a broad cross-section of the population – lack of understanding. In the near future, metadata built with semantic technology will glue all of the disparate services together, with meaning replacing definitions and patterns. Current metadata approaches are largely based on data; with semantics, metadata can marry data, process, application, use, security – virtually everything. Furthermore, it will be able to exchange this information freely because it is based on open standards that conform perfectly with and SOA approach.

## **PRINCIPLES**

Many current analytical tools limit and simplify the models and scale of data that can be investigated because of their client-based architecture. Even in server-based BI, many of the tools offer architectures that scale poorly and simply cannot leverage the volume and richness of the available data. To claim that a product is scalable because it can handle dozens of servers, but only 10-12 users per server, and perform no real load balancing or work distribution across servers, only complicates the result. IT groups are forced to limit access to data warehouses by concealing detailed data, implementing query governors and/or limiting live queries to certain times of the day. As a result, despite huge data warehouses housing the freshest and most detailed data available, the models that are used in BI are typically summarized, simplified and denatured to fit into the limited computing capabilities of most BI tools.

What next-generation analytics promises, particularly as a result of loose coupling and open standards, is the horsepower to finally exercise the investment in data warehouses. However, these limitations are not the only factors that limit the reach of analytics today - they may not even be the most important factors - but they are the ones that can be dealt with easily through technology. Changing the way organizations use information (analytics in particular) to run their business, breaking down the practices that separate knowledge workers from useful tools and information, and educating people that passive receipt of information is not sufficient and levels of self-service and investigation are not beyond anyone’s intellectual reach – all of these are challenging problems to solve.

Manipulating very complex models and walking through reasoning, layers of detail and concepts does not have to be difficult. Quality analytical tools should adjust to the user’s level of skill, and more skilled users will be able to create analyses for others to share. And finally, because all aspects of the analytical workplace, including data, relationships, models, assumptions and report objects will be recorded in a semantics-oriented abstraction layer. This will be similar to but much more robust than today’s metadata. The sharing of knowledge, the ability to use encoded

knowledge in new and creative ways, and the emergence of collaborative analytical environments will become the “killer apps” of the SOA world.

## ***ANALYTICAL FUNCTIONALITY***

Here is a partial list of analytical functionality that is needed in the next generation of SOA-enabled analytical tools. Many of these features already exist today, but the column to the right describes how they need to change:

<b>Analytical Functionality</b>	
OLAP	Separation of navigation/presentation from engine; OLAP-like manipulations, such as drilldown, need to be available as services without human interaction
Descriptive statistics	Common statistical engine across all applications boosts understanding
Threshold-based alerting	Event-based service watches and alerts and is configured for a multitude of processes; this feature should be generalized and not linked to a particular data source or schema
Predictive modeling	Instead of a variety of specialized tools implemented by disjointed groups, predictive modeling tools can be implemented as shared and reusable services for all of those qualified to use them and parameterized models derived from their discoveries implemented as services and embedded in many operations
Optimization	A generic optimization engine with different add-on optimization engines, with domain intelligence, applied to the right problems
Automated Decisioning	Combination of event manager, predictive modeling, optimization, rules engines and other decision tools to take over the task of low-risk decisions and inform those responsible for higher risk ones
Dashboards that learn what you want and how you want to see it	Dashboard software, abstracted from most of its current functionality, focused on learning patterns and needs of clients
Ontologies in place of proprietary rules engines	Semantic technology in open standards such as RDF using conceptual search techniques to draw inferences without explicit rules

## ***APPLICATIONS OF SOA ANALYTICS***

Analytics that were difficult and/or expensive to deploy previously can be deployed in the new architectures. Although they are not necessarily new applications, their impact and reach can be much broader. In most cases today, analytics embedded in operational systems are purpose-built for that single application, so knowledge and insight gained through their operation is not reusable. Service-based analytics can be abstracted from applications and applied where needed and stitched into operational processes. Some examples of these hybrid or composite applications are listed in the table below. All of them exist in some form or other today, but the loosely-coupled standards-based nature of SOA will provide a more economical and practical way to use them:

Some applications of SOA Analytics	
Application	SOA Impact
Data mining with marketing automation	Closing the loop between list creation and list execution through orchestration of services for CRM and analytics
Dynamic pricing	Not only for e-commerce, but any transaction
Supply chain – switching channels	Monitoring channel for shortages, quality problems, switching in midstream
Product replacement	Slow-moving product, limited shelf space (or seats, or beds)
Promotion planning	Optimization of multi location promotions for sites, media planning
All forms of e-business	Personalization, dynamic configuration, collaborative filtering
Visualization	Abstraction of data, metadata and models allows for engine to maximize visualization capabilities and shed other functions

## ***Types of Analytics***

### **PREDICTIVE MODELS**

Predictive modeling is a mature discipline that is continually refreshed with new algorithms and especially new technology to perform it. Predictive models are employed for a full range of challenging problems, from strategy and scenario planning to dynamic pricing. In most cases, the outcomes of a predictive modeling effort are reviewed by experts. For example, a decision to change the product mix, or to enter a new market, is not something that is left to algorithms alone. However, one area where predictive models are deployed in an unattended mode is where there are a great number of small decisions and getting most of them right is what that counts, not every one. For example, when directing a 2,500-person sales force where to place promotional material on a given day, a certain level of error is acceptable, but automatically pricing a popular item on a Web site 50% below your cost can have devastating effects. Predictive models are not useful for unattended operation when even a single wrong occurrence can have devastating effects. Alternatively, a certain level of error is acceptable when each individual mistake has minor and non-recurring consequences. Currently, predictive modeling is implemented as either a standalone application or embedded in each operational application, as part of the application logic. In SOA, standard predictive modeling services can be dispatched and will be able to easily access historical data (data warehouses principally).

### **OPERATIONAL ANALYTICS**

Operational analytics is defined by the part of the business in which it is used, not by the technology that is employed. It can range from monitoring things real- or near-real-time to reviewing metrics and KPIs of the day-to-day results of the business. What distinguishes operational analytical work from strategic analysis is the focus of the analysis – on the effects of small operations, decisions and transactions of the business and the effort made to understand and improve the functioning of the business. Building a business plan for a new product is strategic work; analyzing the facts of the supply chain, customers, logistics and finances of a new product is operational analysis. What is different in an SOA world is the ability to attach analytical processes mid-flight without the need to capture data streams and store them beforehand.

## INTERACTIVE ANALYTICS

The technologies of interactive analytics, as we know them today, are quite useful. OLAP, the ability to navigate through the dimensions of data, with the tool flawlessly performing the aggregating or disaggregating while simultaneously preserving the context and recalculating the derived values in the proper context, has been perfected over 15 years. But it has never spread through organizations to extent that it should have. There are many theories for this, but the most likely cause is that the ergonomics of OLAP are fine, but people either don't wish to investigate data at that detail or they simply don't understand it. But interactive covers a broad range, and parameterized reports, where the report format and query outline is fixed but the ranges of certain values are selected at run time, is still quite popular. However, this is really only interactive report selection, not analysis. The ability of people to assemble requests for information from larger components, selected by concept, not identification, and supported by a semantic model, not relational (or even proprietary) metadata would be a great improvement.

Type of Analytics	Currently		SOA	
	Method	Data	Method	Data
Predictive	Custom standalone logic or embedded in each application	Some real-time transactional, but mostly batch	Logic abstracted to metamodels, reuse base; service; closed loop	Federated: Real-time transactional + historical + external
Operational	Mostly after-the-fact through ODS and data warehouse and specialized tools	Delayed transactional	Same as above	Same as above
Interactive	OLAP, dashboards, ad hoc via many different standalone tools	Data warehouse/marts 1 month – 1 day old	Engines and data abstracted from "skins" and user functions	Same as above

\* There are exceptions, particularly with the use of embedded decision engines, but their use is not widespread and their implementation costs are high.

## VALIDATION FROM REAL COMPANIES

To become and remain globally competitive, businesses need to invent and apply the most useful and productive technologies and processes to continuously innovate their products and services. Boeing and Procter & Gamble adopted this philosophy before implementing SOA, but it was more difficult. But since they already operate in this mode with analytics at the center of the most important applications, moving to SOA will be no more than a technology change.

### Boeing

"As part of Boeing Technology, it's the mission of Boeing IT to transform Boeing into a network-enabled enterprise with a design-anywhere, build-anywhere capability," said Scott Griffin, vice president and chief information officer for Boeing. "This includes establishing a single enterprise process and systems architecture in order to ensure that all our IT investments move us toward that goal. We want to make sure we are investing in the right technologies to meet both the short- and long-term business needs of Boeing, that the business units are not inadvertently duplicating

efforts, and that our technology investments—as well as those of our strategic partners and suppliers—are being leveraged with maximum benefit across all the business units.<sup>9</sup>

## Procter & Gamble<sup>10</sup>

While many companies put advanced analytics in their engineering or research departments, P&G put its OR (Operations Research) group in its IT shop. "IT is one of the very few organizations within a company that has true end-to-end visibility into all the operations of the company," says Glenn Wegryn, one of the analysts on the project and now associate director of P&G's 17-person OR group, known as IT Global Analytics. "Part of our success is that we are a crossover capability; we have as much business knowledge as we do knowledge of our instruments and tools."

P&G applies IT-assisted analytic techniques in three broad areas:

- Optimization models to allocate supply chain resources.
- Simulation models to try out various options. "We've found that the success of a supply chain is not necessarily operating at the absolute optimal solution, but operating at one of the more robust solutions in the real world," says David Dittmann, a Global Analytics manager.
- Decision analysis, such as decision trees that combine the probabilities of various outcomes and their financial results.

But it would be a mistake to give all the credit for the success of OR at P&G to higher math and clever algorithms churning through terabytes of information. "What's really the key to OR is not just the technical side; it's also the very deep understanding of how a business operates," Wegryn says.

P&G recognizes that it is not sufficient to have a facility for using analytical tools - it has to be combined with excellent knowledge of the problem domain as well. These complementary skills do not have to exist simultaneously in the same person. In fact, the best analytical work is usually accomplished by teams of people with a broad cross-section of skills.

## CONCLUSION

In conclusion, analytics are crucial for both making policy and managing it. In the past, different vendors supported different groups within organizations to provide tools for planning, budgeting, forecasting, statutory reporting and other high-level, policy-making and strategic requirements, on the one hand, and analytical and reporting functions after the fact on the other. These efforts were disjointed, incompatible and often conflicting. To the extent that operational systems employed analytics, they were typically embedded in the software and neither reusable nor transparent to the rest of the operations.

Thanks to the onward march of technology, and to Moore's Law and SOA in particular, the gap between strategy, operations and analytics is closing. The SOA environment provides an infrastructure to support dynamic reuse, which requires new classification systems for BI services and policies that are equivalent to operational service policies. If the function of operational and analytical software is going to blend, then these elements will have to come into compliance too. Analytical software as it is today is not entirely ready to be part of a distributed, enterprise computing architecture. Many of the products, including the market leaders, have a lot of work to do to before their products can perform effectively as a set of loosely coupled asynchronous services.

There is still a great deal of work to be done in analytics. The past ten years were focused more carefully on data management than on the application of analytics. Though the problems of data quality, data freshness and data conformity have not been solved, and issues about metadata and master/reference data are still brewing, an enormous amount of progress has been made. Ten years ago, data warehousing was a good idea but implementing one successfully was very difficult. All of the time and attention spent lessening that risk came at the expense of understanding good analytical technique and architecture. Today, there is no time to waste. Because analytics are critical at both the strategic and operational levels, engineering analytics to perform smoothly across the organization is a top priority. Every organization will need to find their own unique way of implementing it, but we will all depend on the vendors to provide the tools and the technology to get it done. . .

## ABOUT THE AUTHOR



Neil Raden is the founder of Hired Brains, Inc., <http://www.hiredbrains.com/knowout.html>. 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 Semantic Technology industries. Based in Santa Barbara, CA, Raden is an active consultant and widely published author and speaker. He welcomes your comments at [nraden@hiredbrains.com](mailto:nraden@hiredbrains.com).

## ENDNOTES

<sup>1</sup> Enterprise Systems, "IT Spending Rebounds," February 15, 2005, <http://www.esj.com/news/article.aspx?EditorialsID=1280>

<sup>2</sup> Thomas H. Davenport, "The Coming Commoditization of Processes," *Harvard Business Review*, June, 2005

<sup>3</sup> Forrester

<sup>4</sup> Loosely coupled refers to the ability of services to be anywhere on a network, without regard to platform or development technology provided they can be discovered and invoked through the standard methods

<sup>5</sup> And some that outsourcing includes analytical processes, too. American Express now outsources its data mining to specialist third-party firms that can process millions of transactions per day to reveal purchasing patterns and other aspects of consumer behavior. Best Buy, the electronics retailer, is outsourcing not just database management but also marketing analysis and the complete execution of the marketing programs for two of its six segments – business and high-end customers.

<sup>6</sup> Thomas H. Davenport, "The Coming Commoditization of Processes," *Harvard Business Review*, June, 2005

<sup>7</sup> Thomas H. Davenport, *Competing on Analytics*, Babson Executive Education Working Knowledge Research Center, June 2005 (available through BPMI – Business Process Management Initiative, [www.bpmi.org](http://www.bpmi.org))

<sup>8</sup> <http://www.intelligententerprise.com/showArticle.jhtml?articleID=19502113>

<sup>9</sup> <http://www.boeing.com/news/frontiers>

<sup>10</sup> <http://www.computerworld.com/printthis/2005/0,4814,99484,00.html>