Advanced Practices Council

Leading Digital Transformation Through Research

A Program of SIM

SIM Advanced Practices Council (APC)

The Society for Information Management's (SIM) Advanced Practices Council (APC) is an exclusive forum for senior IT executives who value directing and applying pragmatic research; exploring emerging IT issues in-depth; and hearing different, global perspectives from colleagues in other industries.

Table of Contents

Executive Summary

Most organizations have or plan for data science teams to parlay data assets into business intelligence, predictive analytics and prescriptive analytics solutions. Such solutions will inevitably be embedded into core applications and business processes essential to running the enterprise because faster speed-to-decision can create a competitive advantage. Effective organizational governance of such infused analytics solutions is necessary to mitigate the risks associated with analytics that may go awry, resulting in unintended consequences. Governance also avoids losing critical information on assumptions underlying analytics solutions when data scientists leave the organization or when environmental changes render solutions obsolete or invalid.

Consider the Target fiasco with coupons aimed at a teen-aged girl algorithmically deemed to be pregnant. Her family did not take kindly to being informed by mailings signaling the situation. Also, think about the mortgage derivatives that were rapidly devalued leading to the Great Recession. It was clear that investment ratings agencies' predictive models were decayed and no longer relevant as the financial mess cascaded into chaos. Also, consider the flash crash of 2010, when the stock market took a 1000 point swing in one day. In 2015, a person responsible was named and indicted, but it was also noted that programmed trading, which relies on predictive analytics-based algorithms, drove the volatility to heights never before seen. In short, embedded analytics can create unintended consequences, and lead to customer relationship disasters and brand damage.

No organization has found the best possible governance approach because we are in a pre-paradigm state of knowledge. However, grass roots approaches used in organizations today, synthesized with some research in the area, provide concrete approaches that can be leveraged by any company at any point on its analytics and business intelligence journey. In this report, we summarize current best practices in a framework that clarifies the levers and design choices available today, helping readers learn how to design a state-of-the art governance approach. Further, there are integrated design choices that can help an organization customize its analytics governance approach. Structural, resource, and model management issues as well as a host of best industry processes are discussed.

Introduction

In the early 1980s, a data scientist analyzed available data sources and found that those customers of a northeastern bank who were sometimes delinquent in their checking accounts were actually more profitable over the long haul than those who always had adequate funds in their accounts. The analysis reflected the overdraft fee structure in place at the time. Armed with this discovery, the evidence to back it up, and approval from the business, this early data scientist requested that IT make changes to applications in order to discontinue the practice of culling delinquent customers. IT responded that the change would have to wait because of higher priority requests to change the existing COBOL-coded applications. Not unlike today's data scientists, this pioneer learned to program on his own and made the changes in the relevant applications himself.

Twenty-five years later, this COBOL code is still running today at that bank, which probably doesn't even know that this analytics solution is embedded in its legacy applications. Data scientists who are currently reviewing the bank's customer data warehouse don't know that ever since the COBOL code change was made, the customers whose transactions are in that data warehouse may have been delinquent account holders. How might that knowledge impact their data mining efforts? Given the impact of the Great Recession, the set of parameters relevant to customer retention, an approach hidden in the COBOL code, may well be outdated.

What does this teach us about analytics? We need governance mechanisms in place to know what analytics are embedded where in business process and workflow applications. Further, we need to ensure that subsequent data mining efforts incorporate knowledge of rules that existed when the datasets being mined for new discoveries were realized. How many organizations are in the same boat as that northeastern bank with respect to its early-day, hidden and mysterious data science dalliances? According to a recent Logi Analytics study on the state of embedded analytics, almost one-third of business intelligence and analytics discoveries are infused in a user interface or business process, and more companies are moving in this direction.

The governance of analytics in most organizations is at a nascent stage of maturity. In the following sections, we review emerging grass roots approaches to analytics governance and extract a framework of governance design best practices. Design options in this framework give rise to levers that companies can engage to create a customized approach to analytics governance. As embedding analytics becomes the du jure deployment approach, the need for governance becomes an important imperative.

Background on Predictive Analytics and Business Intelligence Deployment

Some characterize the current state of business intelligence and analytics in many organizations as one where data science cowboys operate in a wild west of innovation and deployment. Organizations are dipping their toes in the data science water and finding pockets of successes in marketing, finance, supply chain management, and other areas. Most often, organizations undergo a journey-like transition from descriptive to predictive and ultimately to prescriptive analytics. They evolve from building a data warehouse or lake while ensuring data quality before moving to the next stages of the journey. Often the deployment of analytics at the descriptive stage is in the form of user interfaces either in a new application separate from existing applications or there are dashboards and reports provided within applications to provide decision support to those qualified to access them. Note that these applications likely support actions taken by a decision maker. They may provide self-service analytics, but a human remains in the loop. In contrast, the deeper embedding or infusing of analytics into an organization’s core applications is most often associated with the predictive or prescriptive phases of the journey. Models developed through data mining, visualization, and other means become an integral part of programmed business applications and automated workflows. Predictive modeling and how models are deployed are referred to as predictive analytics.

Data scientists construct predictive models using data mining approaches, including classical statistical methods such as clustering and logistic regression, machine learning algorithms, specialized unstructured data mining algorithms, and other big data methods like those associated with Hadoop (e.g., MapReduce) and Spark (combines SQL, streaming and complex analytics in one seamless environment). Once a significant pattern is discovered by a data scientist, the associated predictive model is deployed by embedding it within an application. Embedding a predictive model requires that the algorithm capture the data inputs so the model can make computations using those inputs and parameters that were "learned" in the data mining process. Outputs of a predictive model are decision-oriented and can be made to drive appropriate application logic. For example, a predictive model might be constructed for making a customer an offer in order to entice a sale. Such a model can be used to score potential customers so that those with higher scores can be targeted. By leveraging the model, the organization can generate more profit than if all members of a population were provided with the offer or if the offer had been made randomly.

One might think of a model’s output score in the manner of a FICO score, which is determined by a predictive model that assesses one’s credit worthiness. There may be many inputs to the model and details likely remain proprietary because there is significant value in the discovery of an important pattern. FICO scores can be used to support loan decision making in order to reduce risk and thereby increase a lender’s profitability. The decision rules that leverage the FICO score can be coded directly into applications. In this scenario, a potential customer can quickly be considered for a loan with a decision making engine infused into an e-commerce website’s workflow. A predictive model has what is often called an "uplift," that is, the model provides improved value over a context where the model isn’t deployed. For example, leveraging a FICO score to segment customers provides lift over granting loans randomly or to everyone who might apply. Model lift can decay over time as the environment changes. Also, original assumptions or business circumstances can change in ways that render the model out-of-date.

Figure 1 shows the connection between data mining and deployment processes. The left cycle (SEMMA) depicts the data mining process and the right cycle (DEEPER) depicts the deployment steps. To build a predictive model, a data scientist prepares a relevant data set to use to "teach" a learning algorithm. From that set, an appropriate sample is selected; most often a hold-out dataset is put aside for further model assessment during the final phase of the cycle. The teaching data set sample can be explored using visualization methods, and descriptive statistics are often analyzed to determine if the sample contains, for example, outliers that ought to be removed before further analysis. Data scientists might opt to transform some of the variables in the data set in this modify stage. For example, some logical ratios might be constructed that combine variables in a way that provides information value to a learning algorithm. For example, current assets divided by current liabilities (the "current ratio") is a transformed variable that offers more potential predictive power than would simply asking a learning algorithm to consider the current assets and liabilities measures separately. Different machine learning and statistical model building algorithms might be leveraged in the model phase. The result is a trained model that can be assessed using the hold-out set mentioned earlier. Often there are multiple trained models constructed; a data scientist might designate one that performs the best on the hold-out sample as the champion and the others as challengers. Once a champion model is designated, the deployment cycle can be initiated.

Like the SEMMA cycle, the DEEPER cycle is a generic sequence of steps that should be contextualized to an application domain. For example, suppose a loan application's champion model, like a FICO scoring model, was the result of a SEMMA phase. In this case, DEEPER is where that scoring model would be deployed in business processes. Loan officers would be trained and new automated decision rules implemented.

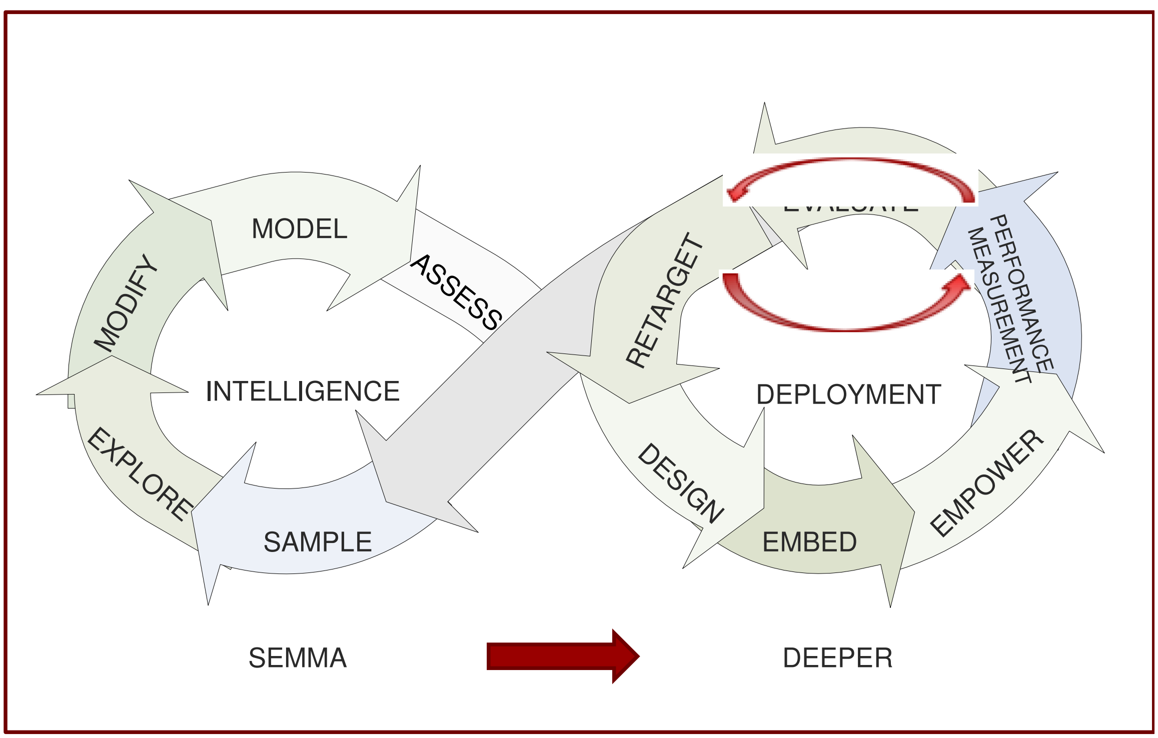


Figure 1: The Connections Between Data Mining and Predictive Model Deployment

Design

The design phase is a planning step that sets a duration or life for the deployment of a predictive model. Model builders and domain experts collaborate to establish the duration because there is potential for model decay and there are likely model life expectations drawn from the business case. A deployed champion model's efficacy may decay over time. This means that the model's uplift is no longer being achieved because of what is called "concept drift" or a context change that renders the model's assumptions invalid. Challenger models may be substituted for a decayed champion model. A strategy for monitoring performance and "model switching" protocols is required.

Embed

During the embed phase, predictive models are infused into core business applications. Consider Amazon's recommendation system that is based on predictive models. Where exactly should the model's results be displayed on the site and at what point in a customer session should those recommendations appear? Should recommended books be treated the same as recommended electronics? Embed decisions involve complex implementation schedules as core applications are modified. In this Amazon case, there are important considerations in how the ecommerce site might cross- or up-sell. For example, trying to up-sell a customer a product that isn't in inventory is a problem.

Empower

The empower phase encompasses employee training as well as relevant process changes. For example, if predictive model deployment requires new hand-held devices for sales personnel, they will be trained on the role of the model in the sales process, how to use the new device, and best practices for customer engagement given the change. In such a case, a mobile hand-held device might make use of a graphical predictive model that assists sales persons in identifying other retail items that a customer might be interested in based on an initial selection.

Performance Measurement

The performance measurement phase involves ongoing monitoring of the performance of a deployed predictive model. Depending on the nature of the model, this may involve tracking the efficacy of the predictions over time to ensure expectations are being met. Since models decay or there is concept drift, the predictive ability of the model is not guaranteed to last. For example, in the changing economic climate of the Great Recession, a type of concept drift, many financial models fell short because their prior assumptions about the economy were no longer relevant or the models did not account for input parameters that became more important in the altered economic situation. For performance management, expected bounds of input variables might be tracked and alerts triggered when such variables fall outside of expected ranges. If the model's impact fails expectations, perhaps because sales decrease for even the most highly scored customers, then retesting the model may be required. There is a cycle within deployment where performance measure and evaluation are ongoing until a retarget decision is made.

Evaluation and Retarget

In the evaluation and retarget phases, predictive models are evaluated for decay or concept drift. One outcome of evaluation may be a switch from the current champion model to a challenger (a retarget using a previously prepared predictive model). In this case, an infused model gets swapped out with another that had been prepared and vetted beforehand. In instances when swapping a new model isn't an option, the entire SEMMA phase may be reinitiated to build a new portfolio of models.

Situations like this arise when there is significant performance degradation, concept drift, or when the environment has changed so dramatically that prior historical patterns are deemed no longer relevant.

Deeper In Context

The level of risk in environments in which predictive analytics are deployed determines monitoring and management. A predictive model deployed to purchase stocks in real-time represents a deployment environment that is riskier than that of an embedded model to help decide who should receive a catalog. A decay, failure, or the realization of concept drift can have significantly different consequences in these two scenarios.

In addition to environmental complexity due to risk, DEEPER process complexity can influence how an organization monitors and manages predictive analytics that have been infused into business processes. On one hand, some phases of the DEEPER cycle may be delivered in a straightforward way; there are clear stakeholders to involve, and the step-by-step phases can be iterated without too much confusion or contention. However, in a more complex DEEPER process, significant cross-discipline coordination might be required at each phase, and implementation of the model into the business process, training employees to use the model effectively, and monitoring model performance across multiple business units may be difficult. For example, a complex DEEPER process might involve the deployment of a predictive model to automate procurement of raw materials where the model's inputs include outputs from both inventory and sales predictive models. In this case coordination is required across sales and marketing, and there is risk that one or both of those models might decay, thereby rendering the procurement model obsolete. There could also be unpredictable interaction effects between the models. The possibilities for unintended consequences are increased in these situations.

Integrating considerations regarding environmental and DEEPER complexities provides a continuum for considering how organizations can mitigate risk through judicious predictive analytics governance. Many organizations are just getting started in analytics; they are experimenting with prototypical projects or working to build analytics depth in one business area, such as marketing. In such cases, projects are typically selected where predictive analytics won't be deployed in mission critical applications, but where they have high potential return on investment. In addition, projects are likely to be low in DEEPER complexity in order to deliver solid proof-of-concept evaluations. These low environmental complexity cases (where DEEPER complexity is also low) are reminiscent of early days of the railroads where building new railways required a fairly straightforward organizational governance approach. This is depicted in the lower-left quadrant of Figure 2. Stakeholders have specific roles, there is a linear progression to completion at each stage, and progress is measured one win (e.g., railroad track link) at a time.

Organizations will not remain in this start-up mode for very long once the advantages of predictive analytics solutions become apparent. There are two trajectories typically followed. One is to delve into predictive analytics projects where the environmental complexity is high but DEEPER complexity remains low. Such is the case with a fairly small analytics team where each member understands his role and the solutions they can deploy become more mission critical. This scenario is depicted in the upper-left quadrant of Figure 2. Here we have a set of railroad tracks that are complex, but governance requirements is not challenging if only one or two trains pass through in a day. This is similar to the small, well-organized data science team that manages one project at a time – even when their projects are mission critical and the environment is complex.

At the other end of the DEEPER complexity continuum, some organizations move from their proof-of-concept stage to deployment across many functional areas and in many different business contexts. The requirements for keeping multiple stakeholders updated on the projects, coupled with the cross-discipline nature of many projects, may require a project portfolio management strategy. By keeping the projects on separate tracks as depicted in the lower right quadrant of Figure 2, the governance of predictive analytics deployment can be separated by project type, domain, functional area, or on some other basis - thereby separating accountability to specific project teams.

Eventually, most organizations will reach the point where deployment falls into the upper right quadrant: there are both environmental and DEEPER complexities in deployed predictive analytics solutions. In this scenario, governance is essential because of the model deployment risks that could, if not mitigated, result in mission critical failures. Moreover, there is greater risk that multiple predictive model interactions could result in unintended consequences simply due to the complexities and subtle inter-dependencies that can manifest. This is depicted in Figure 2 as a situation where organizations must be prepared with the necessary emergency task force to manage situations when things go awry in order to reduce chaos with minimal damage to the organization.

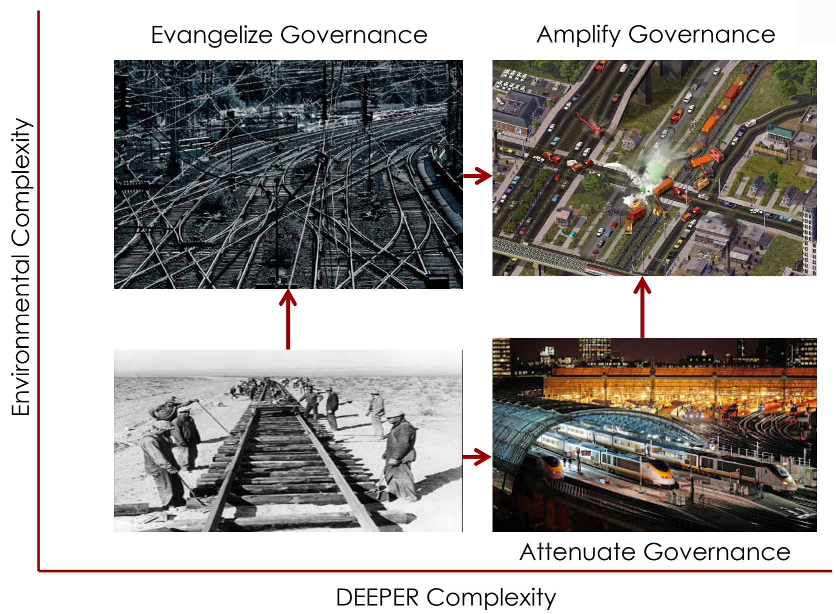


Figure 2: Environmental and DEEPER Complexities

The arrows in Figure 2 depict typical pathways as organizations ramp up predictive analytics deployment. Each quadrant gives rise to a different approach to governing a firm's analytics capabilities. For example, in the initial stage, the lower left quadrant, a firm will encourage well-designed pilot studies and begin to design data management policies to support predictive analytics initiatives. Business experimentation is encouraged, new analytics training is sponsored, and there is experimentation with a variety of different predictive modeling toolsets. Organization leaders will monitor the pilot projects to assess the current workforce for those capable and willing to take the lead on future analytics projects. In many cases, job descriptions and titles will be designed in anticipation of both assigning current employees to new positions and hiring new employees with requisite analytics skillsets. Data quality will be a key concern for data originating from the organization's key application systems.

An organization that follows a transition path to high environmental complexity but low DEEPER complexity will likely opt to evangelize governance. Such organizations will begin advancing learned best practices. Close collaborations between data scientists, business domain experts, and business leadership will emerge. Project management protocols will be realized and there will be an understanding of the best ways to investigate and present analytics solutions' returns on investment to executive leaders. As infrastructure requirements begin to be built-out, there will be increased pressure for dealing with ways to best allocate and share costs. As an organization moves from "low hanging fruit" types of projects to riskier ones, there will likely be some failures. Failing fast will be key, so that other initiatives can get underway. Analytics that assist to identify new lines of business or improve business processes that are no longer achieving high performance will be key.

For some organizations, DEEPER complexity will be high even while environmental complexity remains low. In these cases, there will likely be solutions that require significant cross-functional collaboration and coordination. A key here is to attenuate governance by taking a divide-and-conquer approach to governance protocols (Figure 2). Each project will likely have different governance requirements due to its intricacies and the nature of the stakeholders involved. Like the picture of the trains running efficiently through their respective part of the station in the lower right quadrant of Figure 2, each project type can have its own governance protocols that serve to steer those projects to fruition. Some centralized governance role might still apply. For example, organizations may wish to establish a predictive model asset inventory with metadata like model authorship details and assumptions that went into the model. Such steps can help mitigate issues arising when data science team members leave the organization. Projects should be assessed for risk according to some general framework, but project leadership can be fragmented. Regular reporting of model performance will be important, and process audits may be needed to assess how well DEEPER stages are being accomplished.

Amplifying Governance

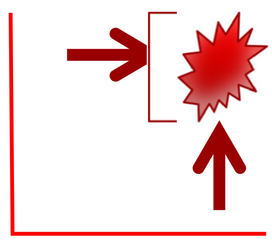
As shown in the pathways of Figure 2, most organizations investing in analytics move into the upper right amplify governance at some point on their analytics journey because successful analytics solutions lead to more complex projects that are higher risk. Given the high environmental and DEEPER complexity, organizations need concerted predictive analytics governance at that point. That governance should include arrangements for making decisions over a multiplicity of stakeholders, projects, SEMMA and DEEPER processes, and work products and resources that involve continuous negotiation, deliberation, deployment, and monitoring. Although there isn't a well-researched set of governance best practices for these scenarios, we can infer beneficial approaches from research on current company practices.

Companies have taken a variety of steps to amplify governance, such as repeated back-testing of predictive models. In other words, they select a model for periodic re-evaluation, retest it against new data and assumptions and, in cases of test failure, they either switch to a challenger model or build a new model portfolio. Some organizations continuously monitor high-risk model performance and often institute a real-time data governance program in addition to their more traditional data governance approach. Some install business process alerts, prepare manual back-up plans, and actively enforce model building and metadata collection standards during the SEMMA stages. There may be frequent and independent deployment audits and new policies are created as experience is gained. More mature organizations are often those in the financial industry, where there are legal compliance issues for deployed predictive models.

There has been significant controversy in the press about predictive analytics governance. *Information Age* wrote that analytics is transforming IT departments into true business partners. *ClickZ* poses the question, "Will this be the year of data analytics in the marketing department? *Edmunds Recruiting* claims there are must have traits of a data analytics department. *Strategy+Business* addresses the question, "Who should own big data? IT, the business or a matrix?" The *Wall Street Journal* ran an article titled, "Why large firms don't need an analytics department." *Forbes* asserts to CIOs and CTOs, "Don't let a chief digital officer steal the best part of your job." *Harvard Business Review* has chimed in with, "Should your CIO be chief data officer?" These articles raise questions about how analytics fits into organizational structures, whether IT should manage the analytics capability, and whether firms even need a central analytics unit. Similar to R&D units engaged in innovation initiatives, analytics projects can change an organization's strategy, alter historical decision making assumptions, and can drive an entirely new evidence-based organizational culture.

At the same time, there is increasing consensus on the business value of analytics governance beyond risk mitigation. Governance allows analytics to be explicitly linked to business performance. And once governance processes are stable, faster predictive analytics deployment is possible, thereby enhancing good decision making across the organization.

Some instances (Figure 3) where poor model governance has created negative business consequences include the "Flash Crash" of 2010. Five years after the crash, regulators blamed a single trader who implemented spoofing algorithms that placed orders the trader planned to cancel later on. This type of algorithm is now banned by the SEC, but some also blamed other analytical algorithms that had been deployed by other traders. High frequency traders deployed algorithms that responded to volatile swings by taking actions of their own. Basically, an unintended consequence of the high frequency trading algorithms was to exacerbate the impact of the spoofing algorithms. The market experienced a swing of over 1000 points on the Dow Jones Industrial Average that day. Millions lost significant value in their investment portfolios and 401Ks on that day.



- ***Flash Crash***
- ***Target***
- ***Real estate***
  'Collateralized Debt Obligations" (CDO's) ratings by Moody's and S&P decayed

Figure 3: Cases Where Amplifies Governance Was Needed

Target launched an ad campaign leveraging buying habits to assess the likelihood of a customer being pregnant. In one instance, the company mailed out coupons based on the prediction, causing significant discord because family members had not yet had an opportunity to discuss the teenager's pregnancy. This situation highlights the many nuances to predictive analytics deployment beyond just infusing a model into a business application. The deployment should be considered in terms of the customer segments targeted and the value of the model vs. the negative business consequences that might arise. Many organizations rush into analytics solutions without in-depth understanding of DEEPER processes.

Many blame the recent Great Recession on the decay of the predictive models used by Moody's and Standard & Poors that rated the risk of subprime mortgage collateralized debt obligations (CDOs), which are notoriously difficult to assess for risk. When the housing market began to collapse, the risk increased rapidly and dramatically. Before the crash, CDOs returned more than bonds, but without understanding the true risk, it was difficult for investors to assess the upcoming market volatility. Most historians and governmental reviews that analyzed the causes of the Great Recession concluded there was plenty of blame to go around. Investors, regulators, governance practices, and many other stakeholders and policies were partially to blame. But predictive model decay, at the most prestigious of risk assessment entities, was among the causal factors.

Previous research identified three options for predictive analytics governance: centralize analytics; assign data scientists to business units; or develop some hybrid model, such as establish a center of excellence and also assign data scientists to business units. Whichever option is taken, a thorough approach includes: addressing people, structure and reward systems; establishing accountability, transparency and traceability to those who are investing in analytics projects; and creating a method for allocating costs to those parts of the organization that derive the most value from analytics resources. Governance should also include identifying and funding analytics opportunities; providing support, infrastructure, and other resources for obtaining the data, and providing guidelines for deployment. Four committees are suggested to handle governance roles.

First, there should be a team addressing analytics data management and quality. This would likely be an extension of data governance teams that already exist in many organizations. Second, there should be a team charged with analytics security and compliance issues. Third, there should be a technical team that develops technical policy and helps to guide analytics and big data architecture. The last team should be an overarching group of stakeholders that oversees the other three teams and analytics governance in general.

Following are some current examples of best practices for predictive analytics governance.

Dell Analytics Platform Management

Dell has amplified governance through a three-tiered approach to its comprehensive analytics environment. Each tier has management ownership over content and governance. IT owns the content and governs all aspects of the production tier. There are strict standards of compliance to data governance policies. IT provides SLAs and operational support for this tier, which contains the mission critical applications that keep the enterprise running and the lights turned on. The semi-production tier is also owned by IT. It is for analytics solutions that have passed proof-of-concept muster and are ready to be institutionalized and infused into business processes and applications. This is where new analytics innovations are hardened, their maturity is enhanced, and they are become standardized, replicable and stable across the enterprise. Significant testing is required to ensure that performance times meet business requirements. IT provisions, automates, monitors and recommends optimizations in this tier. The third tier, the sandbox, is owned by the business. It provides an area for data exploration, discovery, and what-if analysis. IT provides infrastructure, tools support, and monitoring. In some cases, sandbox tier workspaces can be created by business using self-service capabilities.

This three tier approach guarantees that new analytics discovery is driven by the business, that solutions that have the most promise go through a period of hardening and preparation for production, and that the production tier is hands-off to data scientists. Specific processes and protocols for moving discoveries from the sandbox to the production tier guarantee a full vetting of potential solutions before they are deployed.

Turkcell Model Management

Turkcell considered innovative governance approaches for predictive analytics and business intelligence when it experienced challenges in marketing campaign planning and execution. The organization felt it was launching too many campaigns that had low response rates. To deal with this, it introduced model management with organization-wide, customer contact policies and integrity rules. It conceptualized their operation as a model factory, identifying approximately 50 critical models that would be closely scrutinized and monitored. It established performance measurement rules for the models and then set key performance indicators (KPIs) for accuracy. Triggers were embedded in campaign applications to alert governance teams when KPIs went out of bounds, and it designed regular KPI reporting routines. The alerts were implemented in a tracking dashboard and the reports detailed end-to-end model effectiveness and performance of model hierarchies. Actions were decided in advance for handling the various types of alerts, including decision rules for model switching and retargeting.

Adidas Shift to Real-Time Marketing

Adidas launched an ambitious consumer DNA (CDNA) project to capture transaction and interaction data on millions of its customers. Data came from both sales systems and web analytics. The project's goals were to provide the right information at the right time to the right customers, and to select the right target for the right offer at the right time. Data scientists described customer metadata as a consumer DNA protein base. Complex customer analytics records contain information on a customer's preferred communication time, the communication lifecycle preferred, the position of the marketing calendar, and whether the customer had a local vs. global campaign relevant to their particular context. Since pilots of the approach proved successful, CDNA project leaders planned to seamlessly integrate the analytics solution into the CRM infrastructure using in-database capabilities. Prior to integration, the CDNA project was conducted using an independent campaign management platform. To integrate and automate the analytics solution, project leaders needed to collaborate with IT. Governance enabled coordination and cooperation in the development of analytics projects and in their ultimate deployment.

American Express Data/Analytics/Business Strategy Symphony

Similar to Adidas' customer strategy, American Express explored alternatives to presuming that a single interaction with a customer provides sufficient information for a successful marketing effort. The company sought to identify the customer conversations that truly matter. Data scientists felt confident they had the technological means to guide, assess, and dynamically adapt customer conversations. But there were questions as to when a conversation started or ended, as well as the exact point at which an interaction had the most impact in driving the conversation. American Express project team leaders concluded they needed to align business strategy (e.g., the customer engagement strategy), data strategy (e.g., the speed of data collection, storage, aggregation, etc.), and analytics strategy (i.e., the measurement approach and methods for interactive interaction optimization assessments). Project team leaders described their approach to governance as an "instrument that creates a symphony out of any discordance of CMOs and CIOs."

State Farm Testing Culture

Experiments provided the impetus for cross-functional collaboration on American Express' marketing analytics project. Similarly, State Farm established a testing culture in which analytic solutions are assessed to determine whether they are performing as designed and promoted. A testing team comprised of a business partner, a statistical designer, an execution specialist, and a structural and creative design expert treats the business environment like a controlled experimental setting, randomly assigning strategies to customers. Using A/B testing and multivariate analysis, testing teams work across the business where predictive analytics are in use, often infused into applications. The team maintains a test results repository that provides documentation of key hypotheses, findings, participants and actions taken. Testing teams are governed by both strategic and testing council level governance committees.

SAS Consulting Big Data Issues

SAS Consulting dealt with an interesting big data problem. The question they sought to answer is, "What is the probability that a uuid (customer with cookie stored by a browser) with certain traits (captured characteristic of a website visit) belongs to a particular age group?" To address the problem, raw data needed to be converted to one row per subject with one column for each trait. The sheer size of the exploded data would require approximately 885 gigabytes of memory. The SAS Consulting team used visualization and business rules to reduce data in a first step pass. A second step pass performed summarizations and transposition procedures to get to a manageable data set to research predictive analytics solutions. There were two important governance lessons in the project. First, metadata on big data projects must include the reductions, transformations, and business rules used for reduction. In addition, visualizations are an important part of the documentation of the modeling effort. If predictive models decay, it will be important to return to this documentation to avoid a repetition of past efforts. Second, model decay testing and model switching on big data requires periodic recalibration. This is because of the difficulty of capturing, as part of the model development effort, the ranges for variables used in the models and boundary conditions for parameters. The complexity occurs because the raw data undergoes significant transformations and data reductions. Model developers should be consulted when choosing the time period after which recalibration should be performed.

In sum, the approaches described for Dell, Turkcell, Adidas, American Express, State Farm, and SAS Consulting demonstrate an analytics journey that is rapidly moving to the upper right quadrant of the model in Figure 2.

Emerging Levers for Predictive Analytics Governance

Governance model design choices take into consideration such organizational levers as organizational structure, resources, models, and DEEPER processes. The levers make some assumptions about stakeholders, governance evolution, and the likelihood of analytics governance centralization in the highest risk scenarios. Stakeholders will continue to be independent to take advantage of domain focus and expertise while being interdependent at different levels of responsibility and authority. Data science, data management, and domain experts are often required to work together on predictive analytics and business intelligence projects. Domain experts may come from any part of an organization and data science experts may be centralized or assigned to domain areas. Data management experts may be centralized in the IT group or may work with respective data science experts. For purposes of governance, these stakeholders may do parts of their jobs independently, but will be interdependent for many projects. The American Express symphony case is consistent with these assumptions.

Governance will evolve within an organization as wins create spillovers to be exploited and learning is institutionalized. Unintended consequences might be avoided on a particular project and those resolution strategies should be institutionalized. And as competition adapts to industry advantage gleaned from analytics investments, then new analytics innovations will drive new advantage. For this reason, the environmental complexity and DEEPER complexity are likely to increase. Data sources from sensors, data markets, and the government will likely increase over time as will the complexity of those data assets.

Currently the most effective organizational governance levers fall into four broad layers: organizational structure, resource management, model management, and DEEPER processes for amplifying governance. These layers are depicted in Figure 4.

Organizational Structure
Options include centralizing analytics, decentralizing analytics, or creating some matrix structure by combining an analytics center of excellence with the assignment of analytics capabilities to different functional areas (e.g., business units, functional business areas, or product lines). Governance decisions regarding organizational structure directly affect decisions made at other layers.

Resource Management
Resource strategies can be aligned with organizational structure decisions. For example, a hands-off batch approach might be appropriate for projects that can be completed by teams existing within a decentralized structure, such as a marketing group using an analytics appliance whereby data are downloaded from central IT on an ad hoc basis to support marketing campaigns. Adidas took this approach before it chose to move to real-time offers. A one-off projects approach might be appropriate for a company with a single dedicated data science team that conducts one project at a time. A divide-and-conquer approach might be appropriate when the infrastructure is segregated into different areas of governance and management. Dell follows this approach.

Model management
The management of models can be centralized or decentralized. Turkcell takes a centralized approach to model governance. Decentralized model management may be better suited to contexts

where organizational structure is decentralized and predictive analytics by product line or functional area are sufficiently different that a unique model management strategy is needed for each.

DEEPER Process Management

American Express designs synchronized business, data, and analytics strategies for marketing conversation analytics. Adidas embeds CDNA into analytics solutions in its CRM infrastructure using in-database capabilities. American Express manages performance by ensuring that infrastructure costs are a part of model deployment cost/benefit analysis. SAS Consulting performance management involves maintaining metadata on big data projects, including reductions, transformations, and business rules used in reduction. Visualizations are also included as part of the documentation of modeling efforts. State Farm evaluates its models as a regular part of its testing culture. SAS Consulting retargets its models once model decay is detected through periodic recalibration.

Our predictions for the 2020s include:

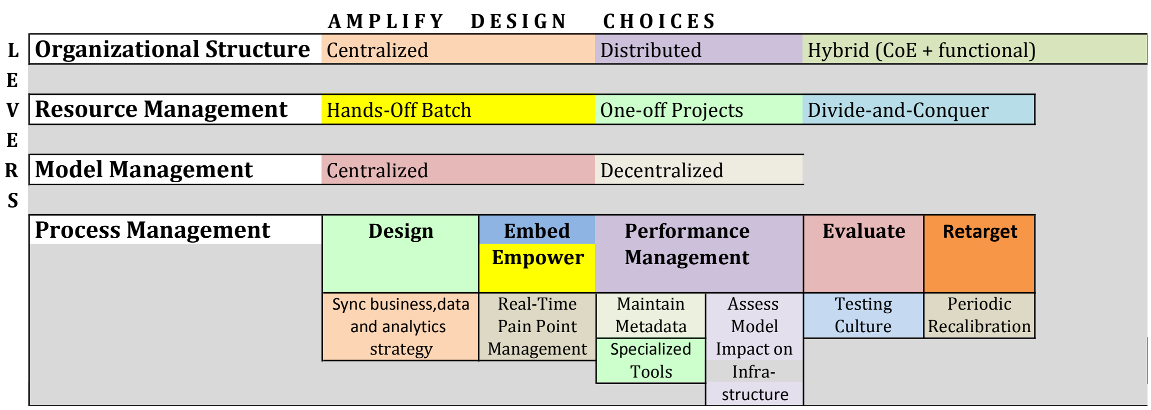


Figure 4: Multi-Layered Predictive Analytics and Business Intelligence Governance Model

Table 1 summarizes the various levers by company.

Company/Source	*Relevant Levers and Design Choices*	*Relevant DEEPER Process*
American Express	Sync business, data, and analytics strategy	Design
	One-off projects	
	Model management decentralized	
	Assess model impact on infrastructure	Performance management
	Center of excellence	
Dell	Resource management is divide-and-conquer	
Adidas	Hands-off batch - migrating to real time	Embed, empower
SAS Consulting	Periodic recalibration	Retarget
	Maintain metadata	Performance management
Turkcell	Centralized model management	
State Farm	Testing culture	Evaluation

Table 1: Summary of Levers by Source and DEEPER Process

When considering choices at each layer, it is useful to explore how different combinations might align an organization's analytics strategy with its overall analytics risk profile. For example, Figure 5 depicts a lower risk, low analytics volume profile where structure is centralized, resource management is managed in one-off fashion, model management is decentralized and two amplify process options are selected: establishment of a testing culture and synchronization of business, data and analytics strategies.

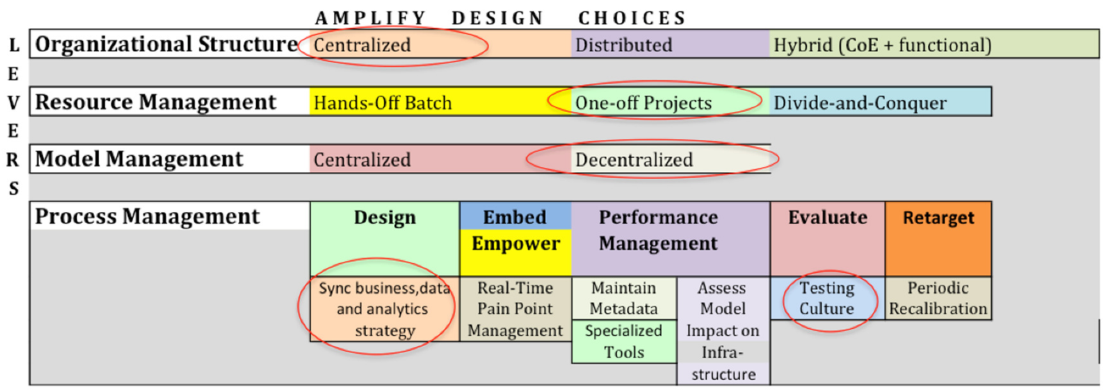


Figure 5: Exemplary Lower Risk, Low Analytics Volume Profile

A higher risk, high analytics volume profile is depicted in Figure 6. Here, the organizational structure chosen is hybrid, resources are managed in a divide-and-conquer mode, model management is centralized and many of the amplify process options have been selected.

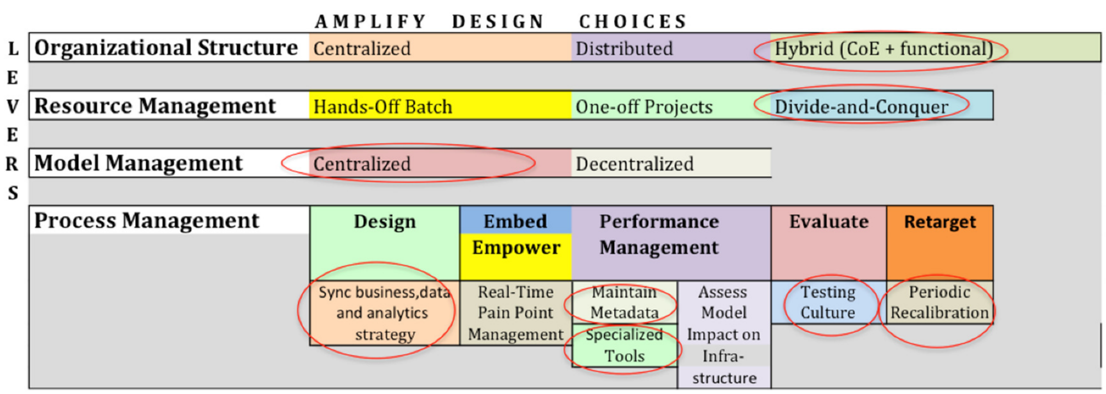


Figure 6: Exemplary High Risk, High Analytics Volume Profile

Conclusion

Most organizations have or plan for data science teams to parlay data assets into business intelligence, predictive analytics, and prescriptive analytics solutions. Such solutions will inevitably be embedded into core applications and business processes essential to running the enterprise because faster speed-to-decision can create a competitive advantage. Effective organizational governance of such infused analytics solutions is necessary to mitigate the risk associated with analytics that may go awry, resulting in unintended consequences. Governance also avoids losing critical information on assumptions underlying analytics solutions when data scientists leave the organization or when environmental changes render models obsolete or invalid.

No organization has found the best possible governance approach because we are in a pre-paradigm state of knowledge. However, grass roots approaches used in organizations today, synthesized with the some research in the area, provide concrete approaches that can be leveraged by any company at any point on its analytics and business intelligence journey. We have summarized current best practices in a framework that clarifies the levers and design choices available today, helping readers learn how to design a state-of-the art governance approach. Further, there are integrated design choices that can help an organization to customize its analytics governance approach. Structural, resource, and model management issues as well as a host of best industry processes have been presented.

We know from experience that there will be analytics failures along the way, and while accountability, transparency, and traceability are import elements of a governance program, they should not be intended to serve as a rationale for doling out poor model performance punishment. Models decay – some faster than others. Assumptions made in their development may not have all been valid, but they may well have represented how much an organization understood about its business at the time. The following quote provides an important context for the long term role of predictive analytics governance as well as a complement to all of the risk mitigation strategies discussed previously:

> "Ironically, the greatest value from predictive analytics typically comes more from their unexpected failures than their anticipated success. In other words, the real influence and insight come from learning exactly how and why your predictions failed. Why? Because it means the assumptions, the data, the model and/or the analyses were wrong in some meaningfully measurable way. The problem—and pathology—is that too many organizations don't know how to learn from analytic failure. They desperately want to make the prediction better instead of better understanding the real business challenges their predictive analytics address. Prediction foolishly becomes the desired destination instead of the introspective journey." [1]

[1] Schrage, Michael, "Learning from your Analytics Failures," *Harvard Business Review*, September 3, 2014.

Recommendations

We suggest that organizations with either no analytics governance or a governance approach that needs reevaluation, begin by examining how framework option decisions have been or are currently being made. How well is coordination taking place and how strong is cooperation? Where are the breakdowns and what causes them? What is covered in current governance plans? How well have design choices been aligned with analytics strategies and risk mitigation profiles?

Understanding the organization's location on its analytics journey in terms of governance complexity can provide a context for why governance decisions made yesterday may no longer be valid for today or tomorrow. Since governance should and will evolve, it is important for IT, analytics, and data management leaders to embrace change leadership. They must ensure that operations exhibit accountability, transparency, and traceability to those who fund analytics projects, those who develop and support analytics resources, and those who leverage analytics resources.

About the Author

Michael Goul

Michael Goul was appointed in July of 2015 to serve as Associate Dean for Research at the W. P. Carey School of Business at Arizona State University. Among other duties, he works with the School's portfolio of research centers, and he is the lead on a cross-university big data and analytics research collaboration initiated by W. P. Carey School Dean, Amy Hillman. For the six years prior, he served as chair of the school's department of information systems. *U.S. News and World Report* consistently ranks the Carey school's programs among the best in the nation. Goul spearheaded the development of the nine-month Master of Science in Business Analytics program that is a collaboration of the School's information systems and supply chain management departments. In addition, he administered the launch of the School's undergraduate Bachelor of Science in Business Data Analytics degree. That program began in Fall, 2014. Michael also administered the launch of the online version of Carey's highly successful Master's of Science in Information Management program. *U.S. News and World Report* recently included that program in its ranking of Carey's online graduate offerings as 2nd best in the nation. Carey's graduate program's information systems specialty was ranked 12th, and the undergraduate program was ranked 18th. Goul is passionate about how the concomitant explosion of big data, the shift to cloud computing and the emergence of the mobile/social web does and will impact the global economy. He has published over one hundred articles and he has authored cases and conducted analytics research at companies including American Express, eBay, Teradata and Intel. For the 2005/6 academic year, Dr. Goul was appointed William J. Clinton Distinguished Fellow by the University of Arkansas Clinton School of Public Service, Little Rock, AR. His Clinton School research focused on open data, cross-agency data sharing and services computing.

References

Balkan, S. and M. Goul, (2010) “Advances in Predictive Modeling: How In-Database Analytics will Evolve to Change the Game,” *Business Intelligence Journal*, 15:2, 17-25.

Balkan, S. and M. Goul, (2011) “A Portfolio Theoretic Approach to Administering Advanced Analytics: The Case of Campaign Management,” *International Conference on Systems Sciences*, 1-10.

Cagatay, T. (2014) “The Journey to Multimodel Environment: A Turkcell Story from Creation to Measurement, “ *SAS Analytics 2014.*

Cailloux, R., (2014) “Consumer Analytics Faster, Higher, Stronger: The Consumer DNA Factory,” *SAS Analytics 2014*.

Casey, T., K. Krishnamurthy and B. Abezgauz, “Who Should Own Big Data?,” *Strategy+Business*, August 12, 2013.

Davenport, Tom, “Why Large firms Don’t Need ‘Big Data Departments,” *Wall Street Journal,* May 16, 2013.

Finta, A. (2014) “Dell Stores Data Securely and gives Business Users Fast and Easy Access,” *Teradata Partners 2014*.

“Five Must-Have Traits of an Analytics Department,” *Edmunds Recruiting*, retrieved Oct. 22, 2015 from http://www.careerealism.com/data-analytics-department-successful-traits/

Goul, M., S. Balkan and D. Dolk, (2015) “Predictive Analytics-Driven Campaign Management Support Systems,” *International Conference on System Sciences*, 4782-4791.

Grossman, R. L. and K. P. Siegel, (2014) “Organizational Models for Big Data and Analytics,” *Journal of Organizational Design*, 2:1, 20-15.

Lara, C., (2014) “What is Relevant? The Question When Working With Big Data,” *SAS Analytics 2014*.

Logi Analytics, “2015 State of Embedded Analytics Report,” retrieved October 22, 2015 from http://learn.logianalytics.com/2015-SOEA

Pulkstenis, A. (2014) “Do you Know or Do You think You Know? Building a Testing Culture at State Farm,” *SAS Analytics 2014*.

Rossi, Ben, “Analytics – Transforming IT departments into true business partners,” *Information Age*, Jan. 27, 2015.

Schrage, Michael, “Learning From Your Analytics Failures,” *Harvard Business Review*, September 3, 2014.

SEMMA, retrieved October 22, 2015 from https://en.wikipedia.org/wiki/SEMMA

Starr, L. and P. van Ispelen, (2014) “CRM: Balancing the Speed, Relevance and Frequency of Your Customer Conversation,” *SAS Analytics 2014*.

Thorogood, Pelin, “Will This Be the Year of Data Analytics in the Marketing department?” *ClickZ*, Jan. 15, 2014.

Westerman, George, “Should Your CIO Be Chief Digital Officer?,” *Harvard Business Review,* August 2, 2013.

Woods, Dan, “Don’t Let a Chief Digital Officer Steal the Best Part of your Job,” *Forbes*, April 14, 2014.

APC Contact Information

Program Director
Madeline Weiss, Ph.D.
301.229.8062
Madeline.Weiss@simnet.org

Research Director
Richard Watson, Ph.D.
706.542.3706
rwatson@terry.uga.edu

SIM Headquarters Staff
Vincenzo Nelli
856.380.6827
vnelli@simnet.org

