ORGANIZATIONAL MODELS FOR BIG DATA AND ANALYTICS

ROBERT L. GROSSMAN • KEVIN P. SIEGEL

Abstract: In this article, we introduce a framework for determining how analytics capability should be distributed within an organization. Our framework stresses the importance of building a critical mass of analytics staff, centralizing or decentralizing the analytics staff to support business processes, and establishing an analytics governance structure to ensure that analytics processes are supported by the organization as a whole.

Keywords: Organizational structures for analytics, big data, analytic governance, organizing data scientists

There is little debate these days about the importance of big data and analytics in supporting the strategic goals of an organization (Davenport, 2006; Manyika, et al., 2011), but there is as yet no consensus about how best to organize analytics efforts within the organization and what core analytics processes the organization must support. In this article, we introduce a framework that breaks big data and analytics into several processes and shows how those processes fit within the organization, and we discuss how an appropriate analytics governance structure can enable an organization to extract business value and competitive advantage from big data.

Following Laney (2001), we consider *big data* as data whose volume, velocity, and variety make it difficult for an organization to manage, analyze, and extract value using current or conventional methods and systems. We use the term *analytics* as the process that extracts value from data through creating and distributing reports, building and deploying statistical and data-mining models, exploring and visualizing data, sense-making, and other related techniques. Data may be internal or external to the organization; processing may be real-time, near real-time, or batch; and any combination of these is possible.

A FRAMEWORK FOR ORGANIZING ANALYTICS

Our organizational framework seeks to integrate analytics, business knowledge, and information technology (see Figure 1), and it is based on four main questions:

1. Does the organization view data and analytics as a key function of the organization, similar to the way that finance, information technology, sales and marketing, product development, etc. are viewed as functions of the organization? Analytics must be perceived as valuable to the business units in order for it to be integrated into operations.
2. Is there a critical mass of data scientists? Without a critical mass of data scientists, there is insufficient domain knowledge to address all the problems of interest. Also, there is not deep enough knowledge of the analytics infrastructure to obtain or create the needed data and to manage the data that is obtained. Finally, there may not be deep enough knowledge to deploy statistical and data-mining models in operations.
3. Are there data scientists with sufficiently deep knowledge of the business unit domains? Without such knowledge, it is difficult to build models that bring value to the business unit. Deep knowledge and complex business problems tend to spawn specialization. It is important for an analytics group to include a mixture of data scientists, some of whom are generalists and others who are specialists.

Journal of Organization Design
JOD, 3(1): 20-25 (2014)
DOI: 10.7146/jod.9799

4. Is there an adequate analytics governance structure? A governance structure helps stakeholders make decisions that prioritize big data opportunities, obtain the required data, deploy analytical models, and support measurement of the business impact of the models.

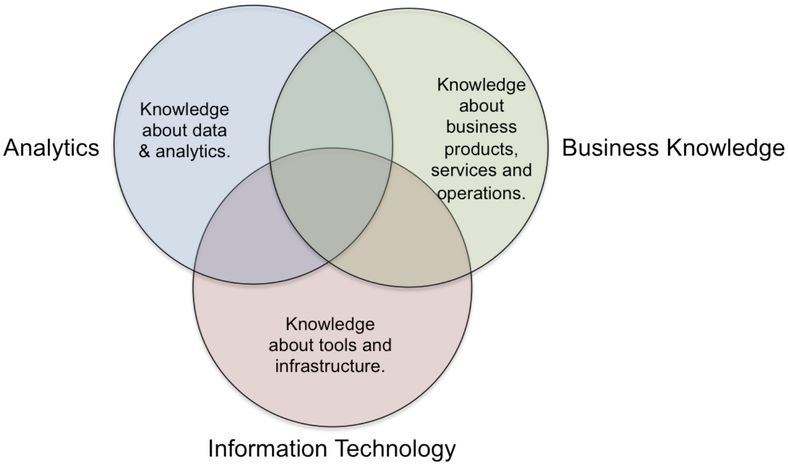
Analytics
Knowledge about data & analytics.
Knowledge about business products, services and operations.
Business Knowledge
Knowledge about tools and infrastructure.
Information Technology

Fig. 1. The knowledge required by data scientists

We call our framework the CSPG framework – for analytics Culture, Staffing, Processes, and Governance (see Table 1). The CSPG framework orients the organization designer to establishing a culture for big data and analytics; hiring, training, and organizing a group of analytics staff; developing the required analytics processes; and setting up a robust analytics governance structure. Starting with culture, corporate-level executives must recognize the need to organize big data and analytics as an organizational function that is given broad responsibility and authority for data assets and which is analogous to other major functions in the organization. The analytics leader has the responsibility for hiring and managing the best data scientists, ensuring that the appropriate analytics opportunities are identified and explored, acquiring the appropriate internal and external data, and setting up and operating the analytics governance structure.

Table 1. A summary of the CSPG Framework

	Analytics at the Department/ Unit Level	**Analytics at the Organizational Level**
Analytics Culture	Are big data and analytics viewed as an organizational function and is there a big data/analytics department or unit to support this function?	Are big data and analytics integrated into corporate strategy? Is there a senior leader advocating for big data and analytics? If not, put a senior leader in charge of big data and analytics with this charge. Is data (both internal and external) that can provide value being used?
Analytics Staff	Does the analytics department have the right people with the right degree of analytic specialization, IT knowledge, and business knowledge?	Are there analytic team members in the right departments within the organization and is there a critical mass of analytic talent? If not, rebalance the analytic staff or change the centralization/decentralization of the analytics staff as required.
Analytics Processes	Does the analytic department have analytic processes in place to build analytic models, deploy analytic models, and measure their business impact?	Does the organization have the analytic processes in place to select analytic opportunities, provide data to the data scientists, build analytic models, deploy analytic models, and measure the business value generated? Is there an analytic governance structure in place to support and to coordinate the correct analytic processes?

The CSPG framework requires that there be a critical mass of analytics staff (data scientists). Analytics staff must be able to obtain and manage data; build statistical, predictive, and data-mining models; and deploy those models. The analytics leader, along with corporate management, must decide on where to locate the analytics function within the organization (discussed in the next section). Essentially, the analytics staff can be centralized or decentralized, with hybrid approaches available as well.

The third component of the CSPG framework concerns the analytics processes themselves. Big data presents many opportunities if those processes can be properly created and managed. Data can be traded among organizations, products can be augmented to produce data, assets can be digitized, data can be combined within and across industries, and so on (Parmar, Cohn, & Marshall, 2014). The more sophisticated the analytics processes become, the more opportunities that can be pursued. The organizational aspects of analytics processes are discussed below.

The final component of the CSPG framework is analytics governance. Because big data and analytics are new to many organizations, analytics governance structures are not well defined. Senior corporate leaders are responsible for setting up the governance structure, and they are responsible for monitoring and improving it as experience accumulates. Analytics governance structure is discussed below.

Broadly speaking, the CSPG framework presented here can be thought of as an application of the Star Model (Galbraith, 2008) to the analytics function. The design of the analytics function must be complete in the sense that it covers people, structure, rewards, and so on, and each component of the analytics function must be aligned with the others and with the larger corporate organization.

LOCATION OF THE ANALYTICS FUNCTION WITHIN THE ORGANIZATION

There are three basic models for locating the analytics function within the organization, all of which involve well-known tradeoffs between centralization and decentralization. One model centralizes analytics by placing the data scientists in a single unit. This model is the easiest way to achieve critical mass, obtain necessary data, drive an integrated infrastructure, and gain the required expertise to efficiently test and deploy various statistical, predictive, and data-mining models. When analytics is centralized, however, the data scientists may be far away from the business units they are supposed to support. The challenge in such a structure is for the data scientists to understand the various business units and their needs. In addition, there is the issue of where the analytics department should report within the organization. Should it report to a functional area such as finance, IT, R&D, or marketing, or should it report to the very top of the organization?

A second organizational approach is to decentralize analytics and place a group of data scientists in each business unit. This approach makes it easier for data scientists to collaborate with their respective business units and to tailor their models to each unit's needs. The main tradeoff is difficulty in achieving critical mass on enterprise-wide problems and opportunities. A closely related question is whether each group has the expertise required to create datasets and to deploy analytical models.

The third model is a hybrid approach in which a critical mass of data scientists is housed in a central unit, and the remaining data scientists are distributed throughout the organization. One common hybrid model is to set up an analytics or big data "center of excellence" that the distributed data scientists can draw on as appropriate. Another is to centralize the data scientists that interact with the IT organization, or those that manage the data, or those that deploy the models.

None of these three models provides a perfect organizational solution; each involves tradeoffs. From a design perspective, managers must recognize the tradeoffs associated with each model and make their location choice accordingly.

ANALYTICS PROCESSES

Generally speaking, the analytics function is composed of analytics models, analytics infrastructure, and analytics operations. Analytics models are statistical, predictive, or data-mining models that are empirically derived from data using generally accepted statistical methods. A key analytics process is building models. This is usually done by statisticians, modelers, or, to use the new name, data scientists (Press, 2013). As discussed above, data scientists may be located within a single department or group, attached to business units, or a combination of both. If the data scientists are centralized in a single unit, it is often called the analytics department. In addition to building models over data, analytics also includes summarizing data in reports (now called descriptive analytics), ad hoc querying of data, exploring data with visualizations, sense-making, and other techniques.

Analytics infrastructure refers to the software components, software services, applications, and platforms for managing data, processing data, producing models, and using models to generate alerts, take actions, and make decisions (Grossman, 2009). The key processes associated with analytics infrastructure are managing the data required by the organization and deploying the models and other analytics that are incorporated into the organization's products, services, and operations. It is becoming more common to use computer languages for describing analytics (Data Mining Group, 2012) so that analytics can be more easily deployed. These processes involving analytics infrastructure are usually performed by the information technology (IT) organization.

Analytics operations refers to the various processes that result in the outputs of models being used to make decisions and to take actions that bring business value. Analytics operations ensures that the results of models are integrated into an organization's products, services, and operations. In an analytics department, data scientists identify the data needed, acquire the data, work collaboratively with business units to build models, and then work with the IT group to deploy the models into the organization's operations. Data can be a combination of data internal to the organization, collected by the organization, or purchased by the organization. The IT department is normally involved if data is generated by the organization or collected by the organization.

An organization requires a critical mass of data scientists so that their expertise as a whole extends across these three analytics processes. The team as a whole must be able to: identify relevant data (both internal and external), manage the data required for analytics, build the needed analytics models, and deploy the models that are built into products, services, and internal systems. Multiple parts of an organization can be involved in analytics processes. Typically, a business unit sponsors the model, an analytics department builds the model, an IT unit supplies and manages the data, and an operations unit deploys the model. With so many diverse pieces of an organization involved, an analytics governance structure is critical.

ANALYTICS GOVERNANCE STRUCTURE

There are three main challenges that organizations face when trying to extract value from big data using analytics.

1. *Identifying and resourcing analytics opportunities.* The first challenge is identifying which analytics opportunities to pursue, building the business case for those opportunities, and obtaining the required resources. Analytics opportunities belong to stakeholders within the various business units and functional areas of the organization. Opportunities can also originate outside the organization and must somehow be identified and subjected to the analytics process.
2. *Obtaining the data.* The second challenge is to obtain all of the necessary data in a consistent and timely fashion. It is usually difficult for most modeling groups to obtain the data they require in a timely fashion unless they have their own datamart, data warehouse, or distributed data processing system (White, 2009). In most organizations, the IT group controls access to the data.
3. *Deploying the models.* The third challenge is to deploy the models into operations or production systems in a consistent and timely fashion. Deployment challenges can directly impact analytics' efficacy. In most organizations, the IT group controls how models are deployed into products, services, and operations.

These are challenges for most organizations since the modeling group must work with other components of the organization to identify analytics opportunities, obtain the necessary data, and deploy the resulting models. The role of an analytics governance structure is to put in place an individual (the analytics leader) with sufficient authority to overcome these three challenges. An analytics governance structure must also include mechanisms for identifying, communicating, and resolving issues that are holding up analytics projects. Lastly, the governance structure requires a mechanism for providing sufficient resources for analytics projects and for balancing priorities between analytics projects and other corporate projects. At this stage of evolution of analytics governance structures, a complete set of parameters for designing a governance structure does not exist. We suggest the following preliminary parameters:

1. Ensure that sound long-term decisions about analytics are reached and that investments in analytics generate business value.
2. Operate in such a way that data, derived data, and analytics products are protected and managed in a secure and compliant fashion.
3. Operate in such a way as to make sure that there is accountability, transparency, and traceability to those who are funding analytics projects, to those who are developing and supporting analytics resources, and to those who are making use of analytics resources.
4. Provide an organization structure to ensure that the necessary analytics resources are available; that data is available to those developing analytics; that analytics can be deployed; that the impact of analytics is quantified and tracked; and that data, derived data, and data products are managed in a secure and compliant fashion.

These design parameters can be achieved by using governance committees:

- An analytics governance committee that includes senior management and representatives from the IT organization and various business stakeholders. This committee helps prioritize analytics opportunities; obtain resources for analytics projects; and ensure that those building the models get the data required, that the models that are built get deployed, and that deployed models measure the business value that they generate.
- An analytics technical policy committee that determines what data, analytics applications, processes, best practices, and standards are used across the organization.
- An analytics security and compliance committee that oversees the security and compliance of data and analytics processes and applications.
- An analytics data management and data quality committee that ensures the organization's data and metadata are accurate, complete, and consistent.

CONCLUSION

Organizations that desire to derive value from big data through analytics are more likely to succeed if they pay attention to the following four aspects of how analytics is viewed and organized: 1) Do senior leaders in the organization recognize the importance of analytics? 2) Is there a critical mass of data scientists who understand the organization and does the breadth of their expertise span not just building analytic models, but also deploying them? 3) Do the data scientists in the organization understand the various processes required for selecting the right models to build; building them correctly; and deploying them into operational systems and processes so that value is generated? 4) Is there an analytic governance structure in place to support analytics and to integrate analytics and big data into the organization's overall strategy?

ADDITIONAL INFORMATION

The views expressed in this paper are the views of the individual authors and do not necessarily reflect the views, opinions, intentions, plans, or strategies of their employers.

REFERENCES

Data Mining Group. 2012. Predictive Model Markup Language (PMML). Accessed April 7, 2014: www.dmg.org.

Davenport, TH. 2006. Competing on Analytics. *Harvard Business Review* 84(1): 99-107.

Galbraith JR. 2008. Organization design. In T.G. Cummings (Ed.), *Handbook of Organization Development*: Sage Publications, Inc., Thousand Oaks, CA.

Grossman RL. 2009. What is analytic infrastructure and why should you care? *SIGKDD Explorations Newsletter* 11(1): 5-9.

Laney D. 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group. Accessed April 7, 2014: http://blogs.gartner.com/ doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Manyika J, Chui M, Brown B, et al. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Accessed April 7, 2014: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Parmar R, Cohn DL, Marshall A. 2014. Driving innovation through data. IBM Institute for Business Value. Accessed April 7, 2014: www-935.ibm.com/services/us/gbs/thoughtleadership/innovation-through-data/

Press G. 2013. A very short history of data science. Forbes. Accessed April 7, 2014: www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/

White T. 2009. *Hadoop: The Definitive Guide*. O'Reilly Media, Sebastopol, CA.

ROBERT L. GROSSMAN

University of Chicago
Computation Institute
E-mail: robert.grossman@uchicago.edu

KEVIN P. SIEGEL

Visa U.S.A.
E-mail: kpsiegel@yahoo.com