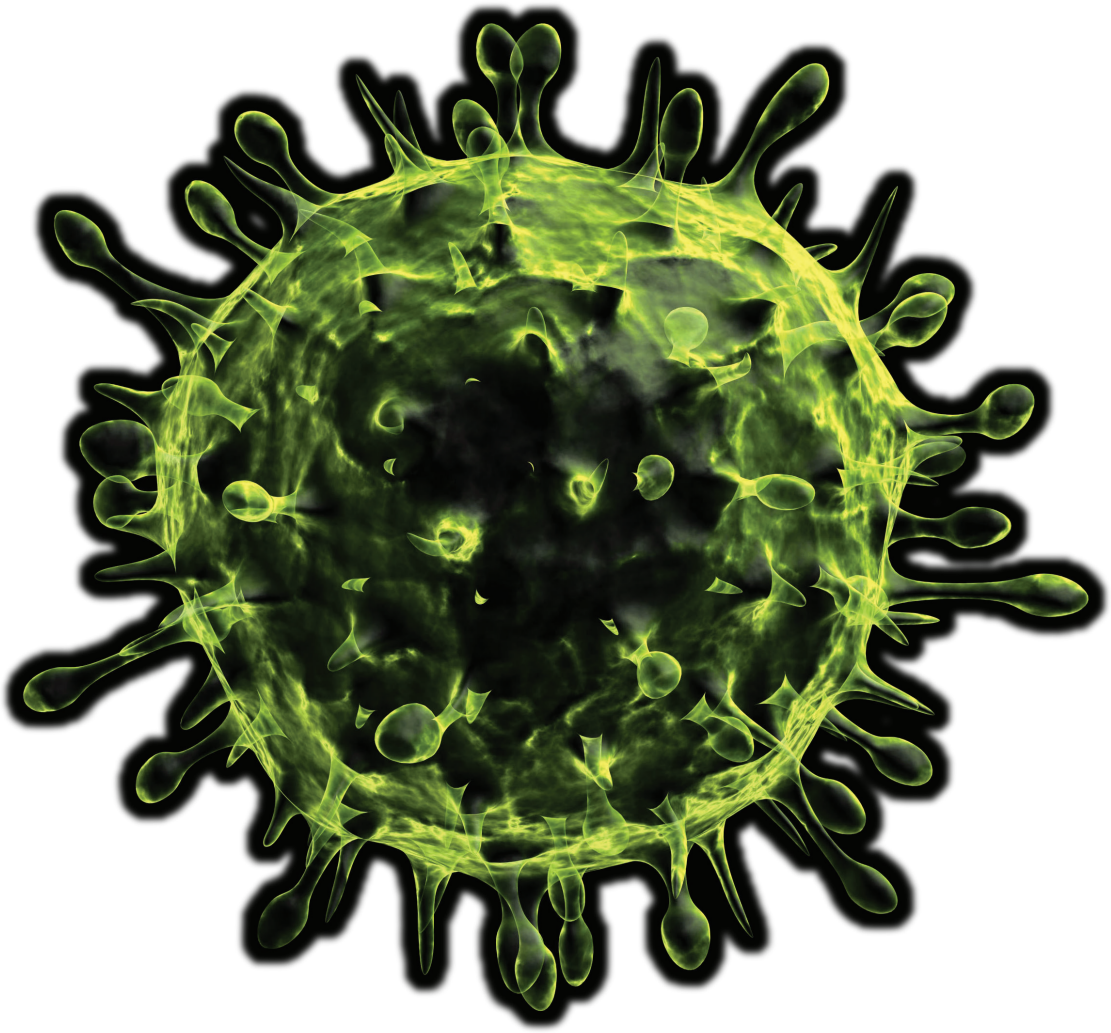
Deloitte.

Qlik



Seeing the big picture about big data

Next-generation analytics for life sciences organizations

Contents

> **"The journey to becoming an insight-driven organization is a three-step process, moving from information to insight to impact."**

Transformative data governance enabled by next-generation analytics for life sciences organizations

In a rapidly evolving marketplace fueled by regulatory changes and global competition, data is possibly the most valuable asset that life sciences organizations have. While most organizations see the potential of leveraging data streams from new technologies such as wearables and connected devices, they also see a growing number of data challenges before them. How can they handle the massive volume of data being produced today? How can they keep data safe and compliant? How can they ensure that the right audiences are using data in the right context? The answers often lie in their approach to data governance and the information management and business intelligence platform they use to enable it.

While almost all life sciences organizations have data governance structures, new technologies, and the "big data" they acquire, produce and consume for decision-making purposes, are straining these legacy environments. In many instances, standard models and processes are no longer sufficient. The journey to becoming an insight-driven organization is a three-step process, moving from information to insight to impact. Since good data governance is a prerequisite for effective analytics, organizations won't be able to arrive at the insights and impacts they desire without a change in focus.

Big data as the big disruptor

Many disruptive forces are converging upon life sciences organizations today, ranging from collaborative business models to healthcare reforms. However, "big data" could be the most challenging, and transformative, of them all. While the definition of "big data" differs across industries, for purposes of this discussion it comprises five main data streams:

1. Web and social media—Through soci al media sites and a number of other channels, customers can share their experiences with a medication or device as well as their sentiments regarding a company or a product.
2. Machine-to-machine—From sensors deployed in a hospital environment to wearable devices consumers use at home, communication between devices through the Internet can produce terabytes of information per day.
3. High-volume, high-velocity transactions—This encompasses major administrative areas such as patient claims and interactions with benefit management organizations.
4. Biometrics—While still evolving, biometrics can produce huge volumes of highly sensitive, personal data such as fingerprints, voice scans and face maps.
5. Personally generated—From emails to blogs, personally generated data represents massive streams of mostly unstructured content.

This deluge of data represents an opportunity if well managed, but a liability if not. Big data, however, is voluminous and complex in terms of its breadth of sources, varied formats and unstructured nature. The rate at which it is being produced is also much faster than anyone could have anticipated even a few years ago, and employees often become overwhelmed in trying to process it—not to mention in trying to extract valuable insights from it. The velocity, volume and variety of big data additionally compound many of

the existing data-related risks and challenges inherent to the life sciences sector. These include difficulties in accessing and managing the complexity of clinical and R&D data, which is typically strewn across many companies and third parties. Indeed, a heavy reliance on third-party data in general poses unique risks to life sciences companies, raising questions not only about which data sets to purchase but also how to use them effectively. Further more, the availability of a wider variety of data sets, both from internal sources and from third parties, raises new concerns regarding how to secure sensitive Personally Identifiable Information (PII) or Protected Health Information (PHI). It also elevates the risk of regulatory non compliance, since patient data is increasingly a component of these data sets. Meanwhile, the ownership of the data is becoming less clear, all while companies are being called upon to share data with collaborators in a safe and compliant way.

In order to address these challenges, while simultaneously unlocking the insights that big data has to offer, many life sciences organizations will need to reassess their data strategies as well as evolve their data governance organizations.

The "democratization" of data governance

With huge volumes of data coming from many different sources both from within and outside the enterprise, a new "democratized" approach to data overnance may be required. In order to be effective, data governance can no longer be driven solely by a centralized governance organization. Instead, various functions of business and IT should be involved, with their roles, responsibilities, handoffs and interactions being guided by a collaborative, responsive, and agile framework. The good news is that taking this new "democratized" approach rarely means starting over. The traditional dimensions of governance still apply to big data, but with a shift in focus—mainly realigning objectives and activities across people, processes and technology. These data-governance dimensions, and proposed shifts, are discussed in greater detail below:

- Organization—As previously mentioned, the data governance organization should become flatter and more agile in order to accommodate big data. This will likely require extending existing business roles or adding new ones. In general, companies should seek

"Big data and master data are highly synergistic, i.e., Big data uses master data, and master data can be enriched based on big data."

to assemble a research-oriented team that includes a mix of data scientists, senior architects and business users, and to implement processes that allow them to collaborate. They should also verify that the team's skill sets align with enabling technologies so that participants can make the most of new and existing tools.

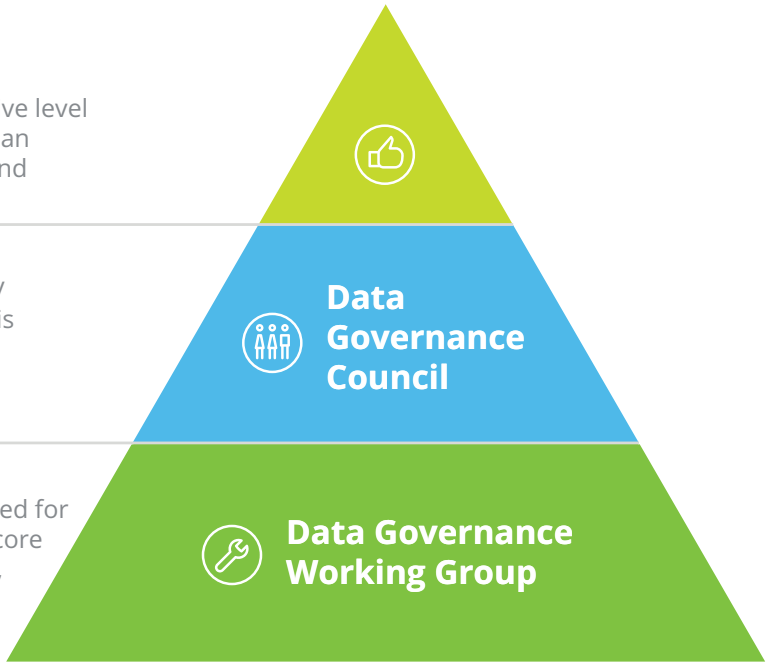
- Master data integration—Big data and master data are highly synergistic, i.e., Big data uses master data, and master data can be enriched based on big data. And quite importantly, low-quality master data can impede big-data analytics. To benefit from these synergies and to enable big-data analytics, new master-data-management roles may need to be added to the existing data governance structure. These individuals will be responsible for defining and executing processes for using big data in order to enrich master data in alignment with business needs. For instance, this might involve leveraging the web or social media to increase the quantity of data or to validate and complete master data. To improve master data management, companies may also wish to consider leveraging new technology capabilities, such as solutions for sentiment and network analyses, customer segmentation, preference management and churn management, along with business-friendly graph databases that scale more naturally to large data sets.
- Metadata management—It is important for organizations to assign a high-priority to the agile development of building and maintaining a metadata infrastructure. Why? Even though it is more challenging for big data, metadata management is critical for deriving meaningful insights from the flood of information. Once again, the business will need to play a more active role in defining the data "context," since big data enters the organization through many paths, and not just through a centralized IT department. Thus, companies will likely need to extend existing roles that sit within the business, such as information stewards and metadata administrators, to encompass metadata-management activities in relation to big data. Processes may also need to be redesigned or established, such as those for creating and maintaining a consistent business glossary and data dictionary, and those for facilitating communication between the business and IT. To enable these expanded activities, the analytics and visualization tools that decision-makers leverage to gain insight should also provide capabilities for metadata management at the point of usage. Tools that enable automated capture, assignment and publication of metadata should additionally be explored.
- Data security and privacy—The objectives of any data security program are to guard against breaches, prevent misuse of data, and to mitigate reputational, regulatory and legal risks. However, big data is a prime target for abuse since it rapidly introduces a huge volume of heterogeneous information. This information could easily be mishandled or misinterpreted, especially considering that laws and regulations concerning data use vary across geographies. To manage the security and privacy aspects of big data, an

Governance Steering Team

Governance Steering Committee includes executive level group from security, risk and procurement who can make strategic decisions considering the scope and attributes of the data to be managed.

As the big data changes rapidly, the responsibility of management and implementation of big data is delegated to the data owners and working group by the council in alignment with the attributes determined by the council.

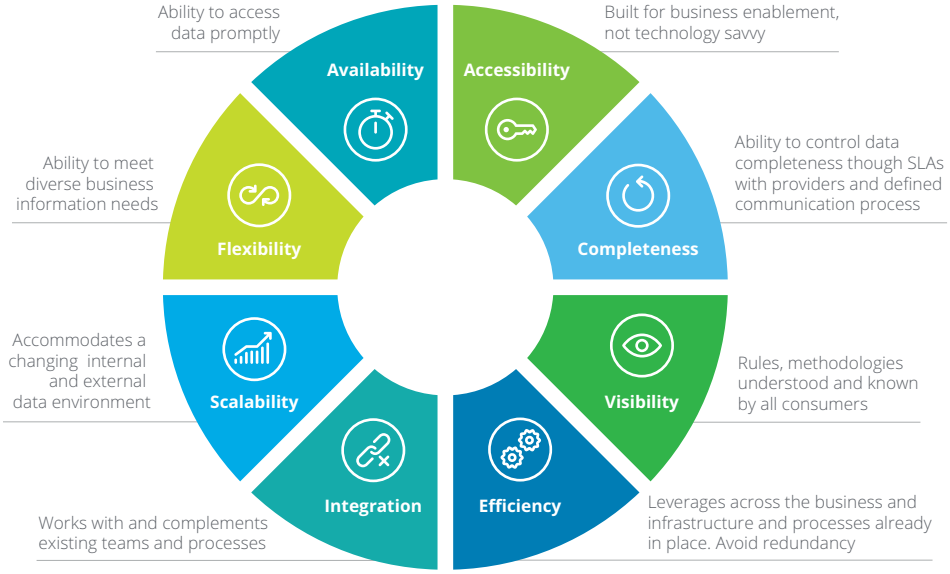
More specialized and varied skillset will be required for executing and taking responsibilities for various core data management roles, including data stewards, metadata lead, data architect, data quality lead, data scientists and data custodians.



expanded approach may be required that includes assignment of security analysts as well as integration with legal and risk-management departments. The role of this integrated team is to minimize security threats by designing processes for identifying and managing sensitive data, while taking the legal and legislative aspects of the data sources into consideration. In addition to upgrading security measures, companies will likely need to upgrade their security technologies as well, with an eye toward deploying adaptive and automated programs to maintain visibility and control over internal and external data, therefore protecting information assets. It is also important to address data security at the information consumption layer through analytics and visualization tools. Self-service analytics should enforce security with data segmentation at the application, metadata, and data access levels all while providing the flexibility of business driven big data analysis.

- Data quality—With big data, the data relationships are not explicit and data may be loaded in "data lakes," or large-scale storage repositories. Because it is largely unstructured, big data demands a different approach in measuring, improving and certifying the quality and integrity of the data. Companies may wish to consider involving business stakeholders to determine confidence intervals for the quality of big data and appointing data stewards who are accountable to an information governance council for improving metrics over time. Processes will likely need to be redesigned for understanding the needs of the business, exploring incoming data, and then defining the data quality rules for big-data elements that are critical to producing the desired business outcomes. In terms of technology enablers, sophisticated data cleansing solutions that are based on machine-learning and predictive modelling algorithms can be helpful in correcting deficiencies and optimizing and updating quality rules. Additionally, data visualization can be leveraged by analysts to help identify data outliers and measure/monitor data quality.
- Data lifecycle management—The volume and variety of big data make it hard to understand the regulatory and business requirements that dictate what data to retain, compress, archive or delete. Within the data governance organization, responsibilities should be assigned for defining retention and disposal rules, potentially collaborating with the legal and risk departments in the process. Ultimately, the retention schedule should be expanded to include big data, and the associated legal requirements should be documented and socialized throughout the

Information governance enables better information usage and increased data adoption



enterprise. In enabling data lifecycle management, data compression techniques can be helpful in reducing data bottlenecks, thus improving performance and decreasing total cost of ownership.

Similar, but expanded On the whole, an effective, big-data governance organization differs from the traditional governance structure in terms of its breadth (i.e., the people involved, as well as their roles and responsibilities) and its collaborative capabilities. This expanded organization essentially serves as a cross-functional gatekeeper of all of the commercial and clinical information that is delivered within the enterprise. As in traditional governance models, the various levels within the transformed data governance organization are responsible for:

- Finding it (data acquisition)—Includes the strategic sourcing of data, managing those who provide or produce it, and setting proper expectations on the completeness quality, timing, consistency, and use of data.
- Storing it (data integration)—Encompasses verifying that the data is current, complete, consistent, compliant, accurate, and actionable; owning the data dictionary and business rules that govern importation and storage; governing how data is accessed, used, and updated; and managing stewardship operations.
- Delivering it (information access)—Comprises owning access and authorization on reporting platforms; ensuring report availability; creating "certified" selfserve data access; ensuring consistent interpretation of key performance indicators (KPIs); and responding to ad hoc requests.

Conclusion

The opportunities and risks associated with big data are much too large to be ignored. In order to capture the potential of big data to fuel growth and enable innovation, while managing the new types of risks that accompany it, life sciences organizations may need to transform their legacy data governance structures. Today, a centralized governance organization is still an important part of the solution, but more is needed. Transformative data governance requires both distributed and centralized participation, and it provides an environment for keeping both big data and traditional data safe and compliant, while simultaneously ensuring that the right users have access to relevant insights in the necessary context to drive innovation.

Data is the currency of the future. For more information on how transformative data governance can help your organization to leverage this currency in the life sciences sector, please contact:

Contacts

Scott Barnes
Director
Deloitte Consulting LLP
scbarnes@deloitte.com

Anjan Roy
Specialist Leader
Deloitte Consulting LLP
anroy@deloitte.com

Peter Leddy
Director
SI Solutions
Qlik
Peter.leddy@qlik.com

Christopher Ferrara
Director
Life Sciences
Qlik

About Deloitte and Qlik

Through a strategic alliance, Deloitte and Qlik collaborate in helping organizations to improve the way they access, visualize, and interact with their data. Together, Deloitte and Qlik offer powerful solutions for analytics and data discovery, leveraging Deloitte's knowledge of leading practices in data management and Qlik's business intelligence and data visualization platform. Industries served include life sciences & healthcare, financial services, and supply chain & logistics, among others. As a result of its significant experience with Qlik delivery projects, Deloitte has received several Qlik Partner of the Year and Developer awards, including 2015 NAM System Integrator Partner of the Year, 2015 Benelux System Integrator Partner of the Year, and 2015 APAC Partner of the Year.