Data Analytics

Part 1: Literacy, Governance, and Management

The Institute of
Internal Auditors

Contents

About the Experts

Steve Mar

Steve Mar has had a long career in auditing technology (IT audit) and information security. He specializes in assessing information security, new systems development, and technology infrastructure risks and controls. He led and developed IT audit teams for Nordstrom, Resources Global, Microsoft, Deloitte, KPMG, and Bank of America, as well as written books and given presentations on technical auditing. Today, he works as the director for Intraprise TechKnowlogies, LLC while also teaching as an adjunct professor at Seattle University.

Trent Russell

Trent Russell is the Founder of Greenskies Analytics and the host of The Audit Podcast. He graduated from the University of Alabama with an MIS degree before joining Ernst & Young's (EY's) IT risk assurance practice, where he served multiple industries. He later joined the financial service office at EY and facilitated the development of data analytics procedures.

Introduction

Adapting to a new data frontier

Since the days before computers, even as far back as when businesses could be run out of horse buggies, people have relied on data to make strategic decisions about their companies. The difference between then and now is the sheer volume of detailed, extraordinarily complex data that modern technology can collect for us. According to a recent study from SeedScientific, it is estimated that there is currently approximately 44 zettabytes of data in the world. For a bit of perspective, one zettabyte equals enough storage to download about 500 billion full-length movies.

By 2025, SeedScientific estimates that the world's data output will reach 175 zettabytes. Although not all of this data is relevant for any significant purpose, let alone relevant to a particular organization or industry, it does speak to the wealth of information available for people and entities to analyze — should they so choose.

Today, data analytics — the formal term for examining data sets to find trends and draw conclusions — is a key component of virtually all business strategies impacting all levels, from executive management decisions down to frontline workers. This means it must be a fundamental part of the internal audit universe as well, both as a tool and as a focus (directly or indirectly) of audits. In this Global Knowledge Brief, the first in a three-part series on data analytics, we will provide an overview of the forms data takes in today's business landscape, as well as how this data must be accounted for in an effective data governance and management strategy. Without a foundational understanding of these topics, internal audit will find providing independent assurance over this critical area next to impossible.

What Is Data?

Know your risk

The forms of data

When people hear the term "data," they likely envision strings of numbers scrolling down a computer screen, perhaps in an extensive spreadsheet or table. That is certainly one form, but the term data goes far, far beyond that image. Indeed, data is any kind of information that can be collected and analyzed.

"To understate all it can encompass, data can come in the form of a tweet or a social media feed, which would be a form of what some people would call public data," said Trent Russell, founder of Greenskies Analytics and host of The Audit Podcast. "It could be on a cloud system, or it could be in paper form. It could be customer trend reports; it could be revenue reports; it could be time data collected and stored by machines on a factory floor. It's a nearly endless list, generally speaking."

There are a few basic data categorizations that can help narrow the list. In 2018, Forbes published a list of 13 data categories, among them:

1. **Big Data**. Refers to the incredible amount of data that will not practically fit into a standard (relational) database for analysis and processing caused by the huge volumes of information created by human and machine-generated processes.

2. **Structured, unstructured, and semi-structured data.** Some data have a level of pre-defined composition and are stored somewhere in some kind of format. Other data, however, such as an organization's social media feed, does not have a composition that fits neatly into such paradigms.

3. **Time-stamped data.** Some data have a concept of time ordering defining the sequence that each data point was either captured (event time) or collected (processed time).

4. **Open data.** Data freely available to anyone in terms of its use (the chance to apply analytics to it) and rights to republish without restrictions from copyright, patents, or other mechanisms of control.

5. **Machine data.** This refers to the digital exhaust created by the technologies and infrastructure powering modern business. Some examples include application programming interfaces (APIs), security endpoints, message queues, change events, cloud applications, call detail records, and sensor data from industrial systems.

6. **Real-time data.** A term to refer to data collected by the instantaneous computing that happens about as fast as a human can perceive.

7. **Dark data.** Digital information that is not being used and lies dormant in some form.[1]

---

1. Adrian Bridgewater, "The 13 Types of Data," Forbes, July 5, 2018, https://www.forbes.com/sites/adrianbridgwater/2018/07/05/the-13-types-of-data/?sh=779ec9723362,

Data Governance and Management

What a sound data strategy looks like

Know your business

Luckily, organizations do not have to have a firm grasp on all possible kinds of available data — a task that in the modern world borders on the impossible. Instead, the more important task for organizations looking to invest in data analytics is to identify the data most relevant to them. The larger the pool of available data and data sources, the more challenging such a task becomes. However, working to overcome this challenge lays the foundations for an organization's data governance and management strategy.

According to Steve Mar, a director at Intraprise TechKnowlogies, LLC and adjunct professor at Seattle University, gaining an understanding of what data is most relevant to an organization begins with a vision for how the data will be used. "If you don't have a vision and plan; if you are not aligned with a corporate strategy; if you don't know what you are capable of doing in your department; you will not succeed," he said.

Critical to the development of such a plan is a keen understanding of the company's business processes. "Try to determine what makes your organization successful and where it is getting the largest return on investment — that will define what data will be most relevant to a data analytics strategy and where," said Mar.

"Some organizations, for example, don't realize they are basically an inventory place," he continued. "What matters the most to them is their supply chain, getting accurate data, inventory management, etc. In this case, you might not have to worry about access to a general ledger account tracking parking fees. You have to worry about your machinery, inventory levels, how part numbers and SKUs are managed, and things like that. And if you don't have that kind of information and understanding of your process, you will quickly get into trouble."

Data storage and protection

A complex regulatory environment

An understanding of how and where an organization's data is stored is vital not just from a data governance standpoint but to comply with an increasingly rigorous regulatory environment, as well.

Although some industry-specific data protection-related regulations existed before, such as the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA), the pioneering legislation regarding data governance is the European Union's General Data Protection Regulation (GDPR). Implemented in 2018, it sparked a global movement to hold private entities accountable for protecting personal data, while enhancing individuals' control and rights over their data. This fundamentally altered business strategy, personnel needs, and the risk landscape.

In the four years since GDPR's implementation, the data protection environment has only grown more complex. As of 2022, more than 120 countries have enacted some form of international data privacy laws.[2] Complicating the picture further, the U.S. has taken action to address data protection at a state level, with five — California, Colorado, Connecticut, Utah, and Virginia — currently

---

2. "Data Privacy Laws by Country 2022," World Population Review, 2022, https://worldpopulationreview.com/country-rankings/data-privacy-laws-by-country.

having comprehensive consumer data privacy laws on the books.[3] According to Gartner, by 2024, 75% of the global population will have its personal data covered under privacy regulations.[4]

To avoid costly penalties, it is imperative that organizations dedicate resources to identifying the regulations applicable to them, keep current on additions or changes to data regulations, and provide appropriate assurance over compliance. Such tasks can be distilled down to asking and answering three basic questions:

1. Where are our customers?

2. What are the corresponding notification requirements?

3. Have any of the areas in which we operate enacted their own data privacy regulations?

Data ownership and access

While not saying that data storage in previous technology generations was simple, it was at least more tangible. Data could be on print documents or in databases, data lakes, or other various digitized architectures, and it could be secure if basic data governance practices such as restrictive access were maintained. With the widespread adoption of Big Data, however, data management became much more nuanced and abstract.

In response, organizations with sound data governance policies and practices have embraced an entire data security ecosystem that includes detailed Big Data governance frameworks; data governance committees; Big Data architects, data engineers, and IT teams; and, in some cases, third-party cybersecurity providers. Together, these entities foster a data security environment founded on collaboration, enterprisewide data education, and clearly defined roles and responsibilities.

Discussion of these roles inevitably includes who has access to the data in question. Accessibility expectations can be established within organizations through updated codes of conduct and policies that contain data governance-related requirements, including:

- An inventory of data covered by the organization's data governance policies.

- A list of security best practices to ensure employees safeguard company equipment and mobile devices.

- Access restrictions tied to job roles and responsibilities.

- A list of prohibited actions that could negatively impact data governance and corresponding penalties.[5]

When considering such requirements, it is important to establish who should have primary responsibility over the organization's data. Typically, said Russell, this should fall to an appointed chief data officer (CDO), who establishes policies and standards for data governance, provides oversight of the implementation of data governance controls, and leads the data governance committee.

"This chief should be keen enough on the strategy of the organization, know how to rank data by priority, and be able to see opportunity within the data," he said.

Internal audit has a vital part to play alongside the CDO.

"Even if the auditor at the table isn't necessarily a data expert, their broad view of both the organization and the risk environment has significant value," said Russell. Not only can they provide assurance for the organization's data governance strategy through the testing of controls, he said, but they also can provide insights on how new data trends, such as machine learning analytics, Internet of Things (IoT), and the increasing adoption of 5G, could fit or not fit into the current governance model.

3. "State Laws Related to Digital Privacy," National Conference of State Legislatures, Jule 7, 2022, https://www.ncsl.org/research/telecommunications-and-information-technology/state-laws-related-to-internet-privacy.aspx#:~:text=Five%20states%E2%80%94California%2C%20Colorado%2C,of%20personal%20information%2C%20among%20others.
4. Gartner Identifies Top Five Trends in Privacy Through 2024," Press Release, Gartner, May 31, 2022, https://www.gartner.com/en/newsroom/press-releases/2022-05-31-gartner-identifies-top-five-trends-in-privacy-through-2024#:~:text=By%202024%2C%2075%25%20of%20the,Data%20Covered%20Under%20Privacy%20Regulations.
5. "Data Governance," The IIA, 2020, https://www.theiia.org/en/content/articles/industry-knowledge-brief/2020/data-governance/.

Conclusion

The importance of buy-in

All these aspects of data governance and management are for naught without enterprisewide buy-in starting with the board. This is true not just for designing and implementing a successful data governance strategy, but also to unlock the potential for greater use of data analytics within the organization.

"You could love data analytics, but without support from the board and executive leaders, you are going to be quite lonely in your data analytics dreams," said Mar. "They may know it's a great thing to do, maybe even necessary, but lots of people say they love to diet and exercise and don't do it behind the scenes."

The secret, said Mar, is to present, or at least play a part in supporting, a powerful data analytics vision. This requires not just restating the benefits of data analytics, but also understanding how to do data analytics enough to show the benefits. "If you don't know what you're doing, you will not succeed. I don't know how to emphasize that more," he said.

With proper understanding of the organization's data needs and strategies, internal audit can present a data analytics audit plan that is detailed, clear, and illustrates to the board a tangible return on investment. Part 2 of this series will highlight what such data analytics plans can look like for internal audit, and how auditors can further adopt data analytics into their audit role in new and exciting ways.