

The Domains of Data and Information Audits

In my early days as a practitioner, the use of computers was referred to as data processing and electronic data processing (EDP), and this covered primarily batch processing of inventory, banking transactions and billing for utility services. Later came transaction processing such as airline reservations and ticketing transactions. At that time, expressions such as first in, first out (FIFO) and garbage in, garbage out (GIGO) were in common use, the first because memory was a precious resource and the second because data was entered manually and programming errors were common.

“Data underlie all processing and poor quality data are now the garbage of the information society.”

The reason for this historical prelude is that, at that time, there was no talk of “information technology” or the “information society” and, in fact, these concepts have hidden the fundamental truth that data underlie all processing and poor quality data are now the garbage of the information society.

To complicate matters, there is semantic confusion among the terms “data,” “information,” “business intelligence” and “knowledge,”

which many, particularly in the user community, frequently use interchangeably. When the terms “enterprise information management,” “master data management,” “data governance” and “IT governance” are added to the mix, real confusion ensues. There are many sources defining their meaning and the reader is invited to pursue them if required.

The information society, populated by billions of people who have computer-like devices, access to the Internet (and to corporate networks, systems and data, and social networks) and the possibility of creating data, has created an environment in which data quality and meaning cannot be taken for granted. It has, in essence, taken us back to GIGO. Unless this is addressed, business intelligence and big data initiatives risk having less value than intended.

Recent audits and discussions have revealed a number of data-related issues such as:

- Some organizations may not know how good or bad the quality of their data really is, assuming they have a full inventory of their data. Huge amounts of data are being created (spreadsheets, personal databases, web pages).
- Such data are “invisible” to the organization and imply that business functions are generating similar, but inconsistent data sets from different sources, i.e., data silos that have barriers to consolidation and integration. A quotation that is pertinent in this context is, “A man with one watch knows what time it is; a man with two watches is never quite sure.”
- Many data creators are not aware of what data already exist, what their attributes are, where they are kept or for what they are used, so they are happy to create their own version.
- There are few policies for data creation, storage and management, including disposal.

Exploring Differences

This column proposes the following definitions:

- **IT governance**—The mechanism that can ensure that investments in technology (including

Ed Gelbstein, Ph.D., 1940-2015

Worked in IS/IT in the private and public sectors in various countries for more than 50 years. Gelbstein did analog and digital development in the 1960s, incorporated digital computers in the control systems for continuous process in the late '60s and early '70s, and managed projects of increasing size and complexity until the early 1990s. In the '90s, he became an executive at the preprivatized British Railways and then the United Nations global computing and data communications provider. Following his (semi) retirement from the UN, he joined the audit teams of the UN Board of Auditors and the French National Audit Office. Thanks to his generous spirit and prolific writing, his column will continue to be published in the *ISACA® Journal* posthumously.

software) create value and support business objectives. There are several standards for this discipline, notably International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 38500 and COBIT® 5.

- **Data governance**—Processes and controls focus on the quality of data (raw alphanumeric characters) and ensure that when the data are used by the organization they are unique, accurate and timely. Master data management seeks to standardize data and ensure there is one single version of the truth.

Figure 1 shows a summary of the components of this discipline, many of which are also part of enterprise information management.

- **Enterprise information management**—Defined in *The DAMA Guide to the Data Management Body of Knowledge (DMBOK)*¹ as covering 11 topics: data management overview, data governance, data architecture management, data development, data operations management, data security management, document and content management, reference and master data management, data warehousing and business intelligence, metadata management, and data quality management

Published by the Data Management Association International and revised in 2013, the DMBOK is a comprehensive set of guidelines of good practice and a worthwhile addition to the libraries of IS/IT auditors as well as data owners/stewards.

What Can a Data Audit Deliver?

A data audit addresses organizational data in repositories such as databases, data warehouses and spreadsheets, and in the cloud. A data audit should deliver diagnostics on:

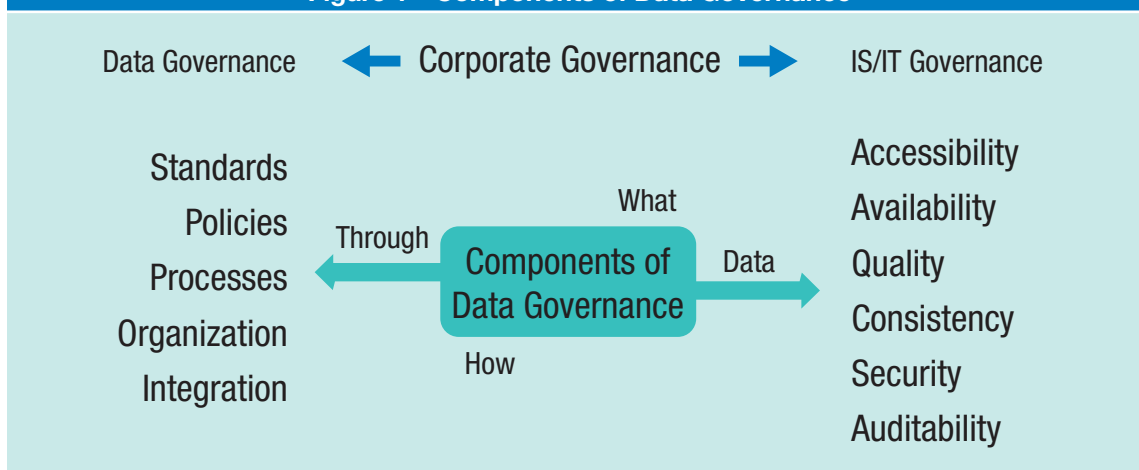
- Incomplete, inaccurate or inconsistent data and appropriate cleansing needs
- Data sets that do not comply with privacy or regulatory laws
- Gaps in data security levels or processes
- The location of the organization's data sources
- Rogue (unverified, untraceable) data that someone may be using without anyone knowing about it
- Gaps in accountability for data and recommendations as to where data stewardship is required
- Changes to processes, revisions to existing policies or areas where new policies are required

This information should provide a better understanding of what data issues exist in the organization and the actions required to achieve an appropriate level of data quality.

A Data Audit Framework

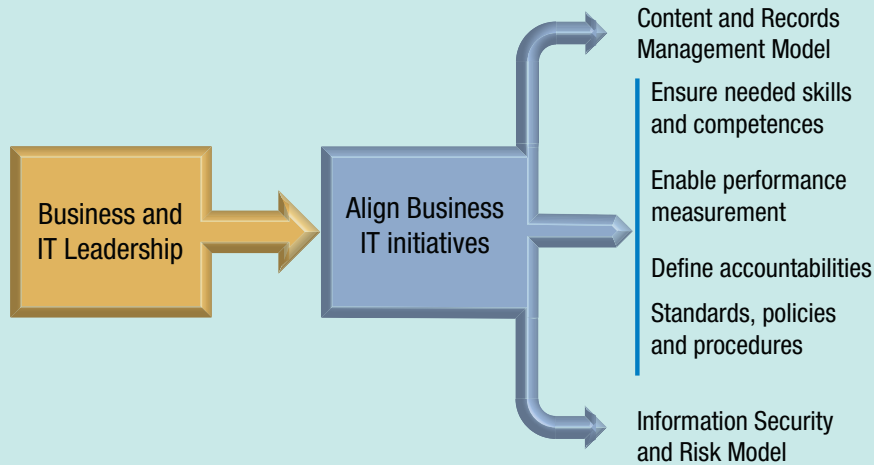
Effective data management and administration have a number of prerequisites, summarized in **figure 2**. When one or more of these elements are missing or incomplete, it will be that much more difficult, or even impossible, to achieve success.

Figure 1—Components of Data Governance



Source: E. Gelbstein. Reprinted with permission.

Figure 2—Prerequisites for Successful Data Management



Source: E. Gelbstein. Reprinted with permission.

The reader may wish to consult the Data Audit Framework² developed by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow (Scotland, UK). It, and its associated methodology, are available as free downloads.³

The Data Audit Framework Methodology proposes four steps for the audit of data governance, which can be summarized as follows:

- **Step 1 Audit preparation**—Prepare in advance to make the audit process smoother and less time intensive.
- **Step 2 Data classification**—Identify and create an inventory of all data sources and resources that end users require and are using to effectively do their jobs.
- **Step 3 Data management**—Determine who, if anyone, is responsible and accountable for data sources.
- **Step 4 Audit results and recommendations**—Summarize audit findings and make recommendations to key stakeholders from IT, the business and management.

Part of the current challenge arises from IT functions having taken responsibility (by default) for data because many business managers lack data

management skills. Outsourcing and data migration to cloud services have also contributed to the problem.⁴

As a consequence, data may not be properly managed, and terms such as “data administration” and “data management” discourage many from taking ownership.

“Terms such as ‘data administration’ and ‘data management’ discourage many from taking ownership.”

Policies, or rather their absence, add to the challenge of having high data quality. A set of data-related policies (assuming these are understood, monitored and enforced) ought to include:

- **Data ownership and stewardship**—Defining accountability for maintaining and operating the data stores, both stand-alone copies and part of production applications

- **Data security**—Keeping data secure regardless of whether data are controlled by an application system, copied to a test or training database, or stored in the cloud
- **Data location**—Keeping track of the data used by the organization, where they are located and how they are secured
- **Data traceability**—Maintaining a record for various types of data, how they are interfaced between systems and what transformations were applied in the process
- **Data quality**—The rules systematically applied in the capture, monitoring and measurement of data assets
- **Service levels**—The required service levels for the timeliness of data delivery or synchronization between copies of the data

There are several other topics that merit formal policies, such as business intelligence, enterprise content management and data architecture, but their discussion exceeds the scope of this article.

The ISACA® Knowledge Center includes among its topics enterprise data management (www.isaca.org/topic-enterprise-data-management), big data (www.isaca.org/topic-big-data) and business analytics/intelligence (www.isaca.org/topic-business-analytics-intelligence). They contain very valuable information.

Conclusions

As business intelligence and big data initiatives mature, the risk of working in a GIGO environment becomes significant when data quality and an understanding of their sources and meaning cannot be guaranteed.

There are many sources of guidance on data governance, data quality and data-related audits. Beyond them, readers may wish to consult the data audit checklist⁵ produced by the US Education Department. Many vendors offer publications on the topic as well.

Endnotes

- 1 Data Management Association International, *The DAMA Guide to the Data Management Body of Knowledge*, 2013, www.dama.org
- 2 Jones, S.; A. Ball; C. Ekmekcioglu; “The Data Audit Framework: A First Step in the Data Management Challenge,” *International Journal of Data Curation*, vol. 3, no. 2, 2008
- 3 Jones, S.; S. Ross; R. Ruusalepp; M. Dobrev; “Data Audit Framework Methodology,” Humanities Advanced Technology and Information Institute, University of Glasgow, Scotland, 2009, www.data-audit.eu/DAF_Methodology.pdf
- 4 Gelbstein, E.; V. Polic; “Data Owners’ Responsibilities When Migrating to the Cloud,” *ISACA® Journal*, vol. 6, 2014, www.isaca.org/Journal/archives
- 5 Data Governance Checklist, <http://ptac.ed.gov/sites/default/files/data-governance-checklist.pdf>