

Big Data - what's the big deal and why is data quality so important?

 By [Gary Allemann](#)

18 Jul 2012

In today's world volumes of data are rapidly expanding, and the nature of this data is more varied and complex than ever before. While this concept of Big Data brings with it numerous possibilities and insights for organisations, it also brings many challenges, particularly around the management and quality of data.

All data has potential value if it can be collected, analysed and used to generate insight, but in the era of Big Data new concepts, practices and technologies are needed to manage and exploit this data, and data quality has become more important than ever.

Big Data has three characteristics, namely volume, variety and velocity. Data today is growing at an exponential rate, with around 90% of all digital data being created in the last two years alone and somewhere in the region of 2.5 quintillion bytes of data being generated each and every day. Data is more varied and complex than ever, consisting of a mixture of text, audio, images, machine generated data and more, and much of this data is semi-structured or unstructured. This data is often generated in real-time, and as a result analysis and response needs to be rapid, often also in real-time. This means that traditional Business Intelligence (BI) and data warehousing environments are becoming obsolete and that traditional techniques are ill equipped to process and analyse this data.

Multiple sources

When it comes to generating Big Data there are multiple sources, some widely known and some that organisations rarely think about. Photographs, emails, music downloads, smartphones, video sharing and social networks account for a large proportion of this data, and are well known sources. But in today's digital world there are many other places that data may come from, including point of sale devices, RFID chips, recording and diagnostic tools in aeroplanes, cars and other transport, manufacturing, and even scientific research.

All of this machine-generated data, along with Internet and socially generated data, can potentially be a source of insight and intelligence. Big Data has the potential to help organisations gain a better understanding of customer and market behaviour, improve knowledge of product and service performance, aid innovation, enhance revenue, reduce cost, and enable more rapid, fact-based decisions which can enhance efficiency, reduce risk and more. The sheer volume of Big Data presents a massive potential pitfall however, and the challenge lies in being able to identify the right data to solve a business problem or address an opportunity, and in the ability to integrate and match data from multiple sources.

You need to know exactly which data is the data that really matters

In order to leverage the potential of Big Data, it is vital for organisations to be able to define which data really matters out of the explosion of information, and assure the quality of internal data sources to ensure accuracy, completeness and consistency and enable this data to be matched with that from external sources. It is also important to define and apply business rules and metadata management around how the data will be used, and a data governance framework is crucial for consistency and control. Processes and tools need to be in place to enable source data profiling, data integration and standardisation, business rule creation and management, de-duplication, normalisation, enriching and auditing of data amongst others. Many of these functions need to be capable of being carried out in real-time, with zero lag.

Data quality is the key enabler when it comes to leveraging actionable insight and knowledge from Big Data. Poor data quality around the analysis of Big Data can lead to big problems. Consolidating various sources of data without data quality will create a mess of conflicting information, and conducting analytics on data without first assuring its quality will likely lead to the wrong results, poor decision making, and other negative consequences, including lack of compliance which has massive legal implications. The three V's of Big Data, namely volume, velocity and variety, need to be met by a fourth V - validity - in order to harness this data for the benefit of an organisation and its customers.

While the concept of Big Data may have been over-hyped, the reality is that it is here, and it will continue to grow in volume, velocity and variety. While it is currently an immature concept, as it begins to mature data will increasingly be viewed as a strategic business asset and data skills will be increasingly valued. Big Data reflects the expectations and needs of the emerging generation, and businesses would do well to pay attention to this, and ensure that their data quality initiatives are up to speed to ensure they can leverage the potential value of the Big Data phenomenon.

ABOUT GARY ALLEMANN

MD of Master Data Management He is passionate about Information Communication Technology (ICT) and more specifically data quality, data management and data governance.

- Major trends impacting 2023 information and development planning - 27 Jan 2023
- Why data management is key to unlocking the digital transition of African banks - 12 Sep 2022
- Can data privacy concerns be managed when moving to the cloud? - 7 Sep 2022
- Effective WFH streamlined through data management as the next frontier - 8 Mar 2021
- Data governance is key to differentiating your customer experience - 20 Mar 2020

[View my profile and articles...](#)

For more, visit: <https://www.bizcommunity.com>