**Big Data Management**

**and**

**Data Analytics**

# I Introduction

# Data analytics, defined as the process of transforming individual data units into actionable information, has been the traditional strength of the Supreme Audit Institutions (SAIs) across the globe. Data analytics, including IT enabled analytics, largely performed on the data generated within the audited entities, has been at the core of any type of audit carried out by the SAIs – Compliance Audits, Performance Audits, and Financial Audits.

# Audited entities are transitioning into virtualized environment, generating and storing voluminous data and their operations are increasingly getting complex by the day. These apart, large amounts of meaningful and relevant data in disparate forms are being incessantly produced by the external eco system- surveys, performance statistics published by the Government, industry/domain specific data, comparable organizational data with benchmarking possibilities etc. When collated, they provide the contextual framework regarding the functioning of the audited entities. The rapidly evolving environment of entities makes it incumbent on the WGITA to encourage SAIs and prepare them for auditing in the big data environment.

# II Big Data and Data Analytics

Big Data refers to extremely large, complex data sets that exceed the traditional processing capabilities of the IT infrastructure due to their size, format diversity and speed of generation. It refers to the large and complex data collated from all imaginable sources, which leverages information as the vital asset and includes structured and unstructured data, internal and external data and formal and informal communication. The three dimensions of Big data that distinguishes it from conventional data are its features, processes and results.

1. Features:
   1. Volume - quantity, the amount of data
   2. Variety – formats, the data types, data from various sources
   3. Velocity – speed, the speed of data generated and processed
   4. Veracity – quality of data
2. Processes, which include capture, curation, storage, search, sharing and transfer
3. Results - visualization and analysis



Big data analytics refers to the process of analyzing big data to provide (i) deeper insights, (ii) discover patterns (correlation and causation) and (iii) to throw up abnormal behaviour, red flags and outliers that are otherwise hidden, which would enable SAIs to:

* Understand the audited entity better
* Analyse, discover, predict and plan audits
* Plan the nature, extent and timing of audit procedures
* Establish benchmarks and
* Assist in drawing conclusions and reporting outcomes

**III Opportunities and challenges**

Conventional Data analytics, including IT enabled analytics undertaken by SAIs have primarily based on structured data generated within the audited entities. Technological and IT infrastructure limitations have thus far restricted the reach of auditors in many ways. Audit as a process was sustained with sampling techniques and extensive substantive procedures to address the perceived audit risks. The advent of big data and its analytics marks a paradigm shift, which by design envisages synthesizing and integration of relevant data not only from various sources but also in various formats– especially unstructured and textual data to transform data into actionable information and derive value by enhancing the efficiency and effectiveness of audits conducted by SAIs. The format diversity which big data could potentially represent is given in the following diagram.

Given that the SAIs audit a variety of public sector entities, the volume and format diversity of data that is accessible to the SAIs are unparalleled. The data accessible to the SAIs, which can be leveraged for data analytics can be categorized in three domains

* **Data created by the SAIs**
* **Data of audited entities**
* **External data**

**Data created by the SAI :** Data under this category provides a greater flexibility for usage and comprises (a) internal databases of information collected by SAIs about audited entities (b) data of Audit Reports placed in the Legislature/Parliament (c) working papers relating to various audits conducted by the SAIs etc.

**Data of audited entities:** The data under this category is available with the SAIs in the professional capacity as Auditors and its usage involves sensitivities, which would have to be appropriately addressed by establishing and implementing a policy for Data storage, access and usage. Such data comprises the (a) financial and non-financial data of audited entities (b) programme specific data including beneficiary databases (c) data pertaining to cases which are under litigation/arbitration/vigilance/sub-judice etc.

**External Data:** This comprises data, which are available in the public domain and can further be categorised into:

1. **Data published by the respective governments and other statutory authorities –** which could include population data, data published by the various Ministries/Departments of Governments, annual budgets etc
2. **Other data available in the public domain –** which could include surveys conducted by non government institutions, information published by NGOs and other not for profit organizations, industry specific information published by various autonomous bodies/institutions, social media and other data available in public domain

While the data published by the respective governments and other statutory authorities are reliable, the accuracy of other data available in the public domain cannot be easily validated and the meaning and context of the data are not necessarily self-evident. The policy for Data storage, access and usage would have to regulate usage of such data.

The volume and variety of data, however, provides immense opportunities for SAIs and at the same time casts responsibilities on the SAIs to be aware of the challenges of auditing in big data environment. WGITA, in its role of championing the cause of SAIs needs to delve on and appreciate the **opportunities and challenges** presented by big data.

The **opportunities** for the SAIsgerminate from the technology explosion that has taken place in the auditing environment, the transformational impact that big data analytics could have on SAIs andtheconsequential aid togovernance.

* **Technology explosion:** With their operations getting increasingly complex and voluminous, public sector entities are transitioning to virtualized environments and processes occurring real time. Proliferation of technology has ensured that capacities for storage, compute and networking are no longer limitations.

From the SAIs perspective as external auditors, technology explosion has transformed the limitations that existed in the conventional data analytical tools - their inability to combine data, incapacity to handle unstructured data and inadequacies in analyzing large and complex data into opportunities by enabling availability and accessibility of ubiquitous and cost effective tools, technology platforms and solutions.

* **Transformational impact for SAIs:** Big Data Analytics leverages the evidence based approach and its deployment at the audit planning stage for a macro level analysis of almost the entire range of data, rather than on a small representative sample is in itself a paradigm shift from conventional analytical procedures. It enables moving away from the computer assisted audits to a more robust digital auditing to enhance the effectiveness of risk assessment by discovery of red flags, outliers, abnormal behavior, frauds, abuse and deeper insights. Thus, it underpins efficiency in determination of the nature, extent (including the extent of sampling) and timing of audit procedures, ultimately leading to providing a greater level of assurance in audits.

Further, it ushers in new age competencies of predictive analysis, advanced statistics and software usage for transformation of data into actionable information. Data visualization and big data analytics are the value added exploratory functions facilitated by big data, which enables discovery of relationships between variables and broader trends of risk and deepens the reach of auditors.

* **Aid to governance:** It enables the SAIs to assume a proactive role in aiding governance by sharing insights with the governance of public sector entities to promote transparency, effective oversight and control.

**Challenges:** To leverage the potential of big data analytics, the **challenges** are to

1. **manage the people** - This involves transforming the mind set of auditors to adapt to digital auditing and auditing in the big data environment and capacity building
2. **manage the data**- This involves defining the scope of data and addressing data sensitivities associated with access and usage of various sources of data such as third party databases, beneficiary databases, data pertaining to audited entities obtained by SAIs as a part of understanding the entities as well as data collected as evidence, addressing the veracity of data involving an assessment of strengths and weaknesses of various sources, privacy issues and compliance with legislative and regulatory requirements. This would also involve establishing data management protocols address the aspects of data quality, integrity and other processes for data analytics.

(c) **manage the infrastructure** – This involves augmentation of appropriate IT infrastructure and establishing the technical solutions.

**Role of WGITA**: Since inception, WGITA has assisted SAIs in developing their knowledge and skills in the use and audit of information technology. In furtherance of this mandate, WGITA may espouse the goal of preparing the SAIs for auditing in the big data environment and encourage the SAIs in establishment, implementation and maintenance of a robust governance framework. A holistic approach to establishing this framework entails addressing the various dimensions of governance and issues pertaining to:

* Availability and accessibility of data
* Protocols for establishing veracity and reliability of data, with their source and format diversity
* Policies for regulating storage, access and usage of data encompassing security and privacy issues
* Enabling appropriate technical solutions and
* Facilitation of data analytics and visualization

WGITA could provide guidance on best practices that the SAIs could embrace toadapt their audit design procedures and methodologies to the changing order of big data environment.

Some of the areas where the practical guidance could be made available to SAIs by WGITA are:

1. **Guidance on Digital Auditing**- It is the obvious next step from the IT auditing. With the audited entities going virtual, it is important that the auditors develop mechanisms that can analyse data in real or near real time. The auditor’s capability of using audited entity’s electronic data repository for looking for outliers, errors or gaps need to be enhanced.
2. **Moving from sample to whole** – With the availability of data analytics tools, it is now possible to look at the whole data, in place of selecting a “representative sample. Using all the data makes it possible to spot connections and details that are otherwise cloaked in the vastness of the information. This can provide the auditor with anomalies, high risk areas for further detailed analysis.
3. **Guidance on data Analytics –** Data analytics involves processes and activities designed to obtain and evaluate data to extract useful information. The results of such analytics in the audit environment could range from identifying high risk areas for effective audit planning to detect frauds and errors to providing assurance on process effectiveness to giving assurance on financial statements. Guidance on usage of such tools, the risks involved, the training infrastructure needed as per the data analytics techniques selected would be very appropriate.