Abstract—Behind any information management practice lies the core doctrines of Data Quality, Data Governance, and Metadata Management along with considerations for Privacy and Legal concerns, which today have become the information landscape of an organization. Big Data, which is the data that exceeds the processing capacity of conventional database systems, needs to be integrated into this information landscape. The reasons are that Big Data information is being generated at a high volume, with a rapid rate of change, and encompassing a broad range of sources and includes both structured and unstructured data. The ability to understand and manage the sources of the data, and then integrate the data from these sources into a larger business application can provide significant value to the organization. Organizations that have already developed its Enterprise Architecture and now want to integrate a Big Data solution will face issues and concerns related to their Information Architecture, Data Architecture, Business Architecture and Technology Architecture. Enterprise Architecture plays a major role in ensuring that an organization maximizes the business opportunities posed by Big Data. Organizations may have the best technology and the best people but if the internal culture has silos and lacks the data sharing as outlined in the current EA, they are less likely to achieve success with Big Data. In this paper we discuss a framework to construct an entire Information Landscape (EIL) and identify the requirements in constructing this information landscape. We also examine the integration of Big Data with EA to develop the EIL.

Keywords—Big Data solutions; Enterprize Architecture; Entire Information Landscape

I. INTRODUCTION

The generation of data has contributed to the emergence of the Big Data phenomenon [1]. As part of the Big Data phenomenon there has been massive growth taking place in data processing power, data visualization, data storage, network speeds, mobility, accessing real time data, and higher semantic capabilities. Examples of technological changes include social networking, being always connected and online blogging. These technological changes are being provided by companies such as Facebook, LinkedIn, Google and Twitter. These changes were not predicted as mainstream activities a decade ago. However these changes have actually defined the evolution of technologies, infrastructures, applications, users, communities, societies and knowledge creation [4]. They are an integral part of any enterprize and Enterprize Architecture (EA) needs to address these technological changes for organizations to maintain their competitive advantage.

Big Data can be acquired, stored, processed, and analysed in many different ways. Every Big Data source (live data, data from internal sources, and data from external sources) has different characteristics, including the frequency, volume, velocity, type, and veracity of the data. When Big Data is stored and processed additional dimensions come into play, such as security, policies, structure, and governance. Data architecture of EA describes the structure of an organization’s logical and physical data assets and the associated data management resources. All of this forms the information landscape of an organisation, a key component of an EA. Choosing an architecture and building an appropriate Big Data solution is challenging because so many factors have to be considered.

As organizations grow and expand they are also likely to gain technologies that duplicate existing capabilities, with workflows in need of significant overhaul. One of the issues that organizations in such situations often encounter and must address is integrating Big Data which can drive the alignment between the business’ operating needs and the processes, applications and infrastructure required to support ever-more dynamic requirements. As the demand for managing information increases, they need to focus their efforts on integrating their data, application, technology and users across their EIL.

Enterprize Architecture provides pragmatic artefacts such as requirements, specifications, guiding principles, and conceptual models that describe the current state and an idea of how to get the next major stage of evolution of an organization, often called the “future state” [16]. Enterprize Architects need to document in the EA the current state, the future state, the gaps and the roadmap to get to the future state which can be used for the integration of Big Data. EA plays a major role in ensuring that organizations maximise the business opportunities posed by Big Data. Big Data disrupts traditional information architectures of an organization. Organizations may have the best technology and
the best people, but if the internal architecture of the organization has silos and lacks data sharing, they are less likely to achieve success with Big Data. Big Data requires IT leaders and technology specialists to acquire and apply tools, techniques and architectures for analysing, visualizing, linking and managing big, complex datasets. Data integration requires capturing and integrating both structured and unstructured Big Data within the enterprise. Big Data as a part of the overall decision support system or Business Intelligence (BI) system can only be realized by Data Integration [2].

Key to success in any Big Data solution is to first identify the business needs and opportunities, and then select the proper fit-for-purpose solution. In fact, some identified use cases from the business might be best suited by existing technology such as a data warehouse while others require a combination of existing technologies and Big Data solutions (such as EMC Greenplum, IBM’s InfoSphere and Oracle Big Data Appliance) which are fuelling business transformation. There could be legacy applications which need to be integrated to Big Data solutions [3]. For this an EA is required to be developed. Enterprize Architect needs to generate the roadmap by looking at the gap between current state without Big Data and future state with Big Data.

The EIL can be constructed if the impact of Big Data on the Business Architecture, Application Architecture, Data Architecture and Technology Architecture is reflected in EA. This paper discusses a framework to construct the EIL for an organization before Big Data solutions can be implemented through understanding different subset architecture such as Business Architecture, Data Architecture, Application Architecture, and Technology Architecture to construct a holistic view. The business goal for Big Data is to optimize customer experience and gain insight upon future products and service, which is consistent with the purpose of EA to bridge the current state and to-be state of business.

The remainder of the paper is structured as follows. Section II describes EA and its role in Big Data solutions. Section III describes an overview of the Entire Information Landscape. Section IV describes Big Data and its impact on EA. Section V describes the framework to construct the EIL. Finally Section VI concludes the paper.

II. EA AND ITS ROLE IN BIG DATA SOLUTIONS

An EA is a conceptual blueprint that defines the structure and operation of an organization. The intent of an EA is to determine how an organization can most effectively achieve its current and future objectives. The Gartner definition of EA is as follows: “Enterprise Architecture is the process of translating business vision and strategy into effective enterprise change by creating, communicating and improving the key requirements, principles and models that describe the enterprise's future state and enable its evolution” [14].

The scope of the Enterprise Architecture includes the people, processes, information and technology of the enterprise, and their relationships to one another and to the external environment. According to Gartner Enterprise Architects compose holistic solutions that address the business challenges of the enterprise and support the governance needed to implement them. Integrating Big Data is a challenge for many organizations. EA can help develop the holistic solution to address this challenge. There are four types of architecture that are the subsets of overall Enterprise Architecture:

- A Data Architecture - this describes the structure of an organization's logical and physical data assets and data management resources. Data governance architecture, Master Data Management (MDM) & Reference Data Management (RDM) Metadata Management, and Data Security come under this.
- An Application Architecture - this kind of architecture provides a blueprint for the individual application systems to be deployed, their interactions, and their relationships to the core business processes of the organization. Reporting Architecture, Integration Architecture, Application Portfolio Management, and Application Security come under this.
- A Technology Architecture - this describes the logical software and hardware capabilities that are required to support the deployment of business, data, and application services. This includes IT infrastructure, middleware, networks, communications, processing, standards, etc.

If all four subset architectures of an EA can show the impact of Big Data, then the EA as a whole will render a holistic view of the information landscape and impacts of Big Data. Big Data solutions can be seen from many perspectives such as:

- Storage: the requirement for lots of storage;
- Distribution: Data stored in lots of places globally;
- Database design: lots of rows, lots of columns, lots of tables;
- Algorithmic or mathematical: lots of variables, lots of combinations, lots of permutations, summed up as one optimal answer among a large number of possibilities.

There are varieties of Big Data Technologies. In distributed processing technology there could be parallel processing, loosely or tightly coupled processing. In compression technology there could be data encoding and least number of bit data function encoding. In proprietary hardware technology there could be performing algorithms at the data persistent layer, massive parallel platforms, networks and quantum computer platforms. In reduced code set, it could be eliminating large amounts of DBMS code and eliminating large amounts of Online Transaction Processing (OLTP) code. The issue is if these technologies and requirements are not depicted in EA the Big Data solution would be uncoordinated, inconsistent and complex to handle. An organization needs to understand its objectives and goals in using Big Data technologies and solutions as they do not come cheap. Handling all the technologies in an organization adds to the complexity of IT assets in the organization and can give rise to uncoordinated Big Data solutions which will cost time and money to the
organization. EA can address the issue of uncoordinated Big Data solution by:

- Assisting the organization in opportunity recognition;
- Architecting the holistic approach of information architecture;
- Evaluating how applications will integrate knowledge via BI, PA and SOA;
- Updating the organization’s infrastructure technology roadmap and watch list [11].

III. OVERVIEW OF THE ENTIRE INFORMATION LANDSCAPE (EIL)

Entire information landscape deals with data landscape, technology, users and applications within and outside the organization which are required for the operations of the organizations. The EIL is shown in Figure 1. There could be number of applications in the organization such as Supply Chain Management, Human Resources, accounting which could be categorized in internal, external, applications on cloud etc., Big Data technologies are Hadoop, Greenplum, Cloudera, etc., Data landscape consists of Big Data, data in motion, file system etc., and users are internal, external, business etc… as shown in Figure 1.

![Figure 1: Overview of the Entire Information Landscape](image)

When EA is being constructed it provides Organization catalog, Business Service/ Function Catalog, Objective catalog, Role catalog etc. under Business Architecture. Under Data Architecture, EA provides Data component catalog, Data Entity, Application/Data matrix, Conceptual data diagram, Logical Data diagram, Data dissemination diagram, Data Security Diagram, Data Migration Diagram etc.. Under Application Architecture, EA provides Application portfolio catalog, Interface catalog, Application/ function matrix etc.. Under Technology, EA provides Technology standard catalog, Technology portfolio catalog, Application/ Technology matrix etc… The Application/Technology matrix documents the mapping of applications to technology platform. This matrix should be aligned with and complement one or more platform decomposition diagrams. The Application/Technology matrix shows: Logical/Physical Application Components, Services, Logical Technology Components, and Physical Technology Components.

There are varieties of Big Data Technologies. They can be summed up as: Database Technologies; Hardware Technologies; Applications; and Reporting Tools. There are two open source foundations. The first is Apache Software Foundations (ASF). The ASF provides support for the Apache community of open-source software projects. The second is Free Software Foundation (FSF). The FSF promotes the universal freedom to study, distribute, create, and modify computer software, with the organization’s preference for software being distributed. The FSF was incorporated in Massachusetts, USA, where it is also based [9].

Apache Hadoop is an open source distributed software platform for storing and processing data. Written in Java, it runs on a cluster of industry-standard servers configured with direct-attached storage. Using Hadoop, petabytes of data can be stored reliably on tens of thousands of servers while scaling performance cost-effectively by merely adding inexpensive nodes to the cluster. Hadoop Distributed File System (HDFS) splits files into large blocks (usually 64MB or 128MB) and distributes these blocks amongst the nodes in the cluster. For processing the data, the Hadoop Map/Reduce ships code (specifically Jar files) to the nodes that have the required data, and the nodes then process the data in parallel. This approach takes advantage of data locality [10] in contrast to conventional HPC architecture which usually relies on a parallel file system (compute and data separated, but connected with high-speed networking). Applications use an assortment of Big Data technologies such as Open source, Proprietary, Traditional databases (OldSQL), Non-traditional (NoSQL), OLTP (New SQL) fully ACID, etc.. Database architectures for Big Data technologies are provided by EMC Greenplum, IBM’s InfoSphere BigInsight, Microsoft’s Big Data Solutions, and Oracle’s Big Data Appliance.

**EMC Greenplum:**

Greenplum Unified Analytics Platform (UAP) is a unified platform enabling agile Big Data Analytics by empowering data science teams to analyze structured and unstructured data in a unified platform. It comprises of three components: the Greenplum MPP database, for structured data; a Hadoop distribution, Greenplum HD; and Chorus, a productivity and groupware layer for data science teams.

**IBM’s InfoSphere BigInsight:**

IBM’s big data platform includes software for processing streaming data and persistent data. BigInsight supports the latter, while InfoSphere Streams supports the former. The two can be deployed together to support real-time and batch analytics of various forms of raw data, or they can be deployed individually to meet specific application objectives [10].

**Microsoft’s Big Data Solutions:**

Microsoft’s Big Data solution brings Hadoop to the Windows Server Platform and in
elastic form to the cloud platform Windows Azure. The Hadoop Hive data warehouse is part of the Big Data Solution, including connectors from Hive to ODBC and Excel.

**Oracle Appliance:** Oracle’s approach caters to the high-end enterprise market, and particularly leans to the rapid deployment, high performance end of the spectrum. It is only vendor to include the popular R analytical language integrated with Hadoop, and to ship a NoSQL database of their own design as opposed to Hadoop HBase.

Along with Greenplum, Aster Data, ParAccel, and Vertica are early pioneers of Big Data Solutions before the mainstream emergence of Hadoop. All these solutions use MPP (Massively Parallel Processing) analytical database and have Hadoop connector available.

NoSQL is Not Only SQL is a new way of thinking about databases, founded on the belief that a relational database model may not be the best solution for all use cases and situations which requires Big Data solutions. The providers of NoSQL Platforms are Redis, riak, CouchDB, membase Cassandra to name few. NewSQL is a class of modern relational database management systems that seek to provide the same scalable performance of NoSQL systems for online transaction processing workloads while still maintaining the ACID guarantees of a traditional database system. The providers of NewSQL are VoltDB, Clustrix, xeround, nuoDB memsql to name but a few. ACID means Atomicity (Transaction cannot be subdivided), Consistency (Constraints don’t change from before transaction to after transaction), Isolated (Database changes not revealed to users until after transaction has completed), and Durable (Database changes are permanent).

The Business Service/Function catalog can be used to identify capabilities of an organization and to understand the level that governance is applied to the functions of an organization. This functional decomposition can be used to identify new capabilities required to support business change or may be used to determine the scope of change initiatives, applications, or technology components. The Business Service/Function catalog contains the following metamodel entities: Organization Unit, Business Function, Business Service/Function, and Information System Service. The purpose of the Functional Decomposition diagram is to show on a single page the capabilities of an organization that are relevant to the consideration of an architecture. By examining the capabilities of an organization from a functional perspective, it is possible to quickly develop models of what the organization does.

The purpose of the Data Entity/Business Function matrix in Data Architecture is to depict the relationship between data entities and business functions within the enterprise. Business functions are supported by business services with explicitly defined boundaries and will be supported and realized by business processes. The mapping of the Data Entity-Business Function relationship enables the following to take place: Assign ownership of data entities to organizations; Understand the data and information exchange requirements business services; Support the gap analysis and determine whether any data entities are missing and need to be created; Define application of origin, application of record, and application of reference for data entities; Enable development of data governance programs across the enterprise (establish data steward, develop data standards pertinent to the business function, etc.).

The purpose of the Application/Data matrix in Application Architecture is to depict the relationship between applications (i.e., application components) and the data entities that are accessed and updated by them. Applications will create, read, update, and delete specific data entities that are associated with them. The data entities in a package packaged services environment can be classified as master data, reference data, transactional data, content data, and historical data. Applications that operate on the data entities include transactional applications, information management applications, and business warehouse applications.

The Technology Standards catalog documents the agreed standards for technology across the enterprise covering technologies, and versions, the technology lifecycles, and the refresh cycles for the technology. Depending upon the organization, this may also include location or business domain-specific standards information. This catalog provides a snapshot of the enterprise standard technologies that are or can be deployed, and also helps identify the discrepancies across the enterprise. If technology standards are currently in place, apply these to the Technology Portfolio catalog to gain a baseline view of compliance with technology standards. The Technology Portfolio catalog contains the following meta-model entities: Platform Service, Logical Technology Component, and Physical Technology Component.

IV. BIG DATA AND ITS IMPACT ON EA

Big Data influences EA design in many ways. Big Data solutions have impact on Business Architecture, Data Architecture, Application Architecture and Technical Architecture of EA.

A. Big Data and impacts on EA Application Architecture

Application architecture is the science and art of ensuring the suite of applications being used by an organization. The application architecture forms pillars of an enterprise architecture or solution architecture. The application architecture shows the internal structure of an application, for its software modularization [15]. The application architecture also show the interfaces that the application provides for integration as well as the integrations to other applications and data sources the application requires to function [12].

IBM Systems Application Architecture provides Document content Architecture which specifies the format for documents to be exchanged among different word processors and other software, Systems Network Architecture Distribution Services for store and forward document transmission, IBM Distributed Data Management Architecture for file sharing and Distributed Relational Database Architecture for sharing relational databases [13]. Application Architecture focuses on defining the applications within an enterprise. This includes identifying how applications interact with each other and the other elements
within the Business Architecture. Application Architecture has number of dependencies such as:

**Services and Processes:** This dependency includes the services provided by the organization, the customers, core processes, value chain, process cuts (insourcing, outsourcing), work packages; tasks involved in the organization etc… Implementation of Big Data solutions will have an impact on Services and Processes of the organization. Services and Processes need to be mapped to each other so that Big Data solutions can be integrated into EA. In services, a Big Data solution will impact client servicing such as Customer Relationship Management (CRM), interactions and dependencies with other systems such as Order Entry/Order Management (OE/OM) etc., factories such as cash/payment transaction factory, security transaction factory etc., cash accounts, custody accounts, wealth management, client reporting, and the way business administration works within the organization. It may also have impact on other service providers which are collaborating with the organization implementing a Big Data solution. In process flows the organization needs to see how client data is being administered so that order cancellations do not occur due to incorrect client data.

**Component Interaction:** The software application interacts with a number of components such as Order management, CRM, Portfolio management, HR and payroll, reporting, etc. There could be case of data overlap. Portfolio management can also have the same data as CRM. A Big Data solution will impact many components in the organization such as Order management, CRM, etc…

**Functional Cluster and Architecture:** This deals with data storage and process execution. Big Data solution will impact the way applications and services are configured as resources on the cluster.

**Value Propositions, Capabilities, and Requirements:** This focuses on client experience, organizations involved, products and services provided, to provide consistent quality over time, automation of the processes, etc. A Big Data solution will have an impact on business value propositions, solution/system capabilities and system architecture requirements.

**Guiding Principles:** This dependency includes strategic principles such as package usage, business principles such as insourcing, technical principles such as platform technology, architecture principles such as service architecture and, data redundancy. A Big Data solution will impact strategic principles, business principles, technical principles and architecture principles.

Application Architecture within organizations needs to integrate Big Data stores, which could be HBase, NFS, HDFS etc… Figure 2 shows an overview of the impact of Big Data solutions on Application Architecture. An integration layer required. An organization needs to identify which applications can be integrated with Big Data solutions. There could be a possibility that not all applications can be integrated with Big Data solutions.

Big Data sources could be from E-commerce and Market Intelligence such as search and user logs, it could be from E-Government and politics such as government information and services; it could be from Science and Technology such as sensor and network content, etc…

B. Big Data and impacts on EA Technology Architecture

The Technology Architecture describes the hardware, software and infrastructure environment that is required to support the development and host the deployment of the application components described in the Application Architecture. Big Data solution introduces new technologies in the organization which may not have been in the organization previously. There will be a need for a new type of storage which can support the Big Data processing needs. Applications with Service-oriented EA in the cloud are emerging and will shape future trends in technology.

Technology domains are based on existing technology, hardware and physical infrastructure categories used. Big Data applications span a broad range of domains which requires change in technology domain of the organization. Expected technology domains are: Service-Oriented Enterprize Architectures for Services and cloud computing, and related Service-oriented technologies and management [19]. The technological impact of the Big Data from cloud vision has multiple aspects. Organizations are extending their technological capabilities to systematically manage their Business Operating Model [17] by developing and managing Enterprize Information Architecture as the architectural part of IT Governance [18].

Technology Platforms include Mainframe, Linux, Microsoft; Big Data solutions need to be hosted on a technology platform. Organizations need to assess the existing technology platform for Big Data solutions. Integration Patterns include Enterprise Service Bus, Message Queueing Web Services (SOAP and REST). Persistent Data storage are the data artifacts such as relational databases, database tables, flat files, xml data stores, message queues etc. that hold persistent data. With the rise of

---

Figure 2: Overview of impact of Big Data Solutions on Application Architecture
personalized and social services, restructuring is occurring in the existing Internet service environment. From the Internet Web service environment that focused on searching/portal sites, there is a demand for personalized and social services in the whole service area, such as telecommunications, game, music shopping etc. This is why scale-out technology is becoming more important than scale-up storage. Moreover, for complicated functions, storage size, and processing requirements, data processing technology beyond the scope of OLTP is becoming increasingly important [20].

C. Big Data and impacts on EA Data Architecture

Behind any information management lies the core doctrine of Data Quality, Data Governance and Metadata management along with considerations for Privacy and Legal concerns[5]. Big Data needs to be integrated into the EIL and not seen as a standalone effort.

<table>
<thead>
<tr>
<th>Information Goal and Principles</th>
<th>Entire Information Landscape Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Governance</td>
<td>Data Asset Planning</td>
</tr>
<tr>
<td>Data Quality and Management</td>
<td>Data Integration</td>
</tr>
<tr>
<td>Metadata Management</td>
<td>Data Models and Taxonomies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Master Data Management</th>
<th>Reference Data Management</th>
<th>Transaction Data Management</th>
<th>Structured Data Management</th>
<th>Unstructured Data Management</th>
<th>Big Data Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Warehousing Business Intelligence</td>
<td>Analytical Data, Documents and Contents</td>
<td>Historical Data</td>
<td>Temporary Data</td>
<td>Big Data Analytics</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: EA Data Management Framework

Data governance helps an organization to take a holistic view and to manage data in the context of business process, and to support application integration needs. Different components of data architecture include data modeling, data design, and data delivery architecture [6]. Data Architecture requires information about Master and reference data management, data warehousing business intelligence, transaction data management, structured technical data management, unstructured data management, Metadata, Analytical Data, Documents and contents, Historical Data, Temporarily Data, and Big Data analytics. All these data sources need to be organized in an information framework which can support data governance and data.

The key element of organizational structure is the business functions. Organizations need to know the kind of data required, captured, acquired, stored and analyzed for business functions such as marketing, sales, supply chain, manufacturing, human resources, strategy making, finance and, information technology, etc.. Data is an asset to the organization. This asset must be managed to ensure competitive advantage and to reduce the complexity of the asset management in the organization. Information Goals and Principles are top priority for any organization but this must be supported by Data Governance and Information Asset Planning [7, 8]. So while constructing the roadmap of Information Management Framework we require Data Governance, Information Asset Planning, Data Quality and Management, Data Integration, Information Life Cycle Management, Metadata Management, Data Models and Taxonomies and information about Master and Reference data Management, data warehousing business intelligence, transaction data management, structured technical data management, unstructured data management, Metadata, Analytical Data, Documents and contents, Historical Data, Temporal Data, and Big Data analytics. Figure 3 shows an EA Data Management Framework.

D. Big Data and impacts on EA Business Architecture

Business architecture is defined as “a blueprint of the enterprise that provides a common understanding of the organization and is used to align strategic objectives and tactical demands”. Business problems can be categorized into types of Big Data problems. This categorization could be used to identify proper fit of Big Data Solutions. Table 1 lists common business problems and assigns a Big Data type to each.

<table>
<thead>
<tr>
<th>Applications</th>
<th>Business Problems</th>
<th>Big Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-Commerce and Market Intelligence</td>
<td>Sentiment analyses</td>
<td>Web and Social Data</td>
</tr>
<tr>
<td>E-Government and Politics, Defense</td>
<td>Policy and Regulation analyses</td>
<td>Web and Social Data, Biometrics</td>
</tr>
<tr>
<td>Science and Technology</td>
<td>Utilities: Predict Power Consumption</td>
<td>Machine Generated Data</td>
</tr>
<tr>
<td>Customer service</td>
<td>Call monitoring</td>
<td>Human Generated</td>
</tr>
<tr>
<td>Retail</td>
<td>Personalized messaging based on facial recognition and social media</td>
<td>Web and Social Data, Biometrics</td>
</tr>
<tr>
<td>Healthcare</td>
<td>Fraud Detection</td>
<td>Machine Generated, Transaction Data, Human Generated</td>
</tr>
</tbody>
</table>

Categorizing the Big Data type helps in understanding the characteristics of Big Data and the technologies associated with it. The Big Data type allows the organization to understand its data categories (transaction data, historical data, or master data for example), the frequency at which data will be made available, how the data needs to be processed (ad-hoc query on the data for example), and whether the processing must take place in real time, near real time, or in batch mode.
V. FRAMEWORK TO CONSTRUCT THE ENTIRE INFORMATION LANDSCAPE

As discussed in Section II, it is not possible to drag and drop any Big Data solutions into an organization. Some of the major risks are:

- The way organizations interpret and manage the information.
- Finding the proper fit of Big Data solutions and technologies for an organization.
- Potential for data privacy violation.
- Potential for spending lots of money and time chasing poorly defined problems or opportunities.
- Big Data tools are evolving and are not taught in many universities.

These risks suggest that there is a need to construct an EIL which can depict Business, Application, Data and Technology issues in an organization. This can only be facilitated with a framework to construct the EIL. This requires a paradigm shift in Big Data solutions where constructing EIL should be treated as preliminary activity while implementing Big Data solutions.

With the changing paradigm of how data can be stored, acquired and analyzed Enterprize Architecture is required for the construction of the EIL. EIL addresses the change in data landscape, technology, users and applications within and outside the organization which are required for the operations of the organizations.

Transaction data reveals past information and may or may not have any bearing on future information. This can be resolved by analyzing social media data. While social media analysis can provide many useful insights, these have largely proved unsatisfactory when used as sole source for analysis because the data is often incomplete and mostly reveals correlation rather than causation [21]. Choosing the “Right Data” source for Big Data solution should enhance any existing analysis projects. It is not merely adding more data but understanding what variables are required for the analysis. Data scientists need to first identify what additional variables could refine or improve the analytical result. By understanding what kind of data is required for analysis will help in locating and choosing the “Right” external data sources.

- **Activity 1: Selection of data sources:** In this activity the “Right” data sources are picked. The value is found in layering data from multiple sources, not in simply adding bulk to the final data count [21]. Internal data sources are helpful in solving number of business problems but by working in isolation and applying analysis only to internal data, it may improve existing processes and products that are on their way to becoming obsolete or unprofitable. Data from different sources are required. However hoarding data either from external or internal sources drives up the cost and potential liabilities.

- **Activity 2: Selection of tools and technologies:** The value of Big Data is always found in the outcome. Organizations have been collecting and analyzing digital data for decades and the resulting storage issues and costs have been an ongoing money drain [21]. Unused data is an opportunity lost. However, buying additional tools and technologies beyond an organization’s existing analytics applications may not even be necessary depending on a project’s particular business goals and functions. The selection of tools and technologies best suited is based on different business functions of an organization to suit Big Data solutions.

- **Activity 3: Assessing the impact on EA to generate the EIL:** Each domain architecture links to the others either directly or indirectly at some point. Business Capabilities and/or Business Processes (Business Architecture), links to the Applications that enable the capability/process (Applications Architecture – COTS, Custom), links to the Information Assets managed/maintained by the Applications (Information Architecture), links to the
technology infrastructure upon which all this runs (Technology Architecture - integration, security, BI/DW, DB infrastructure, deployment model). Figure 5 shows all of the four subset architectures of EA and the impact of Big Data integration on each architecture to generate EIL. Customer service, a business function that involves daily customer interaction can be addressed by assessing the impacts on EA to generate EIL. For customer service business functions, the Data Architecture impacts are Data Quality, Data Needs and Data Flow to get the data that exists in other business functions and divisions such as sales. Data ownership is also impacted as data is being accessed from one division to another. For Technical Architecture impacts are technical infrastructure and hardware so that real time analysis can be done for decision making. For Application Architecture the impacts are integrating different applications which are useful in the generating the required analysis for the customer service. These impacts help us in generating EIL for business function customer service which conveys the requirement to the organization.

Big Data solutions can be implemented by many organizations by using the framework to construct the EIL. The EIL can serve as a single point to look for implementing Big Data solutions.

VI. CONCLUSION AND FUTURE WORK

Data continues a massive expansion in scale, diversity, and complexity. Data underpins activities in all sectors of society. To deal with these Big Data solutions are available and are now fairly commonplace in large organizations. Organizations want to store, process and analyze Big Data for intelligent decision making to get the competitive advantage. Implementing a Big Data solution typically begins as part of an information technology project that will extract, store and analyze large amounts of data in order to get competitive advantage. It is, however very difficult to simply ‘drop’ these solutions into existing IT infrastructure and expect them to run smoothly. There are many different tools and technologies associated with Big Data.

The organization needs to understand if they really require Big Data solutions and the impact it is going to have on the organization’s Enterprise Architecture. New ways of working with Big Data has impact on technology infrastructure components and data architectures. Since most data is directly generated in digital format today these impacts and changes must be captured by an organization’s EA for conducting enterprise analysis, design, planning, and implementation, using a holistic approach at all times, for the successful development and execution of organizational strategy. Our framework to construct the Entire Information Landscape may dramatically improve the way Big Data solutions are being implemented in organizations. It is well understood that at present Big Data solutions are evolving and all organizations are using Big Data as a buzz word without an understanding of the impact of using Big Data solutions. Organizations need to prepare themselves for Big Data integration. This can be achieved by examining their EIL. The EIL deals with the users, technology, data landscape and applications. These are in turn based on Business Architecture, Application Architecture, Technology Architecture and Information Architecture all of which are subsets of their EA. In this paper we have examined areas within the EA that are impacted by Big Data and our future work involves examining this in more detail to clearly identify, document and provide guidance as to the full impacts of Big Data on EA and enable organizations to understand and develop their Entire Information Landscape.

REFERENCES