IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

Data Governance – from Rocks to Diamonds

Jayant Dani[1], Sameer Rane[2]

Life member of CSI, LMISTE, Industry Advisor for Various College,
Principal Consultant at Tata Consultancy Services[1]

Head Presale, TCS MasterCraft Dataplus, Tata Consultancy Services[2]

Abstract: This Paper has been present various changes happen in to Data Governance due change in various new Data type, Data source and regulation changes. This will be describing holistic view of Data governance for future data system. Also it has related the same governance process with Diamond industry to emphasis on value of governance.

Keywords: Data , Data Governance, MasterCraft DataPlus, Enterprise Data Management

INTRODUCTION

Every enterprise source data from various sources and in various formats. There is lot of data available within the data stores of an enterprise, is growing exponentially and shared across multiple stake holders. Though the data is growing, organizations face challenges in managing this big data. Due to this, real value of data is not getting derived out of it. Along with this, organizations need to ensure good quality data that is correct, reliable, available on time and have uniform representation for informed strategic decision making.

There is also increased focus on data privacy globally and with various regulatory compliances and mandates in picture, organizations need to comply with data regulations by making sure the sensitive data is not leaked out of the organization.

Also, at regular intervals, organizations have to undergo audits and significant efforts are involved in collating data for audit requirements.

So considering all these factors, organizations need to have more focus on overall data governance on one day or the other. For this organizations may use any of the existing frameworks or have their own derived frameworks which would suit their needs.

In this article, we have tried to put our point of view on how the entire data governance program should be defined and handled for the benefit of larger audience.

GOVERNANCE

Let us recollect what governance is and have few definitions from best appropriate source according to us.

wikipedia.org: [1]
Governance relates to "the processes of interaction and decision-making among the actors involved in a collective problem that lead to the creation, reinforcement, or reproduction of social norms and institutions".

Governance is the way the rules, norms and actions are structured, sustained, regulated and held accountable.

businessdictionary.com: [2]
Establishment of policies, and continuous monitoring of their proper implementation, by the members of the governing body of an organization.

goodgovernance.org: [3]
Good governance is about the processes for making and implementing decisions. It's not about making 'correct' decisions, but about the best possible process for making those decisions.

The key takeaways based on the above Governance definitions:

- Interaction and Decision Making
- Accountability
- Policy Establishment and continuous monitoring for policy implementation

Data Governance:
With the above understanding of Governance, now let us also understand what data governance is. It is nothing but governance around the data. To understand it better, let us have a look at various definitions from well-known sources.

wikipedia.org: [4]
Data governance is a control that ensures that the data entry by an operations team member or by an automated process meets precisely standards, such as a Business rule, a data definition and data integrity constraints in the data model.

Data governance encompasses the people, processes, and information technology required to create a consistent and proper handling of an organization's data across the business enterprise.

Data Governance Institute (DGI): [14]
Data Governance is the exercise of decision-making and authority for data-related matters.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods

Informatica: [12]
The functional coordination and definition of processes, policies, standards, technologies, and people across the organization to manage data as a corporate asset. This enables managing the availability and controlled growth of accurate, consistent, secure, and timely data for better decision making, reduced risk, and improved business processes.

In accordance with above definition, data governance is a mechanism to protect and manage the data for organization to derive business value and reduce risks. Data governance need to be around data collection, data privacy, data provisioning and data quality. The data governance framework should be able to manage end to end processes, steam line business process definitions, implement policies with continuous monitoring and measure the success through metrics. It also involves defining clear roles, responsibilities and accountability.
For effective implementation of data governance initiative, it is important to have good data governance framework. Today in market, there is ample information available on various data governance frameworks. However, let us consider few data governance frameworks proposed by Informatica, SAS and Data Governance Institute (DGI) and try to map the key takeaways across 3 key attributes – People, Processes and Products and technologies.

People: For effective implementation of data governance program, organizations need to identify and involve people with rights skills and assign responsibilities, decision making rights and hold them accountable. To identify various key stakeholders we can use metrics like RACI (Responsible, Accountable, Consulted, Informed) or DACI (Driver, Approver, Contributors, Informed).

Process: Processes are nothing but ways to govern the data from perspective of data collection, privacy and protection, quality, data enrichment, preparation, dispensation and archiving as well as collaboration for decision making. All these processes needs to be standardized and documented for reuse. These processes need to be digitized in form of business rules, policies, workflows which can be applied to measure effectiveness and ensure data governance.

Products & technologies: In order to implement and execute process related to data governance, right tools and technologies are required for

- Data protection and privacy
- Data profiling and discovery
- Data quality
- Capture business glossary
- Capture Metadata
- Create and execute jobs
- Send notifications

There are various products available in market who caters to these needs independently. However, there is need to have integrated solution which would handle all these aspects of data governance, privacy and quality. Also with exponential data growth and diverse types of data – structured and unstructured, the products should be scalable enough to integrate and apply the existing rules and various business policies.

Example from Jewelry Industry
With all above conceptual knowledge, let us now consider a real life example of the diamond production and jewelry manufacturing process and try to map with the data governance.

Data Governance with focus on data quality:

Raw rocks ~ Raw Data
As diamonds are produced by mining the rocks, the valuable information is discovered from the raw data available with the organization.
As we know, diamonds are formed at high temperature and pressure in the Earth's mantle and available on earth's surface through deep volcanic eruptions. These so called diamonds are not in polished form and do not have market value. In order to derive the market value of these rough stones, it has to undergo various processes like cutting, polishing and re-examination. Similarly, in order to derive meaningful information out of data, it needs to undergo various processes like profiling, cleansing, standardization, enrichment and matching.
Figure 1 demonstrates the mapping between various processes involved with respect to jewelry manufacturing and data management

Figure 1 : Governance process mapping in jewelry and data

Now let us see at each process in details:
Jewelry Manufacturing Process [6, 7, 8]

1. **Identify land:** To start with, organization has to identify and invest in land for diamond exploration
2. **Mining/Exploration:** It is process of extracting rocks, rough diamonds and minerals from the land. This is

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

followed by sorting where rough stones are sorted further into different categories based on shape, size, color, clarity and quality.

3. **Cutting:** As part of this process, preliminary analysis of rough stones is done to identify and address various issues using relevant tools and techniques and cut to provide shapes.

4. **Polishing:** At this stage, rough diamonds are polished to remove further impurities.

5. **Re-Examination:** In this process, diamonds are reexamined to identify flaws, remove non-diamond inclusions like filling of the voids produced. The diamonds are further classified based based on cut, color, clarity and carat weight.

6. **Market Sale and Jewelry Manufacturing:** After repetitive process of cutting and polishing, the diamonds are finally available for sale and jewelry manufacturing.

Data Governance Process [12, 13, 14]

1. **Organizational Data:** Entire data within an organization itself is the area for the data governance program.

2. **Data Collection and Analysis:** The existing data within the organization is analyzed and diagnosed for segregations as rocks, rough diamonds and minerals (rocks – operations KPIs/BO reports, minerals – analytical KPIs/BI report and rough diamonds – new business assets which would have high value)

3. **Data Profiling:** As part of this process, detailed diagnosis of data is done by data stewards to check data for completeness, accuracy and consistency. The process helps in identification of various data elements with high business value which can be of internal or external use. The output of this process is the data quality scorecard which reflects the current data quality index. This data quality index helps in identification of data quality rules.

4. **Data Cleansing, Standardization and Enrichment:** Once the data is profiled and issues are identified, as a next step the data is cleansed by removing unwanted or inaccurate data. This is followed by standardization and enrichment by replacing inconsistent values with standard values and by filling missing values using reference data.

5. **Re Execution of Data Profiling and Quality Jobs:** In this process, the data profiling jobs are executed repeatedly to monitor the data quality score and clean, standardize and enrich the data till good quality data is achieved. After this process the data should be complete, accurate, consistent and trustworthy.

6. **Data Mining, Reports & Dashboards:** After repetitive process of data profiling and cleansing, good quality data is available for decision makers to make informed decisions. The reports and dashboards can be built on this data. Data mining activities can be performed on this clean data to identify the hidden information.

Data Governance with focus on data access and privacy:

Till now, we have considered the quality aspect of data governance. Let us also understand privacy aspect using the same example. As we know, after repeated process of proper cutting and polishing, market value of diamond increases. Similarly, after repeated data profiling and quality management processes, value of data also increases. So it is required, to have sufficient measures in place to protect this valuable asset.

As we know, in case of diamond production process, before commencing the diamond exploration program, the sponsors need to acquire land by investing certain initial amount. There are various risks associated with land like encroachment and trespassing. To protect land from these risks, it is important to have boundary around the owned land.

Also there is possibility that the rough diamonds may move out of the premises without permission, get polished and sold in the market. So, as part of security process, security guards do checking before any employee or contract worker leave the premises to ensure valuable stones or diamonds do not move out of the premises without proper permission by authorized person.

Similarly, in case of data, there is possibility that data would get accessed by unintended and unauthorized people who have access to data. They can analyze this data and get the value out of it for their benefits. To avoid this, proper controls should be there in order to ensure confidential and sensitive data is not accessed by unauthorized person.

In case of diamond processing, as physical movement is involved, physical security is required which can be automated up to an extent. However, in case of data, it can be physical movement like moving data from production instance to development or test instance or non-physical movement like providing direct production data access to support executives in order to support day to day business queries. In both cases, to ensure the privacy, the confidential and sensitive data can be masked or obfuscated using automated tools and processes. The respective masking jobs can also be scheduled to run without manual intervention.

To have proper governance in diamond industry various roles and responsibilities are assigned to individuals based on their skills and experience. Similarly, the roles and responsibilities are assigned to ensure data governance. Let us also look at the mapping.

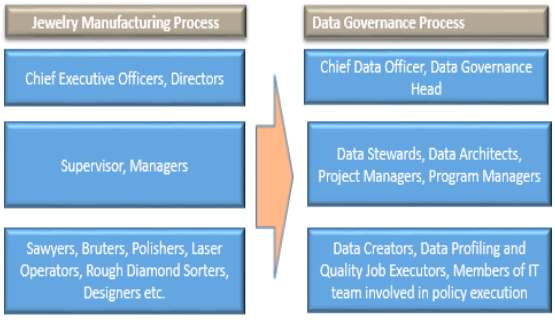

Figure 2: Role and responsibilities mapping in jewelry and data [9]

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 12, December 2016

Impact of Big Data

With advent of big data and analytics, enterprises are more focused towards adopting and investing in these technologies for the competitive benefits and organizational growth. There is need to store different types of data – traditional (structured) and modern (unstructured data like social media, click streams, machine generated), integrate and present in common format to act on it. Unless we know the existing data, its sources, lineage, its quality, it would be difficult to establish integration. Traditional approach of data governance may not work with this modern data and it needs to be scalable. For having data governance around big data, one of the important aspect is to analyze and understand the data as it is not readily available as in case of structured data.

CONCLUSION

To summarize, in order to stay ahead in this completive world, it is imperative to have more focus on data, consider data as an asset, analyze and maintain good quality data, understand its entire lifecycle from collection till retirement. With all this in place, the only step remains is to act on this data to derive hidden information from the data and reap benefits for any organization level initiatives like having data governance program or designing and establishing Data Lake.

It is also important to understand that the data governance is not one time activity and is continuous activity. So once the complete, correct and trustworthy data is available, it is also important to stay clean by ensuring the data quality at the time of data entry or data capture into the enterprise. There should be sufficient control over the data access. To have correct data lineage, the metadata should be maintained whenever there are addition or removal of applications and systems.

In order to implement data governance framework and to digitize this framework, there is also need to have integrated platform to configure business glossary, people, rules, define metadata, ownership, define and monitor policies, define workflows for overall privacy, quality, metadata management, visualize data lineage and all related activities which are part of data governance program.

Also since every organization has some sort of data governance in place and may be using tools and technologies for managing the data governance, the integrated platform should have capabilities to integrate and co-exist with the existing tools and technologies and also scalable enough to address future need of exponentially growing data and diverse technologies.

REFERENCES

[1] Article: Governance - https://en.wikipedia.org/wiki/Governance
[2] Article: Governance - http://www.businessdictionary.com/definition/governance.html
[3] Article: Data Governance - http://www.goodgovernance.org.au/about-good-governance/what-is-good-governance/
[4] Article: Data Governance - https://en.wikipedia.org/wiki/Data_governance
[5] Article: Data Governance - http://searchdatamanagement.techtarget.com/definition/data-governance
[6] Article: Diamond Production Process - https://en.wikipedia.org/wiki/Diamond
[7] Article: Diamond Production Process - http://www.photius.com/diamonds/the_diamond_industry.html
[8] Article: Diamond Production Process - http://www.swissgemlab.com/EducationPages/EducationDetailPage.aspx?pcid=324&AspxAutoDetectCookieSupport=1
[9] Career Opportunities Document: Northwest TerrItorIes Education, Culture and Employment Canada - https://www.ece.gov.nt.ca/files/Career-Employment/Career-OppSeries/Jobs_In_Diamonds.pdf
[10] Document: A roadmap to effective data governance: HIMSS Clinical & Business Intelligence Committee - http://www.himss.org/roadmap-effective-data-governance-how-navigate-five-common-obstacles
[11] Article: Common data governance mistakes: Rick Sherman - http://searchdatamanagement.techtarget.com/feature/A-must-to-avoid-Worst-practices-in-enterprise-data-governance
[12] White Paper: Data Governance Framework: Informatica - https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/white-paper/holistic-data-governance-framework_white-paper_2297.pdf
[13] White Paper: Data Governance Framework: SAS - http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-data-governance-framework-107325.pdf
[14] Document: Data Governance Framework: DGI: Gwen Thomas, The Data Governance - Institute http://www.datagovernance. com/wp-content/uploads/2014/11/dgi_framework.pdf
[15] Data Governance Processes: Dan Power - https://www.informatica.com/downloads/7126_data_gov_web.pdf
[16] Article: Proactive Data Governance: Henrik Liliendahl - https://liliendahl.com/2011/07/28/proactive-data-governance-at-work/

BIOGRAPHIES

Jayant Dani, Chief Architect and Principal Consultant has over 18 years of industry experience. He played a leading role in setting up the Big Data (Hadoop ecosystem) practice at Tata Consultancy Services (TCS). Dani is currently spearheading the conceptualization, architecture, design, and engineering of TCS MasterCraft Data Plus, an integrated data management platform. His area of expertise includes technical leadership of large solution teams, client relationships, and management of large scale implementations. Jayant Dani is also passionate about architecting technology solutions across industry verticals and has many publications to his credit, including a paper on Big Data management presented at Institute of Electrical and Electronics Engineers (IEEE) conference.

Sameer Rane, Consultant has over 13 years of industry experience. Played leading roles in product maintenance, product implementation and project management. His area of expertise includes technical leadership of solution delivery teams and client relationships. Sameer is currently playing role of presales consultant of TCS MasterCraft DataPlus, an integrated data management platform.