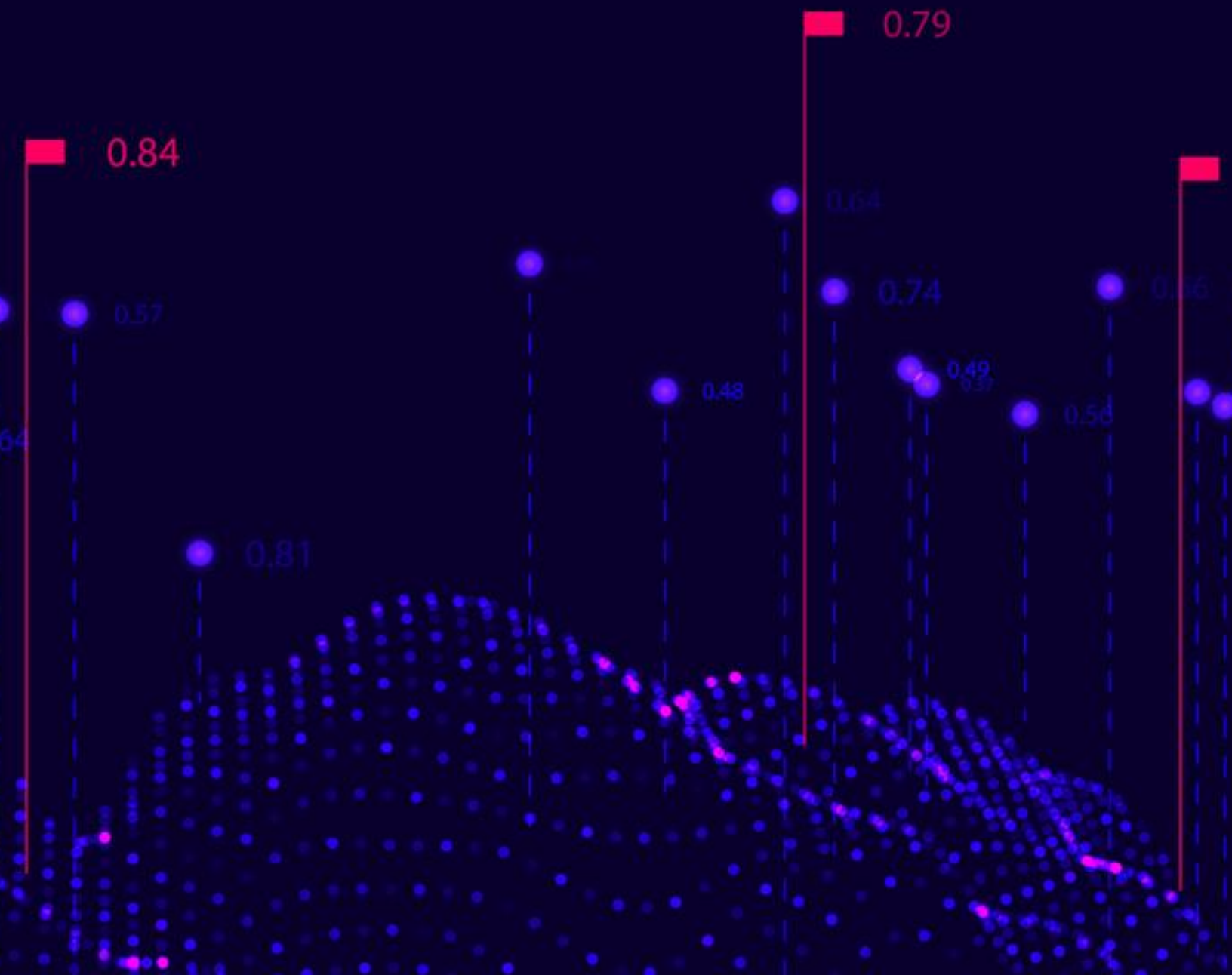Tech
Mahindra

ENTERPRISE
METADATA
MANAGEMENT

Water, **water everywhere**, / **nor** any **drop to drink**. Lines from "The Rime of the Ancient Mariner," by Samuel Taylor Coleridge. The speaker, a sailor on a becalmed ship, is surrounded by salt **water** that he cannot **drink**. Having spent decades in the Data and Analytics domain and having seen it evolve, I can draw a parallel to this age old line - "Data, Data Everywhere; not any value to draw"

We went through a long era of data Collection and MIS Reporting. We are now entering; in fact, a lot of organizations are already into "Data of Data" a.k.a "BI of BI". In very simple and abstracted terms, BI on BI is 'Effective Management of Enterprise Metadata', though analysts can define it in a crisper and all-encompassing manner.

ALCTS (Association for Library Collections and Technical services) created a Task force for Metadata. The formal working definitions for the three key terms deliberated and submitted by the task force subcommittee is listed below; although varying definitions for each concept appeared to depend on intent and context informing and surrounding the definition

- **Metadata** are structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities.
- **Interoperability** is the ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system.
- A **metadata schema** provides a formal structure designed to identify the knowledge structure of a given discipline and to link that structure to the information of the discipline through the creation of an information system that will assist the identification, discovery and use of information within that discipline.

I am attempting to fuse these definitions to have more connected overview (Referred from link below)-

The structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information object or information system for purposes of description, administration, legal requirements, technical functionality or preservation. - Metadata is often called – "data about data" or "information about information" or "structured data about data."

Now let us take a real world example that demonstrates some of the challenges and subsequent opportunities.

A Large Telco will have various lines of businesses e.g. Wireless Prepaid, Wireless Post Paid, Enterprise, products bundled with handsets, Wired and Wireless Broadband, Wireline etc. The business and products that a Telco sells would have been incrementally developed or acquired over a long period of time.

To understand the stored information related challenges let us take the customer profile as an entity in context.

- A Telco would have added the customers and stored their profile over a long period of time. The application stack which serves Post Paid customers and Prepaid customers could be different from each other.
- The customers might be availing wireline or wireless products or both, but if the stack is different, which is most likely the case, the customer's information would be stored in different stacks and in different ways in different formats, in different quality and with different quality KPIs with some variation in the attributes.
- If the businesses were acquired ones and not home grown, again, the application stack would be very different and so would be the way information is stored. In general, the standards to build the applications in two different organizations are not similar.

Another dimension is that every enterprise rushed to create data lakes and DWHs to collect the Data from various sources from where ever they felt data can be useful and have done well in data collection. To have more contextual data, every department of the organization started to build localized Data Warehouses and Lakes for their own requirements due to which department wide DWHs and Lakes mushroomed. The Data Landscape became too complex and too much to handle. The icing on the cake is that different departments are holding and owning the data and they keep it close to the chest. A lot of credit goes to the insecure middle management for gaining controls and scoring some points with their seniors. Remaining challenges are now posed due to requirement of privacy of data, Security issues, CCPA , GDPR compliance etc. One more critical question is how to make data easy to find for non-technical users? So, in a nutshell, while the data has increased and empowered the businesses, the complexities have also increased.

Now comes Mr CEO and Sales Head, who want to monetize the data , cross-sell, up-sell, provide consistent customer experience to their customers and much more, and rightly so! The CEO understands, and all the people lower in the hierarchy will nod their heads in agreement (I will do it too!) that the organization has so much valuable data, however, the CDO is lost with the amount and variety of the data. This is because the applications were either built incrementally at different period of time or acquired in a merger, or developed under different norms are not technologically consistent with underlying guidelines. Data Analysts, Data Scientists, Data Engineers will spend lot of unproductive time finding out which systems have right information which they need to draw value from the information

The IT Manager gets tasked to build a sophisticated model which can fulfill the demands of Mr. CEO to help business flourish. To start building the Cross-Sell up-sell initiative, he needs to tap the information of the customer from the various stacks and ensure right attributes are taken so that there is a good amount of accuracy to select the right customer and sell the right product. The IT Manager will have a lot of challenges in front of him:

- Where is the information of customers stored? Which application Data Base is it in?
- How is it stored e.g. Length, name of the column?
- Is all the required information (e.g. email id, demographics, mailing address, billing address stored in one place within the stack? or, is it stored in multiple systems for different Lines of Business?
- Is the stored information sufficient for the Model?
- Is the data field carrying the Master information?
- Does the field have required quality of information? Is it getting updated with required frequency and not getting updates via multiple channels?
- Where can I have knowledge to get a basic understanding of the data field?
- How do I get a Data Set for my analysis and how are they interrelated?
- How is the information connected while fetching the information? What kind of complexities can be faced while fetching the information? Am I allowed to do it? Is it as per the regulatory compliances?
- When was the data created, by whom was it created and which department created it for what purpose?
- Who can access the data? Are Customer preferences defined for cross-sell and upsell? Was it even captured?
- What is the latency/frequency required
- Who is the end user and consuming application
- Is data gravity higher on premise or on the cloud

A lot of effort goes into searching, identifying and building the above if there is no central repository of the Medadata which can help to reduce the overall effort. Do keep in mind that I have taken a simple use case and there are hundreds of complex use cases which need information about various entities.

The model which could have been done in 2 weeks, now takes 6 months and without assurance of its accuracy.

The CDOs are lost with the amount and variety of the data. Applications that are built incrementally at different period of time, acquired due to merger, developed under different norms, are not consistent, and neither are the underlying technologies. Data Analysts, Data Scientists, Data Engineers will spend a lot of unproductive time finding out the correct systems having required information which they need to draw value from.

Definitely, an Enterprise Metadata platform with the right structure and information can solve some part of the problem. I have had the privilege of meeting various customers across geographies. Most of them are trying to build an Enterprise Metadata Management platform but are in different states of maturity and completion. They are planning to use different toolsets but struggling with the requirement and strategies to tackle. Effective Enterprise Metadata management is very important for an organization to draw the value from the data and step towards becoming a data driven organization. Even the Data Lakes and Data Fabrics which are built within the organizations when any applications have to use the data or build Analytical models, need the right source and information from these sources.

Plethora of issues making Metadata Management implementations difficult

Below are some of my observations on the key issues plaguing this domain

- Absence of a central team that is losing sleep for creating a strategy and ensuring the collaboration across the organization and system owners. Who will bell the Cat? It is nobody's problem apart from CXOs.
- The applications are built in a distributed manner and acquired over a period of time which makes it very difficult to build an inventory of Metadata. If the applications were part of an acquisition, they would have been developed under different standards and technologies. It becomes very difficult to appreciate how all information can be unified and how should the Metadata repository be built.
- The owners of the applications are dedicated to their own applications and don't appreciate the requirement of a robust Enterprise Metadata Management solution or strategy. They often don't get together to discuss and build a central system.
- Due to regulatory requirements, the systems are more restrictive to sharing information; hence, the right approach is needed right from the word go.
- There are lot of security threats and privacy requirements. These challenges make source systems cautious and very restrictive.
- The requirement of accessing the correct data and building cross-functional analytics or getting deeper insights, with appropriate linked information is becoming very important for the CXO communities. However, nobody has a complete understanding of the ecosystem and how to connect.

So what do we do?

Organizations should have a clear strategy to ensure there is a focused Metadata management approach. Success lies in understanding the bigger picture and ensuring that all these factors are considered while building it.

- The platform's requirements and benefits should be understood by the CXOs and department heads. Without participation and collaborative approach, it will fail. As part of the Data Business, we have seen a few large organizations struggling to sell this platform internally and obtain a common agreement on the ownership of such platforms even if agreed upon. Even if it is built, adoptions becomes a problem. It is more of a political angle within organizations.

- We should ensure that there is a top down approach and the key roles are very clearly agreed, defined and EMPOWERED.

    - Ideally, it should be under CDO's ownership and empowerment

- Data Steward roles for central team who can work closely with the Data Stewards of individual stacks for process and quality
- Central Architecture and Technical Lead/Steward
- Central Program Manager to work with Individual stack or source data owners

- The overall strategy should include following topics covered.

  - Core team and org structure – Participation from all stack owners
  - The challenges / business problems which will be solved
  - Clear goals and objectives – Incrementally built
  - Metadata usage plans – Who and how will it be consumed
  - Selected tools and features to achieve the goals
  - Source, elements, standards and processes
  - Data Quality – extremely important
  - Metadata storage and products
  - Adoption and training plans
  - Measurement and success Criteria
  - Project Plans – Major Sprint plans

- As we are proponents of Design Thinking and Usecase led development, we feel that a good amount of time should be spent during the initial foundation build, with right participation. Good amount of time should be spent on the idea of how the platform will be built and leveraged across the Organization. Big Bang approach most likely fails and is difficult to course correct. Every business has different requirements and those should be carefully understood before venturing into the journey.

- The existing Metadata across the applications should be understood with the help of source owners so that when you build an organization wide Metadata platform, it should become a one stop shop rather than being questioned later about the applicability of the platform.

- One of the core requirements that such platforms can help us is with the Data Lineage. However, unless it is necessary in 1st phase, it should be avoided during initial phase. Data Lineage needs significant involvement of the individual application stacks and departments team. The technologies are not as open as those needed for true Data Lineage systems to be built. The ETLs and other tools don't expose their Metadata.

- The overall Governance is very important not only for execution but also for building processes to control inclusion, exclusion, monitoring, access, security and usage of the Data Elements. With all Metadata in one place, it gives an enormous empowerment, understanding and reach to the enterprise information in one place. Hence, governance and security is of umpteen importance.

- The platform should comply to requirements like GDPR, CCPA etc. These requirements should be well understood before new implementations.

- The cross functional departments and analytics team with Data Scientists would be the key users of the enterprise metadata platforms. The key set of users should be identified while building any such platforms. Their requirements can help to ensure that the platform build is heading in the right direction and, after build, it would be easily adopted.

- A good amount of time should be spent on mapping the requirements to the tool's features. Please check with the tool vendors on their roadmap and the speed of new feature's delivery. How easily

can customization be done? How easily can it be connected with the source applications? What kind training plans and technical support can be provided by the tool vendor?

To reiterate, it's very important to ensure that there is a Core team with **empowerment**. The central team should be well equipped with the understanding of the Metadata Management processes, regulatory compliance requirements, business requirements and tool's features to be implemented. The initiative should be equally owned and participated by the department heads. The progress and usage should be reported into CDO and CEO of the organization to bring required ownership and focus.

So what will we achieve?

Managing the Metadata effectively can help Organizations in various fronts

1) Effective and planned Metadata management can help organizations to become more effective Data driven organizations by having better insight about their assets

2) The platform, once implemented, can help to understand the redundant data which can be eliminated to simplify the data estate.

3) Easy and deep understanding of the data will convert it to the asset. The platform can help in monetizing the data efficiently.

4) The effective implementation of the platform can help to understand the GDPR, CCPA and other compliances related gaps. It will be easier to narrow down and prioritize the fixes. E.g. It will be easy to understand where all Customer information is stored; where are the places there are fields to store the user preferences.

5) Once you have the inventory of the Metadata, the organizations should quickly move to the Data Lineage. It can help in streamlining, plugging holes and easy debugging of problems.

6) Every organization wants to have deep analytics capabilities, and rightly so! At the same time, it is important to solve the Cross domain analytics challenges which can be made easy by having enterprise level Metadata management.

7) Data Scientists are a set of users who are spending more time finding and understanding the data fields and wrangling. They can better spend time in actual Algorithm build and test rather than spending time in Data Discovery and data set build.

Over a period i have given my views and experiences on the life and death of Data, Data Driven Organizations, Data Ware Houses, Data Lakes etc. Data management comprises of different aspects such as data integration, data quality, master data management, data governance etc., and metadata management is the aspect that stitches all of them together

About the authors

Manish Kampani, Global Competency Delivery Head | BI, Big Data and Data Management Service – CME, Tech Mahindra

Manish has 20+ years of experience in setting up Practice, running Delivery, Operations and Business Development across the globe. He has deep understanding of the Data and Analytics domain which has evolved over time and has exposure in Big Data, Machine Learning and Artificial Intelligence. He currently serves as a Global Competency Delivery Head for BI, Big Data and Data Management services in Tech Mahindra, focusing on Communications, Media and Entertainment verticals. He is passionate about co-creating solutions with his customers by leveraging not just the leading Data and Analytics products of the market, but also collaborating with the thriving startup ecosystem in this space. You may reach out to him at (@ManishKampani)

Arundeep Sivaraj, Group Manager - Presales | BI, Big Data and Data Management Service – CME, Tech Mahindra

Arundeep has 14+ years of experience in Product Management, Presales and Marketing of enterprise products, solutions and services for the Communications domain. He has built products grounds up from concept to market and lead multi-million dollar consultative selling initiatives. He has experience in BSS, OSS and Data & Analytics solutions. He currently heads Presales initiatives for BI, Big Data and Data Management services in Tech Mahindra, focusing on Communications, Media and Entertainment verticals for the Americas region. He is passionate about working with customers to unlock custom value propositions that provide them with the edge in their business. You may reach out to him at (@arundeepsivaraj)

References: -

https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html

https://www.niso.org/standards/resources/UnderstandingMetadata.pdf

http://dublincore.org/documents/2005/05/26/usageguide/glossary.shtml