Data Governance

Lesson 9

KEY CONCEPTS

■ Data Governance ■ Data Management ■ Business Value of Data Governance ■ Data Governance Quality Index ■ Top-down method ■ Bottom-up method

Learning Objectives

To understand:

- Genesis of data governance
- Meaning of data governance
- Difference between data governance and data management
- Principles of data governance
- Sector wise data governance scenario of selected sectors.

Lesson Outline

- Introduction
- Importance of Data Governance
- Data Governance Challenges
- Difference between Data Governance and Data Management
- Implementing an Effective Data Governance Framework
- Data Governance at government level- The Indian Scenario
- Sector-wise Data Governance Scenario
- Lesson Round-Up
- Glossary
- Test Yourself
- List of Further Readings
- Other References

INTRODUCTION

Data governance is everything one do to ensure data is secure, private, accurate, available, and usable. It includes the actions people must take, the processes they must follow, and the technology that supports them throughout the data life cycle.

Data governance (DG) is the process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage. Effective data governance ensures that data is consistent and trustworthy and doesn't get misused. It's increasingly critical as organizations face new data privacy regulations and rely more and more on data analytics to help optimize operations and drive business decision-making.

A well-designed data governance program typically includes a governance team, a steering committee that acts as the governing body, and a group of data stewards. They work together to create the standards and policies for governing data, as well as implementation and enforcement procedures that are primarily carried out by the data stewards. Ideally, executives and other representatives from an organization's business operations take part, in addition to the IT and data management teams. While data governance is a core component of an overall data management strategy, organizations need to focus on the expected business benefits of a governance program for it to be successful and independent.

Having a brief discussion on data governance, it is imperative to comprehend its genesis. Initially , data governance was strictly considered an IT function focused on data cataloging that was rarely, if ever, seen or used by the business. Next came digital transformation and the big data passion. Organizations realized they could extract value from all of the various data sets they were creating. This became Data Governance 2.0, which ushered in principles to propel modern, data-driven business. This approach embraces collaboration, dismantles organizational silos and spreads accountability across more roles.

In 2018, there was a "global reckoning on data governance," as the Wall Street Journal put it. Massive data breaches at organizations in numerous sectors resulted in serious reputational damage and declining market values for top brands such as Equifax, Facebook, Marriott and Yahoo.

Data governance can be divided into following:

i) Organizing: Identifying all the data sources and getting all the data in one place.

ii) Securing — Ensuring that the data is compliant with data privacy regulations and internal company policies.

iii) Managing and presenting data — After the requisite data is collated, its style of presentation to the team members of the organisation needs to be pondered.

iv) Using methods and technologies — Like modern data governance platforms.

DATA GOVERNANCE PRINCIPLES

The data governance principles espoused globally are as under:

1. **Maintaining the Integrity of the Data:** A principle of the highest importance is integrity. It depends on the entity using your data whether or not it is being used in the appropriate manner. Data integrity is maintained if their means and goals are ethical. In all decisions about the data, the participants must be honest and forthcoming. This can include decisions about actions, impacts, constraints, etc.
2. **Transparency:** In every case where data is used, proper transparency must always be maintained. To use data, as well as whose data is being used, all parties must understand how it is being utilized. Whenever there is a decision about usage or control, it must be communicated effectively to all parties involved. This will prevent any potential conflict in the future.

3. **Accountability and Ownership of the Data:** The ownership of the data must be defined. Appropriate procedures should be followed for defining access rights. Data governance applies to any data that is used across functions. As a result, data governance defines all decisions, processes, and controls related to data, i.e., its accountability.
4. **Data Audit:** Audits are permitted on every piece of data used. Any decision, control, and process about data that relates to data governance can be audited. Therefore, they must contain documentation proving compliance.
5. **Standardization of Data**: A company's data is used by many teams. In this case, the data in one format might not be compatible with another. It is imperative that specific guidelines and rules be defined in order to standardize data. In addition to these, there are rules for data definition, accessibility, security, and privacy.
6. **Change Management:** There may be some discrepancies in the data that require a change. As a result, there is always a risk of tainting the data. Therefore, data governance ensures proper change management activities, whether proactive or reactive. The data will include reference values, metadata, master data, and its use and structure.
7. **Stewardship:** The principle of stewardship should be adhered to. Accountability goes hand in hand with responsibility. It is essential to appoint a data steward in any organization. All rules and regulations must be followed by the data steward. This holds true for groups of stewards as well. It is their responsibility to ensure that the data is stored and used appropriately. It is their responsibility to always follow the best practices when managing data.

IMPORTANCE OF DATA GOVERNANCE

In the "information era" rapidly evolving technologies, and cutting-edge analytics and where massive amounts of data are used every day to drive critical business processes and decisions, data governance plays a vital role in every organization. From setting data management rules and regulations to measuring data quality and defining data interpretation processes, the benefits of following data management best practices are essential for forming proper business planning methods, and accounting approaches, and limiting operational risk. The importance of data governance may be covered under the following points:

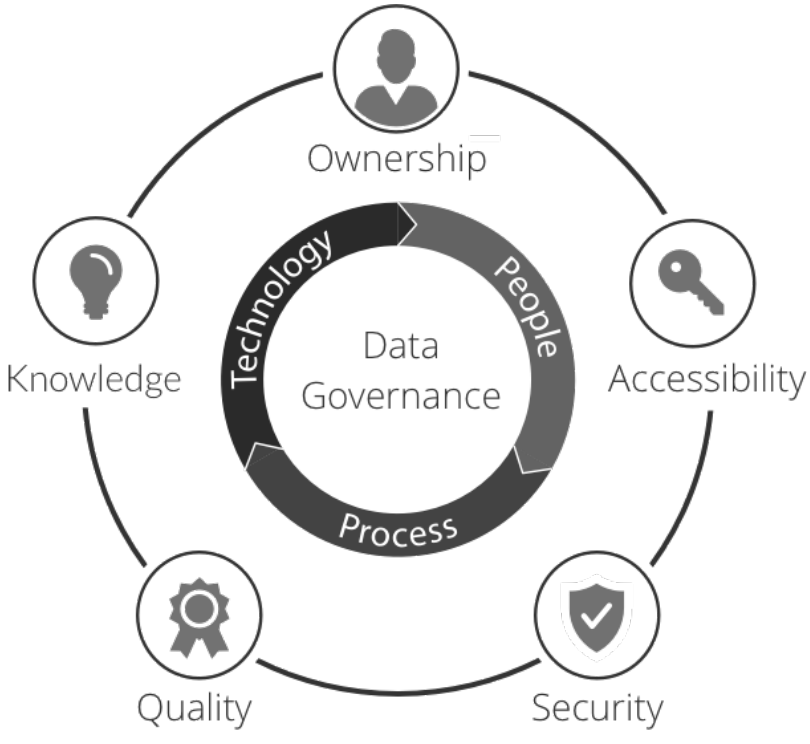
1. **Better Analysis** – Because data governance enhances data quality and find ability, analysts can find, understand, and analyze data more quickly and compliantly.
2. **Easily Defined Goals** – Data governance sets out a clear method to achieve business goals.
3. **Consistent Compliance** – Data governance helps data users stay compliant with regulations, reducing the risk of fees and reputational damage.
4. **Improved Data Management** – Governance minimizes duplicative efforts, which boosts operational efficiency.
5. **Standardized Systems and Data Policies** – Standardizing systems and policies across the organization instills users with ethics and awareness.
6. **Improved Data Quality** – As a business implements data governance, data quality will improve. This leads to more accurate business processes and higher customer trust.

Thus from the aforesaid point, it is evident data governance is indispensable for an organisation. Data governance is key to an organization because it steers a healthy culture with minimal chances of inaccurate, slow data. Processing data without a data governance framework makes it cumbersome for data teams to leverage data integrity and reliability.

Data governance is essential to a business because it fast-tracks an organization's trustworthiness and transparency while helping teams make responsive, data-driven decisions to benefit the company. Please refer the data governance framework provided in exhibit 1.

Exhibit 1

Data Governance Framework



Source: Imperva

DATA GOVERNANCE CHALLENGES

1. Understanding the business value of data governance

Organizations are constantly generating data. The sheer volume of data in modern organizations requires digital transformation initiatives to manipulate and serve up that data. Business managers reflexively see value in, say, getting data a half-day faster, but they don't reflexively see value in getting data that is well understood and properly controlled, with guardrails. One of the key data governance challenges lies in demonstrating to them that the extra half-day of speed doesn't ultimately buy them anything if they don't have data quality or an idea of where that data is being sourced.

Understanding the business value is especially challenging in older, established organizations that are adopting digital transformation more slowly. However, an advantage of implementing digital transformation is the ability to share relevant data with entities like employees, customers and vendors. This data democratization holds potential for moving the business forward, but it relies heavily on having the guardrails of data governance in place.

One solution has been the emergence of the chief data officer (CDO). This role started as a data-focused, right-hand person working for the chief information officer (CIO). With time, the CDO has risen to peer status with the CIO, and is tasked with making sure that the organization's thirst for data is quenched efficiently. That entails rolling out an architecture and a data platform that enforces data governance and keeps it top-of-mind.

The analytics group can also reinforce the business value of data governance. Often, this group is affected the most by low-quality, misunderstood, or misused data, as they are frequently charged with delivering insights, upon which crucial decisions are made.

2. Perception of IT owns the data

In the early days, many organizations relegated data governance to IT, thinking it was a matter of putting rules in place and restricting access to data. But data governance was rarely prioritized, and didn't survive IT budget cuts because it was not treated as having real business value.

One of the related data governance challenges is the widely held belief that IT owns the data and is consequently responsible for its governance. Much of data governance is an effort to clarify who really owns data. The idea that IT owns data — and, therefore, data governance — is an obstacle to digital transformation. In any decision about data, the first consideration is the business perspective, with IT enabling that decision afterwards. To better illustrate, consider the role of the database administrator (DBA). DBAs own the database, but they don't own the data itself. Their remit is limited to administering the vehicle that holds the data.

IT should not be making decisions about implementing solutions that affect data without talking to the business about the impact of those solutions. That would be going backwards, in an unsustainable process.

So, to achieve success, smart organizations have moved data governance back into the business, started showing stakeholders the risk of not having it, and started demonstrating value it can bring to an organization. They've also started defining data ownership, which sits with the managers in the business organization. A large utility in the European Union successfully made this shift, and realized savings of 30% on external data management costs, 50% reduction in data discovery time and 8 million Euros in business impact saved in first 18 months. It's not a matter of who owns Oracle data or who owns SQL Server data; it's a matter of sales owning sales data, marketing owning marketing data and so forth.

To overcome this data governance challenge, a best practice would be to appoint a data steward who works with a DBA to provide access to that data. That ensures that you're neither giving people random access to data nor fighting logjams with DBAs.

3. Limited or misallocated resources

Once the C-suite sees that data governance is important, the business is driving governance and IT is responding, the next data governance challenge becomes limited resources and misallocated resources. Whom will you get to take on the new roles of Data Owner and Data Steward?

Ideally, when data governance is up to full speed and full scale, those roles are filled by dedicated employees. Designating a data owner isn't as difficult because it means having a senior-level decision maker make a few more decisions. But finding a data steward is trickier.

The shortage of skills is a big obstacle to achieving ROI in data governance, so most organizations try to train somebody with tribal knowledge to be data steward. As most organizations start out, the role of data steward is a part-time job staffed by somebody on the business side who frequently works with data. On the IT side, it's usually a data architect or data analyst with additional, part-time responsibilities. Business analysts and business intelligence (BI) specialists are good candidates for data stewards because they're familiar with the data and with the technical folks on the team.

The company's road map and plan should ensure that, once it is demonstrating the ROI of data governance, it should start to dedicate people to these roles. Large organizations further down the road may have eight to 10 dedicated data stewards who cross the lines of business.

For organizations without the resources to have full time data stewards, one solution is to use consultants;

not IT consultants or outsourced developers, but data governance professionals from specialty firms. Some consulting firms are building practices around data governance experts who can fill those roles. At a small scale and for a limited period of time, it's a great way to achieve early success.

4. Siloed data

In many ways, siloed data happens because of different approaches to data operations and versions of technology. One of the things that most companies avoid whenever they modernize is figuring out how to migrate their systems holding legacy data. Those systems still do the job, and ripping and replacing them is prohibitive. Additionally, if the concerned officials don't know what that legacy data contains is, then it's less of an expense to try to solve the problem through downstream integration.

For example, traditional relational databases replace a lot of what was on the mainframe. They speed up transaction processing but require highly structured data. So, what needs to be done with all the new, unstructured data like video and social media? One can put it into NoSQL databases, but that creates a new silo, so people may turn instead to hybrid databases that offer the best of both worlds in one database.

Data silos are a natural phenomenon. Think of them from a business perspective. Often, the business itself is siloed, and the people focused on transactions don't communicate with those focused on strategy. Breaking out of those silos is key to further leveraging the data an organization produces.

One solution is to collect all of the information about the different types of data on different technologies for different uses by different parts of the business. With the right data governance tools, you can connect the metadata of different data types and see it in a uniform fashion. One can then map the architecture of the infrastructure and keep it updated more dynamically than if someone else tried to capture it in static documentation. One can replace the documentation by collecting it from the database instance itself, then bring it into an environment where the concerned person can analyze it and make it useful.

Metadata management is fundamental to successfully operationalizing data governance. Business users, of course, do not care whether the data resides on Oracle, SQL Server, Mongo DB or any other platform. They only want to understand how the data flows through their business and the point at which it comes to them.

5. Poor data quality and lack of trust in data

Data quality is a data governance challenge because poor quality leads to lack of trust, which leads to lack of use and lack of strategic results. If teams don't think they can trust the data they are using to make crucial decisions, it leads to data becoming more of an obstacle than an aid to strategic decision-making.

Even if one understand how poor the data is, he / she can still get good results by making adjustments to account for poor data quality. An important aspect of data governance is identifying where there is poor data quality and so that it may be looked into.

At the technical level, these are examples of poor data quality:

- For a particular field in the customer database, 75% of entries are blank.
- Over time, a vendor table accumulates thousands of duplicate rows.
- To test an application, someone generated hundreds of dummy records and forgot to delete them afterwards.

As part of data governance, one solution is to use data intelligence tools to establish metrics and thresholds for data quality and to monitor them. One can also put in place feedback loops so users can report poor-quality data to be cleaned for future use. When one make his / her efforts to provide transparency on available data the business users can decide whether they're comfortable enough with the level of quality to rely on the data.

6. Poor data context

Pay attention also to the data governance challenge on the other side of the coin: data context. Poor data quality leads to lack of trust, but lack of trust does not necessarily mean that you have poor data quality. It may mean that users aren't easily able to see what the data really means. Suppose a dataset is labeled "Final sales." Does that mean "Gross sales" or "Net sales?" Or a batch of records are presented as sales opportunities, but they are in fact actually marketing-qualified leads (MQL).

These examples are not a data problem; it's a perception problem. The solution is the same feedback loop described above, so that users can communicate questions and doubts about data and one can provide full context.

7. Lack of data control

To return to the relationship between IT and data, organizations that are good at IT but not at data governance generally have too much of the wrong kind of control. Fears about defending the attack surface or protecting the trade secrets leads to shutting out the people who can get the greatest value from the data.

One of the goals of data governance is to avoid turning away the people who will make the most of the organization's data. It's not an exercise in saying "no," but an effort to position oneself to say "yes" with confidence. If one have true control over the data, with the context and understanding that comes from data governance, then one can achieve the right level of control.

Reflexively, companies turn to measures like securing the desktop, the databases and user identities. While those are steps on the right path, one can't apply them effectively across data sources if there is lack of understanding on the data landscape. Data governance tools help a person to map out what each data asset is for, and what they should and shouldn't be for.

DIFFERENCE BETWEEN DATA GOVERNANCE AND DATA MANAGEMENT

S. No.	*Basis of Difference*	*Data Governance*	*Data Management*
1	Definition	Data governance is the process of ensuring that the business rules for data are established and followed.	Data management is the process of making sure that all data is captured, managed, used, and disposed of properly.
2	Scope	Data governance is a process that ensures the quality of data and the data management process.	The scope of data management is to ensure that all data have been collected, stored, and managed in such a way that it can be accessed by authorized users.
3	Benefits	Data governance helps organizations to be more efficient in their operations by providing a framework for managing data. The benefits of data governance are that it ensures consistency in project deliverables and reduces errors related to using out-of-date or incorrect information.	Data management provides many benefits to organizations. It improves efficiency, accuracy, and security of data. It also minimizes the risk of human error when handling large amounts of information.

S. No.	Basis of Difference	Data Governance	Data Management
4	Challenges	There are some challenges when it comes to data governance. One challenge is that it can be hard to know where you should start your data governance process. Another challenge is that there can be some resistance from employees who don't want their work interrupted by new rules and regulations or who are not familiar with how the new system works.	There are many challenges that come with data management. One of the most common problems is that there is no one set standard for how data should be organized and stored. Another issue is that it can be difficult to find information when it needs to be retrieved from various sources and locations.
5	Best practices	There are many best practices of data governance. The first one is the proper classification of data, which means using metadata to identify what kind of data it is, who owns it and what its purpose will be. Another best practice is the use of access control lists to grant or deny access to certain people or groups.	Data management best practices include good file naming conventions, ensuring high quality of data, proper labelling, secure data storage, defining backup plans, and improving data security.
6	Technology	Technology can be used to help with data governance in a number of ways. For example, it can be used to automate processes like records management or to monitor user activity across systems. It can also be used to improve the effectiveness of current processes by automating manual tasks or providing insights into how the business operates.	Data can be managed in a more organized way with the use of various tools and techniques. Some of these tools are- Database Management Systems, Data Mining Tools, Data Visualization Tools and Big Data Technologies.

IMPLEMENTING AN EFFECTIVE DATA GOVERNANCE FRAMEWORK

The most common objective of data governance is the standardization of data definitions across an enterprise or organization. Other goals and objectives depend on the focus of a particular data governance program. Within the commonly accepted data governance framework, one should determine principles that make sense for the environment.

Every organization is guided by certain business drivers — key factors or processes that are critical to the continued success of the business. Your organization's unique business drivers dictate what data needs to be carefully controlled, and to what extent, in your data governance strategy. For example, one of a healthcare organization's business drivers may be to ensure the privacy of patient-related data assets, requiring that sensitive data be securely managed as it flows through the business to ensure compliance with relevant

government and industry regulations. At the same time, patient data must be readily accessible to a patient's healthcare providers. These requirements inform the provider's data governance strategy, becoming the basis of its data governance framework.

A well-planned data governance framework covers strategic, tactical, and operational roles and responsibilities. It ensures data is trusted, well-documented, and easy to find within your organization, and that it's also kept secure, compliant, and confidential.

Some of the most important advantages the framework provides include:

- A consistent view of — and business glossary for — data, while allowing appropriate flexibility for the needs of individual business units
- A plan that ensures data quality, accuracy, completeness, and consistency
- An advanced ability to understand the location of all data related to critical entities, making data assets discoverable, usable, and easier to connect with business outcomes — in other words, ensuring
- A "single version of the truth" that keeps critical business entities aligned across the enterprise
- Well-defined methodologies and best practices for data assets and data management that can be applied across the organization
- Easily accessible data that's kept secure, compliant, and confidential according to the demands of legal or regulatory requirements

The three pillars of data governance framework are as under:

i) **Governance including all data assets**: Everything from dashboards and code to data science models is a data asset. The data governance framework should take into account all data assets, i.e., data and analytics governance.

ii) **A practitioner-led, bottom-up approach:** As the number of data users and consumers keeps rising, making a few people (data stewards or engineers) accountable for data governance isn't a sustainable approach. A decentralized, bottom-up data governance framework that makes every data creator responsible for data governance is the way forward.

An example of a decentralized, community-led approach is the data mesh. The data mesh design proposes a federated computational governance model, where every organization is a federation of business domains. Domain owners fully manage the data they create. However, each domain still follows a set of global (or federal) rules on data definitions, standards, processes, and discovery.

At this juncture, it will be of substantial academic interest to comprehend the traditional approaches of data governance framework at length.

i) The top-down method: Focus on data control

This is the centralized approach to data governance. It relies on a small team of data professionals who employ well-defined methodologies and well-known best practices. This means data modelling and governance are prioritized. Only later is the data made more broadly available to the rest of the organization for analytics.

However, this approach creates a massive scalability issue. In this framework, there's a clear distinction between data providers (typically IT) and data consumers (typically business units). Only data providers are empowered to have any sort of control over the data.

In the past, this was less of an issue because there was a smaller amount of data to be governed, and fewer teams that needed access to it. But today, these small teams of data producers can't cope

with the demand from data consumers. It's now a business necessity to have clean, complete, and uncompromised data available to everyone who needs it, whenever they need it. There are simply too many business users making too many requests for these teams to keep serving as gatekeepers.

ii) The bottom-up method: Focus on data access

Conversely, the bottom-up method allows for much more agility when managing data. While the top-down method starts with data modelling and governance, the bottom-up approach starts with raw data. After the raw data is ingested, structures on top of the data can be created (referred to as "schema on read"), and data quality controls, security rules, and policies can be implemented.

This framework, popularized with the advent of big data, is more scalable than the centralized approach. That said, it creates a new set of data issues. Because data governance isn't implemented until later in the process, and because anyone can enter data, it's harder to establish control. And as we already discussed, lack of data governance can lead to increased regulatory risk, a loss of stakeholder trust in the organization's data, and a higher cost of data management for a sprawling mess of data assets.

What we need is a modern approach to a data governance framework — one that balances access and control. We need to establish control early on in the process without sacrificing the ability for users and subject matter experts to become data owners and curators.

iii) Governance practices embedded within daily workflows: Data governance has always been associated with compliance, control, and risk mitigation. However, it is a business function that can support strategic decision-making by ensuring that everyone has access to accurate, relevant, high-quality, and trustworthy data. That's why it cannot be an afterthought. Instead, it should be embedded within the daily workflows of data practitioners.

Steps Involved in creating Data Governance Framework

i) Revisiting the definition of data governance: Data governance is an ever-evolving project, which is why it is required to revisit and question the idea of data governance before getting to formulating a framework.

While revisiting the definition of data governance, the following questions needs to be focused on-

a) What is the purpose of data governance?

b) Does it cover all data assets across the organization?

c) Does governance also foster organization-wide data sharing and collaboration?

ii) Identification and defining the data domains: Since the data governance framework should cover all data assets, the next step is to identify and standardize data domains across the organization. One can have domains like finance, marketing, sales, etc. corresponding to each function generating data. In this regard, the following questions to be looked into-

a) Which are the important data domains in an organisation?

b) What data is generated?

c) Where is that data currently?

d) Who consumes / uses the data?

iii) Identifying domain data owners and consumers: A key tenet of modern data governance is shared responsibility for data. So, each domain creating data is responsible for managing it and ensuring its security, integrity, and privacy. That's why the next step is to assign data owners to each domain and

understand its data consumption pattern to ensure that the right people have access to the data they need.

In this section, the following questions are relevant-

a) Who is creating data within each domain?

b) Who is consuming that data and how? What do their daily workflows look like?

c) What are the current dependencies to get access to domain data?

iv) Validating and documenting of everything pertaining to the data: By this stage, one must have a clear idea of data flow within the organization. The next step is to standardize data domain definitions, data flow rules and workflows, access policies, and more by documenting everything.

The documentation should address the following:

a) Where does data originate from?

b) What does it mean?

c) How does it flow through your organization?

d) Does it help domains meet their goals?

e) Does it support your organization's business outcomes?

v) Conducting data security and risk assessments for each domain: The last step is to set up processes to conduct frequent data security and risk assessments for each domain. That's because enabling data governance is a journey, rather than a one-time project implementation. This step involves the following questions-

a) What are the existing data access policies and security checks for data from each domain?

b) Who is allowed to access what data and why?

c) Do these policies mitigate risks without creating data discovery, access, and collaboration bottlenecks?

DATA GOVERNANCE AT GOVERNMENT LEVEL- THE INDIAN SCENARIO

Administrative data forms the backbone of decentralized evidence-based decision making in the Government of India. With emerging international evidence of the vital role played by data as an enabler in driving public policy across its lifecycle, the Central and State Governments have paid significant attention to their data systems over the past two decades. Management Information Systems (MIS) and dashboards have been developed for most government schemes and programs. To disseminate this information more widely, Open Data initiatives have also been undertaken. Recently, attempts have also been made to foster data exchange across Ministries/Departments via the Prayas Dashboard at Prime Minister's Office and the Output-Outcome Monitoring Dashboard at Development Monitoring & Evaluation Office (DMEO), NITI Aayog.

In this context, a comprehensive review of present data preparedness levels of all Ministries/Departments was required to chart way forward and suggest measures for improvement. Against this background, the Data Governance Quality Index (DGQI) exercise was initiated with the objective of assessing data preparedness of M/Ds on a standardized framework to drive healthy competition among them and promote cooperative peer learning from best practices.

Several existing data maturity models were studied to develop DGQI's methodology. Three key steps of data preparedness were identified: (a)Data Strategy to lay down systemic guidelines, (b) Data Systems to ensure

smooth processes of data generation, management and its use and (c) Data driven Outcomes where data is utilized and widely shared by institutions to drive decision making.

While DGQI 1.0 focused on only data systems pillar, DGQI 2.0 aims to assess data preparedness levels of Ministries/Departments across the three pillars.

Genesis of Data Governance

Data collection and warehousing started as early as 1881 when the first Census was conducted in India. After Independence, National Sample Survey Organization was established in 1950 and Central Statistical Organization in 1951. Data collected through large scale surveys by these organisations, and the administrative data collected by Ministries and the state Governments led to data-driven decision-making in the Central and the State Governments. Scheme-level information generated and collated at various levels i.e., village, block, district and state levels, assisted programme implementation. However, the whole exercise was done manually on formats individually developed under each scheme and overall scheme progress was mostly tracked inputs (fund releases and budget utilization).

MIS systems and digital data storage facilities became all pervasive in the last two decades. Gradually, activities and outputs started to get monitored. With digitization of data, advent of new techniques and ever-increasing importance of data in public policy, the need for even better management of data was recognized. In order to further India's vision towards Open Government and Open Data initiative, National Data Sharing & Accessibility Policy was adopted and data.gov.in was launched to provide all relevant data from Government at single place for wider public use. Many schemes also migrated to dashboard based and basic analytics-driven systems which make complex information available to decision makers in simple charts and figures. Intra-government exchange and integration of data is now being facilitated using ICT platforms such as DISHA, Prayas and Output- Outcome Monitoring Framework (OOMF).

Current Scenario of Data Governance

As of now, an internal Management Information Systems (MIS) is developed for most government programmes, which provides required information regarding coverage and outputs of the programme, e.g., HMIS for National Health Mission which tracks information uploaded by the States/UTs which enables planning, management, and decision-making based on grading of facilities and various health indicators at block, district, state as well as national level. Such programme MIS typically have capabilities to generate standardized analytical reports on the basis of data collected. Further, Ministry of Statistics and Programme Implementation (MoSPI), through Twenty Point Programme (TPP-2006) and Infrastructure and Project Monitoring Division (IPMD) monitors key infrastructure projects within the Government.

The Government also launched Digital India programme in 2015 to ensure digital availability of government services to citizens. This Programme is being managed by National e-Governance Division (NeGD). NeGD provides project development and programme management support to e-governance related measures taken by Ministries. Some of the State Governments also present the work done by their various departments through dashboard based analytical systems (e.g. Pratibimba by Govt. of Karnataka). These measures have ushered in a new era of accountability.

Overall, it is clear from the background above that governments in India have been quite proactive in ensuring adoption of newer technologies in data management and thereby improving programme outputs and outcomes. However, there still remains lot more to be done with reference to data governance, especially with respect to programme monitoring and management.

Given the above, it is imperative that a comprehensive review of data preparedness is conducted for government data systems for scheme management and decision support information systems. Development Monitoring

and Evaluation Office (DMEO), an attached office of NITI Aayog, has developed DGQI toolkit to enable a comprehensive self-assessment of data preparedness levels to come up with Data Governance Quality Index (DGQI) for the government agencies at the central and state level.

Objectives of Data Governance Quality Index (DGQI)

The intent of the DGQI is to enable Ministries/ Departments and state departments to assess themselves at various levels of data maturity on the basis of a standardized framework, which in turn would facilitate deepening of digitization in the Government of India.

It is hoped that in the long run, DGQI will help in laying the foundation of more integrated monitoring systems, for e.g., a single, online, API-integrable 'Overarching Dashboard' kind of monitoring system of all the CS/ CSS schemes of all M/Ds, ultimately leading to a state-of-the-art data-driven decision making.

The objectives are as follows:

- To enable review and assessment of data preparedness of the data/ MIS systems of the Ministries/ Departments on objective parameters of a standardized framework.
- To prepare a self-assessment diagnostic tool that will enable the M/Ds to internally contemplate the need for improving data systems.
- To enable the commissioning agencies to conduct a comparative assessment of data preparedness and source best practices in IT systems which can enable improved cross-learning between the participating agencies.

Methodology of DGQI

Under the realm of the overall approach, six key themes have been identified under data systems pillar covered by the Data Governance Quality Index:

1. **Data Generation**: Data generation measures the ability of the respective ministries/departments to efficiently generate useful data in the course of their programme implementation. It covers areas related to the level of digitization, frequency and granularity of data generation. It also assesses if mobile phones, location tracking and GIS mapping is used to authenticate the generated data.
2. **Data Quality**: Data Quality covers processes of scientifically and statistically evaluating data in order to determine whether they meet quality benchmarks. The key areas covered under this theme relate to profiling of data, data quality assessment processes (for e.g. data pipeline design, well defined data schema etc.), data cleaning, use of latest technologies and mobile phones in the process.
3. **Use of Technology**: This theme assesses if emerging technologies are being utilized to improve data robustness. It assesses if MIS of ministries/departments have linkages with PFMS for ensuring transparency and Jan-Dhan, Aadhar and Mobile [JAM-trinity (if applicable)] for delivering last mile services. It also explored if other data sources such as remote sensing or social media data is utilized in addition to data collected by ministries/departments to get a nuanced understanding. Finally, it also measures if emerging technologies like block chain, big data analytics, machine learning, artificial intelligence, IoT are being used to collect data or to draw analytical insights from it.
4. **Data Analysis, Use and Dissemination**: One of the core themes, it covers if the collected data is being analyzed and used for evidence creation and decision making. Given the present context, it gauges whether ministries/departments are undertaking basic cross-sectional analyses only or regression and predictive analysis as well. The use of dashboards for visualization of data is also checked to ensure that information is disseminated in a user-friendly manner. It also assesses if other social media

platforms are also being increasingly used for information dissemination and whether websites have features to support multi-lingual interfaces and are GIGW compliant.

5. **Data Security and HR Capacity**: While data security requires an in-depth analysis in itself, the same is briefly captured in the index also to reflect its importance. It assesses if anti-virus updates and internal audit systems are in place to ensure data is not corrupted or prone to threats. These were identified to be the minimum requirements expected to be met and are not meant to be exhaustive in nature. To look at HR capacity, the existence of dedicated data quality teams has been considered. Again, this is by no means an exhaustive measure of capacity development but was adopted as the starting point.
6. **Case Studies**: The present questionnaire for this theme focuses on scheme-level MIS. Any intervention done at the Ministry/ Department level or any innovative approach that may not be captured in the structured questions of the tool can be highlighted through best practices. These best practices can be provided as case studies. This theme is expected to help unlock the hidden potential not only in terms of enhanced decision making through inter-ministerial collaboration but also by opening doors for learning from challenges faced and the solutions devised by peer ministries.

National Data Governance Policy

To unleash innovation and research by start-ups and academia, the Union Budget 2023-24 mentioned about National Data Governance Policy to enable access to "anonymised data" to unfetter innovation by startups and academia. The current draft Data Protection Bill 2022 published by MeitY (Ministry of Electronics and Information Technology) does not cover non personal data. The mentioned document sought to set up an India Data Management Office (IDMO) under the Digital India Corporation and frame certain guidelines for sharing of non-personal data by private entities. The concept of sharing non-personal data is different from Personal Data Protection Bill, 2022 which was designed with online personal privacy in mind. A suitable framework will allow non-personal data to be equitably utilized by start-ups and academic institutions in the country to bolster research and innovation.

As part of the policy, the Indian government will also build the India Datasets program, which will consist of non-personal and anonymised datasets from Government entities that have collected data from Indian citizens or those in India. Private entities will be encouraged to share such data, according to the policy.

SECTOR-WISE DATA GOVERNANCE SCENARIO

In this section, data governance of selected sectors has been studied to comprehend various dimensions of data governance prevailing in those sectors. The sectors considered for the study are as under:

1. Banking
2. Automobile
3. Energy
4. Hospitality
5. Information Technology (IT).
6. Telecom
7. E-Commerce

1. Banking Sector

In banking, data governance is about meeting both regulatory and internal requirements. Banks in particular need to derive value from data for both the innovation and modernization of their operations as well as for continued compliance and ethical management of the data they work with.

Today's economic landscape requires most if not all industries to enhance their data-driven capabilities in the market to maintain a competitive edge. The banking sector is no exception. The introduction of data governance models in banking gives banks the resources they need to upgrade their current procedures and policies to improve their data protection mechanisms. Through following ways data governance provides value in the banking sector.

i) **Regulatory compliance:** It is a requirement for banks to keep all the data they have secure, based on a variety of federal and state compliance regulations. Regulatory requirements continue to pressure the banking industry to get data governance under control as the consequences of data violations become more costly. With the right data governance plan in place, banks always know exactly what data they have access to. They also always know where data is located, which ensures they can enforce the right controls — even during complicated projects like cloud migrations.

   Awareness of the location of data, the regulations it is liable to and the correct approach to protection is key for successful cloud migration and other digital transformation projects. Data governance provides pertinent tagging to ensure banks satisfy regulatory requirements with the correct access and security controls.

ii) **Cost cutting:** Manual data management is tedious, inefficient and expensive. The responsibility of manual data management is often placed at the doorstep of IT teams, which means financial institutions frequently foot the costs of maintaining active IT teams.

   Data governance relieves the manual burdens of discovering, granting access to and implementing security to data through centralizing technologies, effectively ending the need for multiple costly third-party systems and sprawling IT teams. The self-service capabilities of many data governance tools ensure that organizations maintain secure data access without incurring unnecessary costs.

iii) **Market insights:** The financial sector is now characterized by relentless competition between institutions and saturation for newcomers. As a result, market insights have become a necessity for a competitive advantage. Through data analysis initiatives, banks can confidently approach their data and derive actionable insights.

   Data governance supports company-wide analysis and its processes, ensuring that there is easy access to data and that it is well organized. This makes it easier to innovate and use data across the organization as opposed to leaving the responsibility solely to leadership teams.

iv) **Data-driven culture:** Data-driven models are increasingly transforming how organizations handle business goals and objectives. A data-driven culture is proving to be of great benefit to organizations, as it intuitively improves approaches to cost-cutting, innovation and customer insights. Data governance supports and encourages a data-driven culture so banks can more effectively run their operations and make customer-experience-focused decisions.

v) **Collaboration and risk management:** Banking institutions work with hundreds of data sources and require a way to log the data they have. They also need to utilize data for managing and acquiring new customers, discerning fraud and reducing risk. With support from data governance processes and procedures, banks create data catalogs to ease both data discovery and quality assessment. The result is better collaboration and decision-making and improved productivity.

vi) **Improved compliance and customer service:** As the customer experience and secure data controls become more important to consumers, institutions such as Fifth Third Bank are evolving their data governance approaches to become more effective and less invasive, enhancing both compliance and the customer experience.

Financial entities are also looking to deliver more personalized customer experiences to their customers but struggle as they encounter patchy and segmented data. NCBA[1] is a financial services institution that handles this challenge by adopting various customer experience platforms to enable the organization to follow the customer journey from beginning to end. Through this approach, they are able to derive insights into customer patterns and improve experiences for their clients.

At this juncture, it would be of paramount academic significance to explore the necessity of ESG data governance in banks. As it is a well-known fact that banking industry is facing mounting pressure to meet fast-changing demands in environmental, social, and governance (ESG) issues. New and evolving regulations call for greater transparency and disclosure of ESG-related data (see sidebar, "ESG regulatory and disclosure requirements"). Stakeholders and investors are increasing their scrutiny of the effects investment decisions have on the climate and society.

Now a day, banking preferences of consumers are massively influenced by concern about purpose and sustainability. To meet these expectations it is imperative that banks should integrate their IT systems to systematically collect, aggregate, and report on a broad range of ESG data. However, many financial institutions still do not have a comprehensive approach to integrating ESG data into their existing risk reporting.

Moving toward this goal will require significant changes to the IT infrastructure, from applications to data integration, architecture, and governance. New applications include not only the management and capture of ESG data but also financed emissions models, climate risk models, ESG scorecards, climate stress tests, and climate-adjusted ratings. ESG data must be woven into existing processes, such as credit approvals and decision making. And banks will need to adjust their data architecture, define a data collection strategy, and reorganize their data governance model to successfully manage and report ESG data.

Embedding of ESG requirements into core banking processes

- Integrating new workflows into existing processes, such as using artificial intelligence to incorporate ESG data into decision-making processes (for example, credit decisions).
- Communication of ESG requirements across the organization and bring all employees on board with an intentional change management approach.
- Review and revise existing data processes to comply with changing ESG requirements (for example, increasing the frequency of data updates).
- Developing a clear plan to support the integration of new ESG policies (such as how to add new certificates to investments).

2. Automobile Sector

The global big data market in the automotive industry size was valued at USD 4,500 million in 2021. It is expected to reach USD 15,800 million by 2030, growing at a CARG of 17% during the forecast period (2022-2030). Embracing technology, applications, and services ranging from sensors to artificial intelligence to big data analysis is transforming the auto industry. Consequently, the ecosystem is witnessing a steady influx of new participants, leading to the continuous evolution of automobiles.

Big data analytics facilitates the automobile manufacturing sector to collate data from ERP systems and combine data from different corporate functional units and supply chain participants. It is heartening to note that automobile industry is getting ready for industry 4.0 with the emergence of IoT, a networked system, and M2M communication.

1. East African financial services organization NCBA has embraced digital technology to serve its 32 million global customers. However, its sales process was comparatively antiquated. Data was segmented and patchy, making it difficult to deliver a first-class customer service.

With reference to data governance in automobile industry, it is worth mentioning the four data usage categories provided by German Association of Automotive Industry-

a) Traffic data: Traffic related data can be released anonymously to public services, such as police and fire brigades, supporting them in keeping the public space safe. Early alert systems are able to operate in real time.

b) Usage data: used for different services and business models, the required data is released anonymously to third parties in order to develop new products. Telematic data can be included, as well as traffic-related data. This category will drive future innovations.

c) IP data: this category contains data sets relevant to intellectual property. This data is available solely to the original equipment manufacturer (OEM) and its contractual partners; its use is in order to further improve the OEM's vehicles and to gain valuable insights on the lifecycles of its products under real conditions. The new dimensions of evaluable data are likely to optimize certain mobility concepts to a new level, also enabling the creation of more brand specific services.

d) Personal data: in order to offer a service tailored to an individual, it is necessary to gain information on that driver's behaviour, only possible if the individual gives their consent. With privacy more valued than ever, personal data requires special treatment; since the introduction of the General Data Protection Regulation (GDPR), severe penalties for data breaches can be imposed on companies.

These four categories consist mostly of technical data gathered through the telematics system of the connected vehicle and from roadside infrastructure. Communication between connected vehicles and infrastructure leverages the positive effect of data usage on traffic safety; in addition, fuel consumption is expected to decrease through intelligent analysis of operating states and following recommendations for travel speed and in advance phased traffic lights.

OECD: Connected and Automated Vehicles – A Case Study of Data Governance Issues

Connected vehicles can connect to networks, like the Internet, to send and receive data with other networked devices, both inside and outside the vehicle. They communicate with other systems, including other vehicles, roadside infrastructure and third-party service providers.

Automated vehicles are vehicles that operate on a spectrum of declining input from the driver. Many aspects of vehicles manufactured today already involve some automation, including adaptive cruise control and active lane-keeping assistance. However, a fully automated, or autonomous or self-driving, vehicle would include the "full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver".

Automated vehicles periodically connect to the Internet, e.g. to receive software updates. In some models of automated vehicle deployment, achieving increasing levels of automation is expected to require increasing levels of connectivity, as vehicles use short-range communication technologies to communicate with road-side infrastructure and other vehicles.

Connected and automated vehicles highlight how different digital technologies increasingly converge in their application. Connected vehicles commonly use both cellular and non-cellular communication technologies to connect and share data with other vehicles, infrastructures, road services, satellites and other third parties. Given they are intelligent devices using sensors to collect data, connected and automated vehicles are part of the "Internet of Things". Connected vehicles on the market already display some forms of automation enabled by artificial intelligence (AI), including driver assistance systems. As technologies continue to develop and vehicles become more automated, the use of data processing technologies like big data analytics and AI systems within vehicles are also expected to increase.

Connected vehicles can collect and transmit data differently than previous generations of vehicles. As vehicles become increasingly automated, data generation, collection, storage, transfer and use are expected to become increasingly essential to their function. The types of data collected or generated by such vehicles will include the following:

i) Locational data – data about the precise geographic location of a vehicle, including direction and speed. Such data will interact with both static data (such as high-definition 3D maps) and semistatic data (on temporary events like weather, accidents and traffic jams).

ii) Sensor data – data about how the car perceives the external environment, including infrastructure, traffic signals and other road users. Such data are typically derived from radar or light detection and ranging (LIDAR) sensors or cameras.

iii) Diagnostic data – data on how the vehicle is performing with respect to fuel consumption, energy emissions, engine operation, battery status and performance, among other indicators.

iv) Driving behaviour data – data on driver behaviour, such as seatbelt use, speeding and frequent stopping.

v) Identity and biometric data – identity data, like names and other identifying information. This may also include biometric information, like fingerprints.

It is to be noted that vehicles are increasingly gaining the ability to collect a greater variety and volume of data. In most jurisdictions, these data are primarily controlled by the original equipment manufacturer following consent from users as part of the terms of service. Some authors describe original equipment manufacturers as in a "gatekeeper position" with control over data generated by connected and automated vehicles.

Data collected by connected and automated vehicles can be valuable to other vehicles and infrastructures to support their basic function. Some models of automated driving rely on real-time flows of data to achieve real-time two-way communication. This is true both among connected vehicles and between them and connected infrastructures, like parking spaces and motorways.

In these models, effective automated driving relies both on short-range communications technologies in cars, as well as mechanisms to enable data access and sharing between a diverse set of stakeholders. These could include vehicle manufacturers, regulators, communication service providers and third-party service providers.

Similarly, data generated by vehicles could improve public policy development and delivery. Potential applications for connected car data include improved traffic management and safety, more dynamic management of traffic control and public spaces, and better and faster responses to incidents in real time. Data from connected and automated vehicles could also help authorities plan the deployment and improvement of new infrastructures and public services, like roads or bike paths.

Data generated by connected vehicles may also emerge as an input into production for firms. Automotive ecosystems are already characterised by a wider variety of actors. These include suppliers and downstream sectors with strong linkages, like the wholesale and retail trade and repair of motor vehicles sector.

Data from connected vehicles could have value for vehicle manufacturers, enabling new business models and revenue streams associated with improving customer experience and operational efficiency. Similarly, data from connected vehicles could provide the basis for the provision of a range of other data-driven business models to car users. These might include predictive repair and maintenance services or dynamic adjustment of insurance services based on driving behaviour

Emerging approaches to data governance for connected and automated vehicles

Data from connected and automated vehicles are the subject of policy debate across the OECD. Discussions on how the data should be collected, processed, shared and reported are ongoing across OECD jurisdictions. In this regard, the following points are discussed in brief-

a) Approaches to data collection and management at the vehicle level: Data collected by connected and

automated vehicles are likely to be personal in nature. Much technical data like vehicle movement and condition are likely to concern the driver or the passengers. In view of the privacy risks associated with the data generated and collected by vehicles, both the public and private sectors have developed principles for data collection and management at the vehicle level.

These complement data protection regulations that usually do not directly address automotive data. Among original equipment manufacturers, alliances of industry actors have promulgated codes of conduct and privacy principles. The European Automobile Manufacturers Association has developed the Principles of Data Protection.

In the United States, the Alliance for Automotive Innovation has developed Consumer Privacy Protection Principles. The Alliance represents the manufacturers of 98% of personal vehicles sold in the United States. Other players in the automotive ecosystem have also developed privacy principles. Automotive data aggregator Otonomo, for example, released a "Privacy Playbook" in 2019 (Otonomo, 2019[56]). Some data protection and privacy enforcement authorities have also released principles for data management and collection at the vehicle level for connected and automated vehicles.

b) Data processing near or within vehicles: Questions of where and how to process data from connected and automated vehicles are informed by both technical requirements, including the need to ensure vehicle operation, as well as regulatory and other requirements and obligations of the data controller, including with respect to the management of personal data. In some models of connected and automated driving, data are expected to be shared and processed with a widening ecosystem of actors.

These range from equipment manufacturers, regulatory authorities and cloud service providers to network operators and third parties providing in-vehicle services. Noting this expansion, the European Data Protection Board has encouraged local processing of data, notably within the vehicle. It also recommends avoiding external cloud computing of data generated in connected vehicles . The Board notes that such an approach can increase user control over personal data, as well as "mitigate the potential risks of cloud processing".

c) Access to in-vehicle data from connected and automated vehicles: Data from connected and automated vehicles are of value to a wide range of stakeholders. To date, such data have been under the effective control of the original equipment manufacturers after consent from the vehicle owner and/or driver. This consent is typically required to use the vehicle and related services. Manufacturers often argue that tight controls on this data are necessary for privacy and security. They contend such data are often personally identifiable and subject to appropriate laws.

Other actors, including those in the motor vehicle aftermarkets and services areas, argue that access to such data could enable a more competitive ecosystem of services. In addition, the owner and purchaser of a vehicle may not be its sole user. This can further complicate questions of how collected data should be appropriately managed and shared.

Efforts to enhance access to in-vehicle data in a mechanism compliant with the European General Data Protection Regulation are ongoing at the European level. Meanwhile, some jurisdictions like Austria, France and Germany have already set rules on access to vehicle data by third parties, including regulatory authorities and insurance companies.

d) Models of data sharing and access for connected and automated vehicles: Data from connected and automated vehicles have value for a wide variety of actors. These range from players directly involved in the automotive industry ecosystem to third-party players like entertainment service providers and aftermarket service providers of repair and maintenance.

In the context of automated vehicle trials, data sharing among innovators could help speed the rate of experimentation. However, unlike other valuable inputs into production, several economic characteristics, including non-rivalry, economies of scale and information asymmetries, preclude the emergence of markets for data.

In the absence of widespread data trading, and since data might be used repeatedly without depletion, new frameworks are required that enable greater data sharing among actors. This is particularly relevant for the automotive industry, which features a variety of upstream and downstream industrial activity and where incentives may not exist for data sharing.

e) Data reporting and mandatory data collection: Data from connected and automated vehicles hold great potential for regulators and public authorities to improve development of public policies – from transportation to planning. The use of such data can create public value by enhancing network operations, investment, maintenance, planning and road safety in jurisdictions across the OECD.

Initiatives across the OECD encourage or oblige the reporting of some kinds of data generated or collected by vehicles to relevant public authorities. For example, in Europe, the Data for Road Safety initiative enables data sharing between vehicle manufacturers and national road authorities and service providers. This initiative recognises that vehicles are already equipped with technologies that can detect dangerous road conditions and warn drivers.

It also points out that greater sharing of these data could benefit road operators and the wider public. The International Transport Forum recently released good governance principles for frameworks for reporting mobility data to public authorities. In some jurisdictions, authorities mandate the collection of some kinds of data.

For example, event data recorders are compulsory in vehicles in some jurisdictions. These devices record technical vehicle and occupant information for a brief period before, during and after a crash. Developing similar data storage systems for automated vehicles was the subject of international negotiation in 2020.

3. Energy Sector

From smart meters to equipment sensors, energy companies are gathering data from connected devices at an unprecedented rate. But as data grows—and BI, predictive analytics and artificial intelligence practices evolve—energy companies face a challenge: Innovation intended to improve the customer experience or operational efficiency can easily conflict with privacy laws, corporate data guidelines and regulatory mandates. So the million dollar question is how can the energy industry balance this data deluge with regulatory compliance and data protection laws?

Given the quantum of data being collected daily, there is a growing likelihood that organizations will “cross the line,” either intentionally or, as in most cases, unintentionally—especially as they begin to leverage new technology like big data analytics, artificial intelligence (AI), machine learning (ML), Internet of Things (IoT), and sensor-driven devices. So, as these technologies and use cases evolve, regulatory requirements must also shift in response.

Regulatory agencies are tasked with protecting consumer privacy, enforcing existing regulations and constantly updating codes to adapt alongside changing technology, but they still require a healthy balance to maintain innovation: If agencies overregulate, they may unnecessarily hamper growth; too little, however, and they risk failing the public or corporate trust.

Limitations of Data Governance in Energy Sector

It is been observed that most of the power and utilities companies are not utilising data to its full potential nor meeting the soaring needs for data sharing. Across the energy sector, organizations are discovering that their data strategy will not enable their future. In fact, the data capabilities at many power and utilities companies do not even meet today's needs for reliability, efficient costs and employee productivity, let alone the evolving data-sharing needs related to emissions reporting and sustainability efforts.

Quality data is needed for all day-to-day operations for power and utilities companies, for compliance, grid reliability, financial reporting and to improve customer experience. Companies depend on data to keep lights on and homes warm during a storm, plan for peak energy demands despite volatile weather, get ahead of maintenance and mitigate damage by more quickly knowing the impact of natural disasters.

More than just a tool to avoid trouble, data can be an asset that strengthens the future of the organization. Business users need the latest data to innovate and build an agile data marketplace. With access to accurate data systems, users can develop self-service analytics capabilities to explore new ways to address operational and energy supply challenges.

In a survey conducted by EY across a segment of the electricity industry revealed the following significant points:

i) Power and utility executives see the value in data as an asset, but many companies are at the nascent stages of establishing a modern data plan. None of the respondents has a chief data officer, and less than one-third have a formal data strategy that serves both IT and business needs.

ii) A direct relationship between senior leaders' commitment to improving data governance and the organization's success rate with digital maturity and data accuracy, reporting and strategy was observed. The EY report explored the potential of the chief data officer in power and renewable companies in aligning data strategy to business strategy.

iii) While a CDO, specifically, isn't mandatory, but it is important to have a formal oversight, process or centre of excellence that works collaboratively to manage how data is collected, protected, shared and controlled. Centralizing the data, determining who is going to own it and empowering them to become the "data governor" of the organization is required. Moreover, creating one source of truth should be the goal so that from there information can be organized and extracted by each function of the business and stakeholder group.

iv) Power and utility data executives must also look ahead to compliance needs and the changes that will come with renewable energy integration and global warming, including mitigation strategies related to environmental disasters and climate change.

v) A free flow of information across the sector will also be necessary for utilities to meet customer expectations and to effectively embrace the future of the sector — such as when renewable energy (wind, hydro, solar) is purchased from customers or the increased adoption of electric vehicles (EVs).

For having a robust data governance, the energy sector based companies may consider the following questions while formulating their data strategy?

a) How to measure data quality?

b) Are key internal stakeholders are involved, including senior leaders and IT team?

c) Whether ESG sustainability trends have been identified that will impact the business?

d) Does the business have right data strategy at place to give business users the right access to the right data at the right time?

e) Is right data strategy is in place to advance self-service analytics?

f) How to comprehend the needs of all the stakeholders to deliver as per their expectations?

g) How data can be utilised as the most valuable product?

h) Who is entrusted with the responsibility of reviewing the quality data results?

4. Hospitality Sector

Numerous hotel companies are embarking upon analytics programme to improve the guest experience, maximise revenue and profits, optimize operations to control cost and enhance the value of the guest relationship. Brands are hoping to drive values for owners through robust guest databases and advanced targeted marketing programs. Hospitality companies are looking to other industries like retail, CPG and banking for inspiration on business analytics programs and analytics cultures.

Data security in hospitality sector is one of the major concerns in view of the nature of the data collected by companies operating within hospitality. Hotels, motels, resorts, and rented apartment complexes all gather and electronically store a range of sensitive personal guest data, such as names, phone numbers, addresses, and credit card details.

From the perspective of cybercriminals, hospitality appears to offer an ideal target vector for conducting crimes such as identity theft and credit card fraud due to the existence of multiple databases and devices containing both Payment Card Information (PCI) and Personally Identifiable Information (PII). In view of this, it is important to look at the following data security concerns in the hospitality sector:

a) Complex Ownership Structures: Restaurants, hotels, and other companies in the hospitality sector often have complex ownership structures in which there's a franchisor, an individual owner or group of owners, and a management company that acts as the operator. Each of these groups may use different computer systems to store information, and the information can also frequently move across those systems.

   A case in point was the Wyndham Worldwide breaches of 2008 and 2010. Hackers gained access to the systems of an individual operating company through easily guessed passwords, and the attack easily proliferated through the entire corporate network, with the result that 619,000 customers had their information compromised.

b) Payment through Cards: The nature of the hospitality industry is such that it is extremely reliant on cards as a form of payment. Restaurants and hotels alike often require credit card details for reservations, and final payment is also frequently made by the same card.

   Cybercriminals use this reliance on cards to infect point-of-sale (POS) systems with malware that steals credit and debit card information by scraping the data. In fact, it was reported in 2017 that out of 21 of the most high-profile hotel company data breaches that have occurred since 2010, 20 of them were a result of malware affecting POS systems.

   Because this malware can often proliferate or move between POS systems run by the same operator, multiple individual and groups of hotels can be afflicted by these types of attacks, and they can go unnoticed for months.

c) High Staff Turnover: A vital part of protecting data is training staff to securely gather and store personal information. Well-trained staff also know how to recognize social engineering attempts and they understand an organization's compliance requirements. The risk is that the hospitality industry involves lots of seasonal work in which people might move on after only a few months, or they might be transferred. In the U.K., for example, the job turnover rate in hospitality is as high as 90 percent.

The high level of turnover and high degree of staff movement between different locations makes it a real challenge to maintain teams of well-trained staff. All it takes is one person who isn't familiar with the importance of data security for a cybercriminal to exploit a hospitality company's systems and gain access to sensitive data.

d) Compliance: Data security risks in the hospitality industry extend far beyond the reputation hit that a hotel can take if guests' data is compromised. Industry and political regulators are becoming stricter in governing how organizations process and store personal data.

The GDPR regulation was introduced by the EU in May 2018 as a landmark legislation that aims to return control over personal information to individuals while simultaneously enforcing stricter rules for organizations in protecting such information during any period in which they possess it.

While GDPR protects individual data within the EU and EEA, its ramifications have rippled through industries globally, and organizations are realizing the need to put greater compliance measures in place.

PCI DSS is another important global regulation that protects credit card data, and fines for non-compliance begin at $500,000 per incident. The risk here is not just to data security but to the future survivability of hospitality companies, many of which would not be able to absorb the substantial losses resulting from non-compliance fines.

e) Insider Threats: This type of data risk is more subtle and it involves employees selling data to third parties without the knowledge of the organization that employs them. Such insider threats typically occur to data on customer preferences and behaviour, which hospitality companies can collect at multiple touchpoints, from interactions with their website, to form data on booking systems, to review data. This data could be potentially lucrative when it ends up in the hands of those who know how to use it to gain a competitive advantage.

Best Practices for Data Security in Hospitality Sector

Best practices for companies in the hospitality sector to protect data include:

a) Always encrypt payment card information.

b) Operate a continuous training program in cybersecurity to maintain a well-trained workforce.

c) Always adhere to relevant regulations, such as PCI DSS.

d) Use cybersecurity measures such as firewalls, network monitoring, anti-malware, and traffic filtering to protect against common threats.

e) Conduct tests against your organization's cybersecurity defenses in which you mirror the behavior of an actual hacker.

f) Know where the data is and enforce the principle of least privileges to limit access to sensitive information.

5. Information Technology (IT) Sector

With data being the new oil, there's an emerging mandate among global giants – to put into practice a data governance policy that not only restricts access to data but also governs which information should be made available and to whom. From a security and privacy point of view, data governance has gained credibility over the last couple of years with breaches and hacks becoming increasingly commonplace and organisations grappling with ransomware.

Examples of some of the leading IT companies embracing data governance models are as under:

a) **IBM:** IBM has one of the most robust data governance models and is also one of the biggest vendors of data governance solutions, which is now a strategic priority for companies of all sizes. Pegged as a quality control discipline for introducing rigour and discipline to the process of managing, using, improving and protecting organizational data, the most prized asset IBM data governance model possess the potential of significantly improving the quality and integrity of the company's data through inter-organisational collaboration and policy-making.

According to the IBM model, the core disciplines outlined are Data Quality Management, Information Security & Privacy and Information Life-Cycle Management and the supporting disciplines are Data Architecture, Audit Information & Logging and Classification & Metadata. The key enablers are organisational structure, awareness, policy and stewardship.

b) **SAS:** SAS's data governance maturity model has been around for decades and presents a data governance model that can even drive the organizational philosophy of data management, acquisition, and even archiving. Termed as more of a cultural shift that requires both business and IT sides of the organisation to come together to define data elements and the rules that should eventually govern data across applications.

According to SAS whitepaper, pockets of "data governance" emerged over the years and even grew as grassroots effort, it took off as a hot topic in the IT industry only in the early 2000s to keep pace with next-gen technology such as Hadoop and Internet of Things (IoT). And the data governance models also matured rapidly and evolved into business process management tools to deliver automated business rules.

Some of the major key constraints in data governance models outlined is a) data being considered an IT issue; b) fragmented, siloed data driven culture wherein the processes are not well-designed or understood and c) Key resources are not allocated properly.

c) **Oracle:** Oracle's data quality solutions are based on the principle of optimizing, leveraging information as an enterprise asset**.** The California-headquartered computing giant's Enterprise Data Governance solution helps identify, secure, manage and even discover sensitive data in the database. From visualisations to sensitive database discovery results to automatic metadata discovery jobs, Oracle's data governance functions provide improved quality and access and security to the core enterprise asset. The company data governance policy outlines a) establishing enterprise data strategies; b) identifying the right stakeholders and assigning accountabilities; c) outlining the report status for data-focused initiatives.

d) **Microsoft:** Redmond giant, Microsoft, has been at the forefront of data governance issues and even charged the shift from governance being a "data management" to "data as a strategic advantage" issue. According to a post by Geoff Clarke, Regional Standards Manager, Microsoft Asia Pacific data became a boardroom discussion with businesses rapidly embracing digital innovation. Over the years, data has evolved from being merely an operational "IT" issue to becoming the lifeblood of the organisation and this critical asset poses high risk, constraints and even benefits. Microsoft's data governance framework helps organizations better understand the data protocols, aligning data strategy with business goals and outcomes and how to secure data as it is rapidly moved into the cloud.

6. Telecom Sector

As the Technology, Media, and Telecommunications (TMT) industry quickly grows in the face of remote working, 5G networks, and other innovations, vast amounts of data are gathered every second, yet are rarely used in an optimal way to help an organization reach its goals.

There are missed opportunities when companies let this data pass them by, not using it to strategically

make data-driven decisions–especially at the enterprise level. Accurate, easily-accessible data can help the organization to reduce operational costs, increase sales, and get faster reporting to make informed decisions.

Technology is quickly evolving from streaming services to fiber internet, everything is moving faster. Big data in telecom is vital to understanding and acting on new technologies, adapting business strategies, and how to harness it all for both the company and customers.

With so many business opportunities, having a data governance strategy for the company's data analysts to create real-time forecasts can positively impact the business of the company. With clean, centralized data, one can increase visibility among the customers thereby creating more business opportunities, increase sales, and make better informed decisions leading to growth of the company.

The need for data governance in telecom sector is evident from the following significant statistics

Gartner reports that 80% of data governance initiatives fail to deliver on their expected outcomes. Harvard Business Review has found that over 65% of the target users for data governance initiatives do not sufficiently understand how those programs impact their roles within the organization. Finally, according to Forbes, over 74% of data leaders struggle to calculate the return on the investment of their data governance projects.

Hence, telecom sector needs to embrace a successful telecom data governance which is built on the following four pillars:

1. Link Data Governance to Business Goals.
2. Prioritize the data sets to be governed, based on the business value they provide.
3. Engage and communicate value to strategic, operational, and tactical teams
4. Leverage cultural engagement processes and people to reinforce data governance value on a day-to-day basis.

Telecom Data Governance in India- Telecom Regulatory Authority of India (TRAI) recommends Apex Body for Data Governance

TRAI has recommended the formation of the Data Digitization and Monetization Council (DDMC), an apex body to oversee all issues related to data digitization, data sharing, and data monetization in the country. The Authority further suggests that DDMC should also be entrusted with the responsibility of putting in place an overarching framework for the ethical use of data both by the Government as well as by the corporates in India.

As part of recommendations on Regulatory Framework for Promoting Data Economy through the Establishment of Data Centres, Content Delivery Networks, and Interconnect Exchanges in India, TRAI noted that a statutory body be established at the Centre that will have suitable representation from the Department of Telecommunications (DoT) and the Ministry of Electronics and Information and Technology (MeitY).

Data Digitisation and Monetisation Council (DDMC), would be formed by either amending the present law or enacting a new law, to oversee all the issues related to data, including digitisation, monetisation, sharing and storage.

CASE STUDY

Example of Airtel's Data Governance

Airtel integrates data protection safeguards into product development. All products undergo application security assessment and compliance review at the development stage. Non-compliance of any third party/ partner with the privacy practices followed at Airtel is escalated for disciplinary actions up to and including termination of the contract. The company have also been offering cybersecurity services to B2B customers, under 'Airtel Secure' which includes security monitoring through multi-layered safeguards for enterprises against threats.

Since the company deals with mammoth personal data which relates to the private, professional or public life of the customers and in light of the fact that in the online environment, where vast amount of personal data are shared and transferred around the globe instantaneously, data protection is given utmost priority.

"Bharti Airtel Information Privacy Policy (BIPP)" is in alignment with the Information Technology (IT) Rules 2011 and best practices of industry and GDPR (General Data Protection Regulation). Airtel's privacy policy provides management direction and support to ensure privacy of personal information collected by Airtel, in order to allow collection, processing, retention, dissemination and destruction of the personal information in accordance with the appropriate laws, regulations and contractual obligations.

Bharti Airtel Information Privacy Policy (BIPP) is applicable to all employees of Airtel and all third parties (including strategic partners) of Airtel who have access to personal information of customers, employees and vendors. The BIPP is applicable across all business functions of Airtel and across all geographies of Airtel in India including Airtel center, all circles and other Airtel locations.

7. E-Commerce

Today, data determines the success of both online retail and e-commerce. High-quality data is required to enable these industries to speedily make critical and accurate decisions. E-commerce businesses also capture lots of sensitive consumer data, from payments to patterns, reiterating the importance of data governance in terms of regulation and compliance.

E-commerce businesses collect data from many sources, including web analytics, email and marketing tools, online transactions, surveys and more. This information is consolidated and consumed by the relevant stakeholders. Correctly implemented data governance initiatives ensure that these stakeholders have access to data and that the data is of high quality.

Governing and integrating these data sources in e-commerce means that teams can generate actionable insights on customers, trends, products, regions and more. These insights have a huge part to play in:

- Informing gaps to adapt to consumer and market trends
- More effective customer retention and engagement
- Optimization of pricing, inventory and labour allocations
- Informing innovation as well as new and untapped market opportunities

Significance of Data Governance in E-Commerce Companies

1. **Visibility, relevance and consistency:** Digital outlets have a huge influence on the client of today. For instance, online reviews go a long way toward swaying customers. Customer journeys have also advanced with time. There is more data for e-commerce companies to record.

   Data such as customer purchases, shipping details, inventory and more are interlinked and need to be updated seamlessly all through multiple platforms and systems. This responsibility is placed on various teams that may ultimately introduce data silos or irrelevant and obsolete data.

   An efficient and unified data governance system helps e-commerce companies efficiently manage such distributed data. Such a system guarantees that data remains relevant and consistent across platforms and prevents data silos. This gives these companies greater visibility to escalate their operations.

2. **Limiting data exposure:** Data should seamlessly circulate all through the ecosystem of stakeholders in the e-commerce and retail industries. Even though restricting data may limit the effectiveness of most

industries today, the circulation of such data may give rise to safety and security concerns. Security breaches involving sensitive customer data often destroy relationships between customers and brands.

Data governance systems provide e-commerce brands with security features like two-factor authentication, data encryption and tokenization to limit the accessibility to sensitive data.

3. **Dealing with data inconsistencies:** The data repositories and warehouses kept by e-commerce companies may suffer from data inconsistencies. A change in one repository means that all the other repositories need to be updated to reflect the same change, which may become complex and overwhelming over time.

   Inconsistent data impacts sales, revenues, productivity and overall strategy. However, a robust data governance system takes advantage of data pipelines to help curate, modify and thoroughly validate raw data. E-commerce businesses resultantly enjoy better data visualization and easier, faster and more accurate data analysis.

Merits of Data Governance for E-Commerce

1. **Overall performance:** By improving efficiency and saving time, data governance raises the overall efficiency of e-commerce companies. Various teams can easily find the correct data and generate insights faster. They can also get much more accurate answers faster.
2. **Data quality**: Data governance consistently tracks data quality and usage metrics. These metrics help e-commerce companies gain visibility into how various teams and stakeholders are using data.
3. **Better business insights:** E-Commerce teams can highlight areas of subpar performance, gain a competitive advantage and discover new revenue streams using analytics.
4. **Improved decision-making:** With data governance, companies can arrive at accurate decisions speedily using high-quality data. Staff can access the correct data under the right controls to ensure that privacy and compliance are maintained.
5. **Data ownership, responsibility and accountability:** For teams to be able to use data assets effectively and appropriately, they require data governance to help them capture and share ownership, responsibility and accountability of data. Additionally, staff know who are the experts to be contacted when questions and issues arise.

   Data governance is proving to be a necessary component of e-commerce today. Correctly implementing a robust data governance solution will position e-commerce companies for new opportunities, greater security and customer confidence, more sales and e-commerce growth.

Product Data Governance Strategy for E-Commerce- A Crucial Element

For E-Commerce businesses, establishing a product data governance strategy is a crucial aspect as it has a direct effect on brand reputation and revenue. When it comes to working with product data, the important part is to list out the key elements. Once that is done, they need to be treated as strategic resources.

Some of the key points to take away when coming up with a strategy are:

1. **Ease of use for all stakeholders**

   Make sure that business users (everyone except developers who are hands-on with the product data) can utilize the product data easily. Avoid over-restricting data access.

2. **Transparent product data lifecycle**

   When product data is moving through the stages of its entire lifecycle, transparency is a key component

to the best outcome. From the initial stages of determining attributes for a particular item that will be manufactured and sold, to creating the description listing and distributing it on third-party platforms, everyone who is involved in the process should have clear communication in place that guarantees transparency.

3. **Pre-validation & assets adaptation**

   To ensure high quality of product descriptions, it's important to review the information listed. Removing outdated details and updating new ones will provide an accurate product feed that in return boosts your customers' confidence in the buying decision.

4. **Controlled delivery & coverage monitoring**

   The last step is the actual delivery process that requires accurate formatting for each endpoint the information is being uploaded to. Once all of the descriptions are successfully distributed throughout the external sales channels, it's important to double-check them. Review if there are any errors in the descriptions, confirm that the correct product pictures were chosen and make sure that all of the intended products were listed.

DATA GOVERNANCE – IMPORTANT REGULATORY DIMENSIONS

Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011

Body corporate to provide policy for privacy and disclosure of information.— (1) The body corporate or any person who on behalf of body corporate collects, receives, possess, stores, deals or handle information of provider of information, shall provide a privacy policy for handling of or dealing in personal information including sensitive personal data or information and ensure that the same are available for view by such providers of information who has provided such information under lawful contract. Such policy shall be published on website of body corporate or any person on its behalf and shall provide for— (i) Clear and easily accessible statements of its practices and policies; (ii) type of personal or sensitive personal data or information collected under rule 3; (iii) purpose of collection and usage of such information; (iv) disclosure of information including sensitive personal data or information as provided in rule 6; (v) reasonable security practices and procedures as prescribed in this rules.

As per Rule 2(i) "Personal information" means any information that relates to a natural person, which, either directly or indirectly, in combination with other information available or likely to be available with a body corporate, is capable of identifying such person

As per Rule 3 Sensitive personal data or information of a person means such personal information which consists of information relating to;— (i) password; (ii) financial information such as Bank account or credit card or debit card or other payment instrument details ; (iii) physical, physiological and mental health condition; (iv) sexual orientation; (v) medical records and history; (vi) Biometric information; (vii) any detail relating to the above clauses as provided to body corporate for providing service; and (viii) any of the information received under above clauses by body corporate for processing, stored or processed under lawful contract or otherwise: provided that, any information that is freely available or accessible in public domain or furnished under the Right to Information Act, 2005 or any other law for the time being in force shall not be regarded as sensitive personal data or information for the purposes of these rules.

The Digital Personal Data Protection Act, 2023

The Digital Personal Data Protection Act, 2023 received the assent of the President on the 11th August, 2023. The mentioned Act provide for the processing of digital personal data in a manner that recognizes both the right

of individuals to protect their personal data and the need to process such personal data for lawful purposes and for matters connected therewith or incidental thereto.

With reference to the application of the aforesaid Act, it shall-

(a) apply to the processing of digital personal data within the territory of India where the personal data is collected in digital form; or in non-digital form and digitized subsequently;

(b) also apply to processing of digital personal data outside the territory of India, if such processing is in connection with any activity related to offering of goods or services to Data Principals within the territory of India.

However, it does not apply to-

(i) personal data processed by an individual for any personal or domestic purpose; and

(ii) personal data that is made or caused to be made publicly available by—

(A) the Data Principal to whom such personal data relates; or

(B) any other person who is under an obligation under any law for the time being in force in India to make such personal data publicly available.

The salient facets of the Act are- Focus on General Obligations of Data Fiduciary, Processing of personal data of children, Additional obligations of Significant Data Fiduciary, Right to access information about personal data, Right to grievance redressal, coverage on duties of Data Principal etc.

The act is expected to have an impact on the majority of organizational areas, including legal, IT, human resources, sales and marketing, procurement, finance, and information security because of the type and volume of personal data that is collected, stored, processed, retained, and disposed of in India.

For details: *https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf*

AI Complementing Data Governance

The development of AI and machine learning in everyday business reflects the eminent role of data in management development strategies. To function effectively, AI depends on vast sets of data, which must be the subject of methodical and rigorous governance. Behind the concept of data governance lies the set of processes, policies, and standards that govern the collection, storage, management, quality, and access to data within an organization.

The role of data governance is to ensure that data is accurate, secure, accessible, and compliant with current regulations. The relationship between AI and data governance is a close one. AI models learn from data, and poor quality or biased data can lead to erroneous or discriminatory decisions.

The benefits of AI powered data governance are as under:

1. **Improve Quality of Data:** Data quality is a key to any data strategy. The more reliable the data, the more relevant the lessons, choices, and orientations that emerge from it, and AI contributes to improving data quality through a number of mechanisms. In fact, AI algorithms can automate the detection and correction of errors in datasets, thereby reducing inconsistencies and inaccuracies.

   Moreover, AI can help standardize data by structuring it in a coherent way, making it easier and more reliable to use, compare, and put into perspective. With machine learning, it is also possible to identify trends and patterns hidden in the data, enabling the discovery of errors or missing data.

2. **Automate Data Compliance:** At a time when cyber threats are literally exploding, data compliance must be a priority in an organization. But guaranteeing compliance requires constant vigilance, which can't depend exclusively on human intelligence. Especially as AI can proactively monitor potential violations

of data regulations by performing real-time analysis of all data flows – detecting any anomalies or unauthorized access, triggering automatic alerts, and even making recommendations to correct any problems.

In addition, AI facilitates the classification and labelling of sensitive data, ensuring that it is handled appropriately. Finally, AI systems can also generate automatic compliance reports, reducing the administrative workload.

3. **Strengthening Data Security:** Through its ability to proactively detect threats by analysing data access patterns in real time, AI can alert about suspicious behaviour, such as attempted intrusions or unauthorized access. To take data governance even further, AI leverages machine-learning-based malware detection systems. These systems can identify known malware signatures and detect unknown variants by analysing behaviour. Finally, it contributes to security by automating the management of security patches and monitoring compliance with security policies.
4. **Democratize Data:** At the heart of the data strategy lies one objective: to encourage employees to use data whenever possible. In this way, it can foster the development of a data culture within the organization. The key to achieving this is to facilitate access to data by simplifying the search and analysis of complex data.

AI search engines can quickly extract relevant information from large datasets, enabling employees to quickly find what they need. In addition, AI can automate the aggregation and presentation of data in the form of interactive dashboards, making information ever more accessible and easy to share.

General Data Protection Regulation – European Union

The General Data Protection Regulation is a European Union law that was implemented May 25, 2018, and requires organizations to safeguard personal data and uphold the privacy rights of anyone in EU territory. The regulation includes seven principles of data protection that must be implemented and eight privacy rights that must be facilitated.

It also empowers member state-level data protection authorities to enforce the GDPR with sanctions and fines. The GDPR replaced the 1995 Data Protection Directive, which created a country-by-country patchwork of data protection laws. The GDPR, passed in European Parliament by overwhelming majority, unifies the EU under a single data protection regime.

Any organization that processes the personal data of people in the EU must comply with the GDPR. “Processing” is a broad term that covers just about anything you can do with data: collection, storage, transmission, analysis, etc. “Personal data” is any information that relates to a person, such as names, email addresses, IP addresses, eye colour, political affiliation, and so on. Even if an organization is not connected to the EU itself, if it processes the personal data of people in the EU (via tracking on its website, for instance), it must comply. The GDPR is also not limited to for-profit companies.

The GDPR allows the data protection authorities in each country to issue sanctions and fines to organizations it finds in violation. The maximum penalty is €20 million or 4% of global revenue, whichever is higher. Data protection authorities can also issue sanctions, such as bans on data processing or public reprimands.

Organizations can comply with the GDPR by implementing technical and operational safeguards to protect personal data they control. The first step is to conduct a GDPR assessment to determine what personal data they control, where it is located, and how it is secured. They must also adhere to the privacy principles outlined in the GDPR, such as obtaining consent and ensuring data portability. You may also be required to appoint a Data Protection Officer and update your privacy notice, among other organizational measures.

(For more details on GDPR, please refer: https://gdpr.eu/faq/)

Regulatory Trends in AI Regulations

Recognizing that each jurisdiction has taken a different regulatory approach, in line with different cultural norms and legislative contexts, there are six areas of cohesion that unite under the broad principle of mitigating the potential harms of AI while enabling its use for the economic and social benefit of citizens.

1. **Core principles:** The AI regulation and guidance under consideration is consistent with the core principles for AI as defined by the OECD and endorsed by the G20. These include respect for human rights, sustainability, transparency and strong risk management. The OECD principles for AI have been discussed in the ensuing paragraphs.
2. **Risk-based approach:** The jurisdictions are taking a risk-based approach to AI regulation. What that means is that they are tailoring their AI regulations to the perceived risks around AI to core values like privacy, non-discrimination, transparency and security. This "tailoring" follows the principle that compliance obligations should be proportionate to the level of risk (low risk means no or very few obligations; high risks mean significant and strict obligations).
3. **Sector- agnostic and sector-specific:** Because of the varying use cases of AI, some jurisdictions are focusing on the need for sector-specific rules, in addition to sector-agnostic regulation.
4. **Policy alignment:** Jurisdictions are undertaking AI-related rulemaking within the context of other digital policy priorities such as cybersecurity, data privacy and intellectual property protection – with the EU taking the most comprehensive approach.
5. **Private-sector collaboration:** Many of these jurisdictions are using regulatory sandboxes as a tool for the private sector to collaborate with policymakers to develop rules that meet the core objective of promoting safe and ethical AI, as well as to consider the implications of higher-risk innovation associated with AI where closer oversight may be appropriate.
6. **International collaboration:** Driven by a shared concern for the fundamental uncertainties regarding the risks to safety and security posed by powerful new generative and general purpose AI systems, countries are pursuing international collaboration towards understanding and addressing these risks.

Other factors to consider in AI policy development include:

Ensuring regulators have access to sufficient subject matter expertise to successfully implement, monitor and enforce these policies

Ensuring policy clarity, if the intent of rulemaking is to regulate risks arising from the technology itself (e.g., properties such as natural language processing or facial recognition) or from how the AI technology is used (e.g., the application of AI in hiring processes) or both

Examining the extent to which risk management policies and procedures, as well as the responsibility for compliance, should apply to third-party vendors supplying AI-related products and services

In addition, policymakers should, to the extent possible, engage in multilateral processes to make AI rules among jurisdictions interoperable and comparable, in order to minimize the risks associated with regulatory arbitrage – that are particularly significant when considering rules governing the use of a transnational technology like AI.

Artificial Intelligence – OECD Principles

The OECD principles for Artificial Intelligence is covered under two categories-

A) Value-based Principles

B) Recommendations for Policy Makers

A) Value-based Principles

Principles	Principles Brief Description	Principles Detail
Principle 1.1	Inclusive growth, sustainable development and well-being	Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.
Principle 1.2	Human- centred values and fairness. AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and should include appropriate safeguards to ensure a fair and just society.	AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.
Principle 1.3	Transparency and explainability This principle is about transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes.	AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: ● to foster a general understanding of AI systems, ● to make stakeholders aware of their interactions with AI systems, including in the workplace, ● to enable those affected by an AI system to understand the outcome, and, ● to enable those adversely affected by AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

Principle 1.4	Robustness, security and safety AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed.	i) AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. ii) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art. iii) AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.
Principle 1.5	Accountability Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the OECD's values-based principles for AI.	AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.

B) Recommendation for Policy Makers

Principles	Principles Brief Description	Principles Detail
Principle 2.1	Investing in AI research and development Governments should facilitate public and private investment in research & development to spur innovation in trustworthy AI.	i) Governments should consider long-term public investment, and encourage private investment, in research and development, including inter-disciplinary efforts, to spur innovation in trustworthy AI that focus on challenging technical issues and on AI-related social, legal and ethical implications and policy issues. ii) Governments should also consider public investment and encourage private investment in open datasets that are representative and respect privacy and data protection to support an environment for AI research and development that is free of inappropriate bias and to improve interoperability and use of standards.

Principle 2.2	Fostering a digital ecosystem for AI Governments should foster accessible AI ecosystems with digital infrastructure and technologies, and mechanisms to share data and knowledge.	Governments should foster the development of, and access to, a digital ecosystem for trustworthy AI. Such an ecosystem includes in particular digital technologies and infrastructure, and mechanisms for sharing AI knowledge, as appropriate. In this regard, governments should consider promoting mechanisms, such as data trusts, to support the safe, fair, legal and ethical sharing of data.
Principle 2.3	Providing an enabling policy environment for AI Governments should create a policy environment that will open the way to deployment of trustworthy AI systems.	i) Governments should promote a policy environment that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate. ii) Governments should review and adapt, as appropriate, their policy and regulatory frameworks and assessment mechanisms as they apply to AI systems to encourage innovation and competition for trustworthy AI.
Principle 2.4	Building human capacity and preparing for labour market transition. Governments should equip people with the skills for AI and support workers to ensure a fair transition.	i) Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills. ii) Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programmes along the working life, support for those affected by displacement, and access to new opportunities in the labour market. iii) Governments should also work closely with stakeholders to promote the responsible use of AI at work, to enhance the safety of workers and the quality of jobs, to foster entrepreneurship and productivity, and aim to ensure that the benefits from AI are broadly and fairly shared.

Principle 2.5	International co-operation for trustworthy AI Governments should co-operate across borders and sectors to share information, develop standards and work towards responsible stewardship of AI.	i) Governments, including developing countries and with stakeholders, should actively cooperate to advance these principles and to progress on responsible stewardship of trustworthy AI. ii) Governments should work together in the OECD and other global and regional fora to foster the sharing of AI knowledge, as appropriate. They should encourage international, cross-sectoral and open multi-stakeholder initiatives to garner long-term expertise on AI. iii) Governments should promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI. iv) Governments should also encourage the development, and their own use, of internationally comparable metrics to measure AI research, development and deployment, and gather the evidence base to assess progress in the implementation of these principles.

Data Protection Seal of Data Security Council of India

The Data Security Council of India (DSCI) is planning to devise a data protection seal (DPS) to verify and check secure use of people's data by platforms across the country. The project, currently piloted with partner organisations, will help users know which organisations are using their data safely and following the basic standards of data privacy. This will be similar to the ISI mark that conforms to a product in accordance with the Bureau of Indian Standards.

The data protection seal will provide some level of assurance about the application, website, or product, according to expectations of privacy, and whether it behaves responsibly. Such a process will allow companies to better comply with the Digital Personal Data Protection (DPDP) Act and also any other upcoming rules.

One of the main challenges in today's digital landscape is the rise of deepfakes, manipulated audio or video content that can deceive viewers. DSCI aims to tackle this issue by training and certifying Data Protection Officers (DPOs) through their DSCI-certified Data Protection Officer program. These DPOs will play a crucial role in identifying and addressing deepfake-related concerns, ensuring the security and authenticity of user data.

Other several major cybersecurity challenges include the growth of ransomware, attacks on multi-factor authentication, and the use of artificial intelligence. To address these challenges, DSCI collaborates with governments, agencies, regulators, industry sectors, associations, and think tanks to advocate for cybersecurity and privacy policies and capacity-building.

The data protection seal aims to help platforms comply with the Digital Personal Data Protection (DPDP) Act and upcoming regulations. Preserving privacy while analyzing deepfake content related to sensitive issues is crucial to combat misinformation and protect user data. However, analysing content authenticity without revealing it to the platform is a significant challenge, but essential in the fight against deepfakes.

The data protection seal program is currently being piloted with partner organizations and is operational in Delhi and Bengaluru. DSCI plans to train multiple batches of DPOs to help organizations comply with the DPDP Act. This initiative will enhance data privacy practices and build trust among users, knowing their data is handled responsibly.

As the digital landscape evolves, organizations must prioritize user privacy and data protection. The introduction of the data protection seal by DSCI is a significant step towards this goal. With the increasing prevalence of deepfakes and other cybersecurity challenges, platforms must adhere to strict privacy standards to maintain user trust and protect sensitive information.

The onset of the data protection seal by the Data Security Council of India plays a crucial role in safeguarding user privacy and promoting responsible data handling by platforms nationwide. As the program expands and more DPOs are trained, it is expected to make a significant contribution to the fight against deepfakes and the overall improvement of data privacy practices. With DSCI's commitment to creating a secure and ethical data protection ecosystem, users can trust that their personal information is handled with care and responsibility.

Cyber Security breach – The Case of Sun Pharma

Sun Pharma's operations got affected by ransomware attack and a group claimed responsibility for the mentioned 'IT security incident' whose effect included breach of certain file systems and theft of certain company data and personal data, the drug manufacturer mentioned in a stock exchange filing. Sun Pharma first reported the incident on March 2, 2023. Back then it said that the incident did not affect Sun's core systems and operations. The five facts of the mentioned incident are as under:

1. On March 2, 2023, Sun Pharma reported an "information security incident" at the company, adding that the impacted assets have been "isolated".
2. 25 days later, a ransomware group claimed responsibility for the information breach. The infringement of the IT systems includes a breach of certain file systems and theft of certain company and personal data, Sun Pharma said.
3. "The Company promptly took steps to contain and remediate the impact of the IT security incident, including employing containment and eradication protocols to mitigate the threat and additional measures to ensure the integrity of its systems infrastructure and data," Sun Pharma said in a statement.
4. As part of its containment strategy, the company isolated its network and initiated a recovery process, resulting in the company's business operations being impacted.
5. As a result, revenues are expected to fall, Sun Pharma said. The company added that it is currently unable to determine other "potential adverse impacts" of the incident, including other security incidents or the possibility of litigation.

This comes amid growing threats of such attacks on Indian healthcare sector, which is the most attacked sector and is followed by education, research and government, and the military. A study by Check Point Research in January 2024 said healthcare saw the maximum number of attacks among all sectors in India, with an organisation in India being attacked 1,866 times per week on average in 2022. Global cyberattacks increased by 38% on year in 2022, it added.

From Sun Pharma's Cyber Security breach, it creates substantial academic interests to explore the reasons for vulnerability of pharma sector to cyber-attacks. Some of the reasons of cyber-attack are as under:

i) Research and development (R&D) are a top priority for pharmaceutical companies. If they want to stay ahead of their competitors, they need to constantly innovate when it comes to new drugs, treatments, and therapies. However, the IP from their clinical trials, manufacturing, and patents is especially

valuable. Cybercriminals might target these assets to sell them on the black market, to forward them to a competitor, or to use them for their own advantage.

ii) Pharma companies access a huge amount of sensitive data, including:

- Patient information
- Clinical trial results
- Proprietary research
- Regulatory filings

This valuable data is subject to stringent regulations which makes it even more appealing to people who want to monetize it. For example, cybercriminals could use this data for fraud, blackmail, or identity theft.

iii) Pharmaceutical companies work via a complex network of partners, vendors, providers, and suppliers. With so many parties involved, there are countless insider threats and opportunities for cybercriminals to take advantage of, such as by accessing databases or compromising the integrity of the products. Unfortunately, it only takes one player to compromise their data security, and the entire supply chain will experience disruption.

iv) Majority of pharmaceutical companies operates globally. This means that cybercriminals can have a significant impact across multiple countries and regions via an attack. Thus, when it comes to attack scale, the pharma world has huge potential.

v) Although more pharmaceutical companies are starting to understand cyber risks, their cybersecurity solutions aren't always as developed as in other industries. This may be due to limitation of budgets and companies may not always be proactive regarding mitigating cybersecurity challenges. The result is limited cybersecurity measures makes them more vulnerable to phishing attacks, ransomware attacks, and other cyber-attack malware.

vi) With so much sensitive data, cybercriminals have lots of opportunities to exploit pharma companies. For example, they might use ransomware attacks to encrypt valuable data and demand a "ransom" for its release. Or, they might engage in insider trading, where they access secret information on regulatory approvals or treatment research.

WAY FORWARD

Data governance is a journey, not a destination. To ensure its success, an organization needs a way to measure its progress and identify areas for improvement.

The three maturity levels that organizations go through as they become more data-driven companies are:

i) **Data integration**: application integration, data integration, and data loading

ii) **Data integrity**: data preparation, data stewardship, and data quality

iii) **Data intelligence**: data cataloging, data lineage, and metadata management

Because organizations require trusted data to empower data users, improve customer experiences, and make decisions with confidence, data quality metrics must be a core component of any data governance program. The further along an organization is on this maturity curve, the more it can take advantage of powerful technologies like data profiling and data matching with machine learning. This helps position an organization to get the maximum value out of all of its data assets while maintaining the necessary level of control and trust in that data.

It's important for business units across the organization to recognize the data governance team as a friend and ally in their business processes. After all, governance is about more than data protection and control of sensitive data. Data governance policies give business users access to the data they need, when they need it. Data governance is ultimately a tool to help optimize decision making.

While data governance is important, it should not hold back innovation or take it hostage. Modern technologies usually work best when used in concert with other technologies—like using blockchain to secure data generated by non-traditional energy producers (solar, wind, private), or integrating machine learning and semantic processing to configure energy storage or retrieval options via an intelligent chatbot. But for these kinds of "breakthrough" use cases, developers and companies need room to experiment and innovate; seasoned developers and architects are more likely to resolve potential issues during design or development phases, so data governance should have some flexibility in these phases.

As Jeff Goldblum famously said in Jurassic Park in 1993: "Life finds a way." Transformational technologies, if employed to the advantage of both customers and providers, will continue to evolve and contribute integrity to the data framework. Software developers and data architects will specialize beyond coding languages and platforms, focusing more on certain industries or business models where they can differentiate through knowledge of non-technical challenges. And new data governance approaches will have to be considered to allow innovation and protect data simultaneously.

LESSON ROUND-UP

- Data governance is everything one do to ensure data is secure, private, accurate, available, and usable. It includes the actions people must take, the processes they must follow, and the technology that supports them throughout the data life cycle.
- Data governance is the process of ensuring that the business rules for data are established and followed.
- Data management is the process of making sure that all data is captured, managed, used, and disposed of properly.
- The most common objective of data governance is the standardization of data definitions across an enterprise or organization. Other goals and objectives depend on the focus of a particular data governance program. Within the commonly accepted data governance framework, one should determine principles that make sense for the environment.
- Top-down method is the centralized approach to data governance. It relies on a small team of data professionals who employ well-defined methodologies and well-known best practices. This means data modelling and governance are prioritized. Only later is the data made more broadly available to the rest of the organization for analytics.
- The bottom-up method allows for much more agility when managing data. While the top-down method starts with data modelling and governance, the bottom-up approach starts with raw data. After the raw data is ingested, structures on top of the data can be created (referred to as "schema on read"), and data quality controls, security rules, and policies can be implemented.

GLOSSARY

Data Governance: Data governance means setting internal standards—data policies—that apply to how data is gathered, stored, processed, and disposed of. It governs who can access what kinds of data and what kinds of data are under governance.

Master Data Management: Commonly referred to as MDM, Master Data Management is a technology-enabled discipline that comprises specific tools, processes, policies, and rules to ensure one point of reference for the entire organization. MDM ensures timely, consistent, and accurate data management and distribution across your business departments, entities, and applications.

Data warehousing: It involves the storage of the organization's various data sources in internal or external databases.

BI management: or Business Intelligence management ensures that the tools, processes, and units involved are following the guidelines outlined in the data governance strategy. While BI is the process of analysing data that has been checked for accuracy and validity and delivering actionable business insights that help organizations make better decisions.

Document and content management: They are both different processes that intersect. A DMS or a document management system is used to store and retain different document formats while a content management system can handle unstructured and structured data such as web content.

Data security management (DSM): It ensures that there are measures to protect data from theft, breaches, and corruption. There are also laws in place that vary from region to region that organizations have to keep in mind while ensuring DSM.

Data operations management: It is the management of DataOps or data operations and focuses on data delivery to the organization. DataOps deals with implementing, planning, and managing a distributed data architecture that will support a wide range of tools and guidelines that have been outlined.

Data development: Itis the collation of data sets with a common objective. This means the way the data is collected has no consequence on this process. An ideal data developmental process would help the organization chart out data standards that are aligned with consistent data collection.

Data architecture management (DAM): It keeps a track of the organization's data assets and charts out the data flow. On the basis of the data flowing through several systems, DAM aims to provide a strategy for managing this data flow.

Data integration and distribution: It ensures that data is synching and integrated across all business systems, applications, and the ERP. A data integration tool can be used to integrate all data so that there are no data silos that could slow down operations or result in issues. This is also a means to distribute data within the ERP and to other legal entities.

Data quality management: It is the process of adding rules and validations to ensure that data is meeting the set criteria for accuracy, consistency, timeliness, integrity, validity, and completeness. Consistent data quality is required to ensure that any analytics performed on the data is accurate and meaningful. It would be advisable to have periodic assessments to your data to ensure data quality even with changes in validation rules.

Data stewardship: It is responsible for the accessibility, usability, and security of the organization's data. A data steward oversees all functions that come under the data lifecycle from creation to storage to deletion.

Reference data management: It is the management of classifications that distinguish data across business systems. This include tracking any changes, the creation and distribution of reference data.

Product information management: It is the management of all product-related information that is used to market and sell products through existing channels.

Workflow automation: is the process of automating task flows for documents and data across business functions adhering to the set business rules. You can use a data entry workflow tool to quicken this process.

Data management: Data governance is a subset of data management but all the areas that data governance sets out a strategy for, come under data management as well. Therefore, data management involves the collection, storage, protection, organization, correction, management, and distribution of the enterprise's data. Data management processes ensure that the data is ready to be analysed for extracting business insights that impacts the growth of the organization.

Master data: It is what we like to call the single source of truth. While master data is the content, MDM is the practice area. It is the data that is absolutely critical for day-to-day operations within a business unit or organization.

Metadata: shares distinct attributes that help describe and categorize other data within a database. There are various types of metadata such as descriptive, structural, administrative, reference, statistical and legal.

Reference data: It is a subset of master data that is used to classify other data throughout the organization.

Data migration: It is the process of moving or migrating data between systems, formats, or servers.

Data protection and compliance: It is an important process to safeguard and protect important business information from corruption or loss. Compliance ensures that there are strict guidelines that are followed to protect data and in keeping with international and local data privacy laws.

Bulk data transfer: A mechanism, usually software-based, which is designed to move large data files, supporting compression, blocking, and buffering in order to cut down on wait times.

Data staging: It ensures a place for data to be stored where it can be validated or corrected.

Database: It is the collection of the organization's data listed out that can easily be retrieved or searched via data catalogues or other means to categorize data.

Data Lake: It is a storage repository for all categories of data regardless of its size. A data lake acts as a large container for data coming from various sources into an organization, internal or external.

Data Warehouse: It is the central location of data that is integrated across systems and applications. A data warehouse stores real-time and older data and can be used to create reports and for analysis.

TEST YOURSELF

(These are meant for recapitulation only. Answer to these questions are not to be submitted for evaluation.)

1. What do you understand by data governance and explain its significance for an organization?
2. Distinguish between data governance and data management.
3. Elucidate the data governance scenario in telecom sector.

LIST OF FURTHER READINGS

- Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success First Edition by Robert S. Seiner
- Getting in Front on Data: Who Does What First Edition by Thomas C. Redman
- Data Governance: How to Design, Deploy and Sustain and Effective Data Governance Program (The Morgan Kaufmann Series on Business Intelligence) 1st Edition by John Ladley
- Get Governed: Building World Class Data Governance Programs by Morgan Templar

OTHER REFERENCES

- https://www.techtarget.com/searchdatamanagement/definition/data-governance
- https://www.forbes.com/sites/forbestechcouncil/2019/03/11/the-evolution-of-data-governance/?sh=5425c23a5ef7
- https://www.ovaledge.com/blog/what-is-data-governance
- https://www.analyticsvidhya.com/blog/2022/02/importance-of-data-governance-and-its-principles/
- https://www.spiceworks.com/tech/big-data/articles/what-is-data-governance-definition-importance-and-best-practices/
- https://www.alation.com/blog/what-is-data-governance/
- https://www.acceldata.io/article/benefits-of-data-governance
- https://blog.quest.com/the-top-7-data-governance-challenges-organizations-face-and-how-to-address-them/
- https://theecmconsultant.com/data-management-vs-data-governance/
- https://www.varonis.com/blog/data-governance
- https://www.talend.com/resources/data-governance-framework/
- https://atlan.com/data-governance-framework/#data-governance-framework-best-practices
- https://www.talend.com/resources/data-governance-framework/
- https://dmeo.gov.in/content/dgqi-overview#
- https://www.techrepublic.com/article/data-governance-in-banking/
- https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/esg-data-governance-a-growing-imperative-for-banks
- https://straitsresearch.com/report/big-data-market-in-the-automotive-industry
- https://s3-eu-west-1.amazonaws.com/xapix.io/whitepapers/Xapix+Data+Governance+2020.pdf
- https://one.oecd.org/document/DSTI/CDEP/GD(2022)3/FINAL/en/pdf
- https://www.bitraser.com/blog/data-privacy-concerns-of-automotive-industry-pave-way-for-data-destruction/

- https://agilethought.com/blogs/modern-data-governance-in-energy-sector/
- https://travel-hospitality.cioreview.com/cxoinsight/developing-a-common-business-language-the-importance-of-data-governance-in-hospitality-nid-12405-cid-40.html
- https://hospitalityinsights.ehl.edu/data-security-in-hospitality-best-practices
- https://analyticsindiamag.com/whats-data-governance-strategy-check-top-data-governance-models-leading-companies/
- https://www.to-increase.com/master-data-management/blog/data-governance-glossary
- https://www.precisely.com/blog/datagovernance/telecom-data-governance-strategies
- https://www.varindia.com/news/trai-proposes-formation-of-an-apex-body-to-oversee-data-digitization-issues
- https://economictimes.indiatimes.com/tech/technology/airtel-denies-claims-that-data-of-2-5-million-users-was-leaked/articleshow/80660207.cms
- https://www.techrepublic.com/article/data-governance-in-ecommerce/
- https://gepard.io/product-information-management/product-data-governance
- https://moneyweek.com/investments/stockmarkets/china-stockmarkets/605136/the-fallout-from-alibabas-huge-data-breach