Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.10 Issue 1 Page No 19-25

Survey on Data Management Challenges by Utilities and Role of Data Governance

[1] Mukesh Kumar Bansal1, Rahul Vijayvargiya [2], M. K. Gupta [3]

[1] Technology Partner and Business Leader, MNC UK, mkbansal.edu@gmail.com
[2] Senior Project Manager, MNC UK rahulvijayvergiya01@gmail.com
[3] Department of Electrical Engineering, Suresh Gyan Vihar University, Rajasthan, India, mkgupta72@gmail.com

Abstract:

In the modern business landscape, data governance has emerged as a critical component in digital transformation across different industries, as it can facilitate decision-making, enhance operational efficiency, and increase regulatory compliance. The global data governance market size is projected to grow from $3.66 billion in 2023 to $14.53 billion by 2030, at a CAGR of 21.8% during the forecast period [17]. The urgent need for decarbonisation, the strategic move towards achieving legally binding net zero targets, need for open data and various new regulatory compliances are driving transformation in the utilities industries. Data governance emerges as a critical enabler of this transformation. Various organizations and research industries have come up with data governance solutions /frameworks. This paper examines the current landscape of the utilities sector, highlighting the key trends and challenges that are driving the transformation. It then provides details about the key capabilities of a comprehensive data governance solution. This is followed by an analysis of data governance solutions offered by different vendors in the market and their suitability for the utilities sector. Through a comprehensive exploration of data governance, this white paper provides valuable insights for utilities leaders, data professionals, and policymakers.

1. Introduction

Utilities play a pivotal role in our daily lives, powering our homes, businesses, and essential services. They form the backbone of modern society, ensuring the smooth functioning of our day-to-day activities. However, the integration of distributed energy resources, such as solar panels and wind farms, the regulation to provide flexible tariff options such as Static and Dynamic Time of Use Pricing to customers and the EV revolution are putting new demands on the grid. Additionally, utilities face stringent regulatory reporting and compliance requirements.

In this new landscape data is key to unlock system and consumer benefits. As per Energy Data Taskforce report "Effective storage, sharing and management of data will allow the markets to develop that will put consumers at the heart of this change while allowing networks to support the proliferation of new business models and technologies. Interoperability, the virtue that allows different organizations to share and understand information, is critical too if we are to solve the complex challenges of decarbonising energy, heat and transport [1].

The first part of this paper examines the current landscape of the utilities sector, highlighting the key trends and challenges that are driving the need for transformation. The second part covers building blocks of a data governance solution along

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com

with examples of how data governance capabilities are being used in utilities. The third part provides an overview of data governance solutions from industry leaders along with a comparative study of these solutions. The fourth part provides the author's view on an ideal framework for a data governance solution to address challenges faced by utilities.

2. Need for Transformation of the Grid

The urgent need for decarbonisation and the strategic move towards achieving legally binding net zero targets are driving several changes in the energy ecosystem. The traditional model of big power stations connecting to one end of the transmission system and the distribution networks connecting to the other end is being replaced. Instead, a greater number of smaller generation assets, like wind farms, are being connected directly to the Grid. Additionally, the rise of prosumers (consumers who also produce electricity) and the increasing prevalence of microgrids and renewables are changing the energy ecosystem.

This shift represents a significant change in the way electricity is generated and distributed, requiring a new approach for managing and operating the grid. It is estimated that an investment of $1.1 trillion is required every year till 2050 to reach the net zero goals. The world will need four times as much electricity generation as it has today, and three times as much transmission capacity [2].

The role of utilities is evolving in this new energy landscape. Governments in different countries have introduced new regulations to digitalise the energy sector. As an example below is a list of few key recommendations from the UK Government in 'Energy Data Taskforce report'[1]:

1. The energy system data should be presumed open and data should be discoverable, searchable, understandable', with common 'Structures, Interfaces and Standards'. Additionally, data should be 'Secure and Resilient'.
2. A Data Catalog should be established to provide visibility through standardized metadata of Energy System Datasets across Government, the regulator and industry.
3. A unified Digital System Map of the energy system should be established to increase visibility of the energy system infrastructure and assets, enable optimization of investment and inform the creation of new markets.

As per an IBM study [13] utilities are behind industry peers in aligning their data strategy with business strategy. So, there is an opportunity for utilities to increase strategic focus on data governance and use AI to automate decision making.

3. How Data Governance can facilitate Transformation

Utilities often face challenges in providing reliable, high-quality data promptly due to increasing data privacy and industry regulations. Data governance enhances data transparency, trust, and understanding, promoting faster insights while ensuring data protection. It simplifies access to data for appropriate users while meeting privacy and compliance needs.

Key benefits of Data Governance:

1. Empowers data consumers to find and access the relevant data by enabling self-service data consumption
2. Helps organizations address compliance requirements by reducing the time and effort required to comply with new industry-specific regulatory policies and governance rules.
3. Helps facilitate collaboration on data assets between business and technical users
4. Higher levels of data quality and trust, through data governance, can help utilities to leverage Artificial Intelligence to optimize operations and improve customer experience.

There is a lack of agreement in the industry regarding the key capabilities required to implement data governance. Some organizations focus on implementing governance to comply with regulatory requirements whereas others implement data governance to derive value from data for business growth. So, the way Data Governance is implemented might differ based on organizational policies and priorities, but in our opinion below mentioned are the key roles and capabilities required to implement a comprehensive Data Governance program.

3.1 Data Governance Roles

Any successful Data Governance initiative requires collaboration between Business and IT teams, hence requires representation from both the areas. Below mentioned are the key roles, in our opinion, for a Data Governance program:

- **Data Governance Council:** The data governance council is responsible for allocating resources and budgets, and monitoring data governance performance and outcomes in an organization. It consists of senior executives and stakeholders, usually from business, who provide strategic direction, sponsorship, and oversight for data governance initiatives.

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com

- **Data Owners**: Data Owners are responsible for approving data policies, standards and procedures. They decide who should have access to the data. Additionally, they have the authority to grant or revoke data access and usage rights.

- **Data Stewards:** Data Stewards are responsible for implementing and enforcing data policies, standards, and procedures as part of day-to-day operations of a Data Governance program. Data stewards are usually subject matter experts or business users who understand the data needs, quality, and usage from a business perspective.

- **Data Engineers:** Data engineers are responsible for providing tools that help secure data, manage data quality, and integrate data from a variety of sources.

3.2 Data Governance capabilities:

While implementation of Data Governance can vary depending on the industry and business priorities, below in our opinion are the key capabilities for a successful Data Governance program [4]:

3.2.1 Data Profiling: Data Profiling involves analyzing data, generally using software-based tools, to identify anomalies, inconsistencies, relationships, and redundancies. It's a crucial step as it helps in understanding the quality, consistency, and integrity of data. Data Profiling is generally performed by Data Engineers.

As utilities collect data from many systems, Data Profiling can help identify data quality issues such as missing values in data collected from smart meters or a wind turbine sensor, highlighting a potential issue with the device. Moreover, it can also highlight data format issues between different tables/ data sources which might hinder the ability to combine the data from these sources. Furthermore, it can help to understand if energy usage is very high in certain areas by identifying the relationship between usage data and location which can help to plan for future needs at these locations.

3.2.2 Data Quality Management:

Data Quality Management (DQM) refers to the processes and technologies used to measure, improve, and control the quality of data. For example, if Data Profiling has revealed issues with data from smart meters or wind turbines, DQM process can be invoked manually which might involve triggering asset maintenance workflow to remediate or replace faulty devices. It might also involve setting up rules and alerts around the device data and in case data from devices or sensors is missing for a specific duration they might trigger alarms and an automated workflow for maintenance.

3.2.3 Data Integration:

Data integration involves combining and harmonizing data from various sources into a unified, coherent format which can be used for various analytical, operational, and decision-making purposes. Utilities have access to many data sources, such as operational technology data from Supervisory Control and Data Acquisition (SCADA) & Geographic Information System (GIS), asset data from Enterprise Resource Planning (ERP) systems, customer data from CRM systems, environmental data from weather stations and climate models, and technical data from equipment manufacturers and industry standards. But this data exists in silos across different departments, in different systems and formats. They can combine these data sources to create a common data hub which can be used across the organization as a single source of data for any operational decision making. Engie, the French energy company, has built a common data hub by combining various data sources to drive innovation across the whole organization [5].

3.2.4 Data Catalog:

A data catalog is a detailed inventory of all data assets in an organization, designed to help users quickly find the most appropriate data for any analytical or business purpose. The Energy Data Taskforce Report highlighted the need for Data Catalog as one of its key recommendations. For utilities, a Data Catalog can improve visibility of data assets across the industry, helping increase innovation and optimisation of assets. For example, using near real time data from grid operations during times of congestion, a supplier can help in alleviating strain on the grid by incentivizing its customers for voluntarily flexing the time when they use electricity. This service is currently used in the UK by NG ESO to manage demand on the grid during peak times [6].

Another example of how visibility of assets and open data can drive innovation is through understanding of the Australian Renewable Energy Mapping Infrastructure (AREMI) tool. Using the geo-spatial data from the tool a business can understand average solar coverage, local transmission infrastructure, grid capacity and local buildings in a specific location to make a decision regarding planning a new solar farm [7].

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com

3.2.5 Master Data Management:

MDM provides the ability to intelligently connect multi-domain data such as customers, assets and location. By delivering access to accurate views of master data and their relationships, MDM provides a trusted view of data that can be shared across the business to promote accurate reporting, reduce data errors, and facilitate cross-sector collaboration.

Utilities have to manage a wide range of assets like transformers, cables, poles, meters, fibers etc. These assets are often managed in different systems. With MDM, utilities can have a unified view of each asset, linking data like the asset's specifications, its location, its maintenance history, and its performance data. This can help them to manage and utilize their assets more effectively. Let's understand through an example of how better visibility of assets is facilitating cross-sector collaboration and ultimately better asset utilization. To ensure future water supplies, by 2030 the water industry must reduce leakage by 1 billion liters/day. Existing leak detection solutions such as acoustic loggers (sensors that detect water leaks and raise concerns) are not only expensive to deploy but also suffer from a high false alarm rate (as high as 60%). The Dark Fibre project, funded by the Ofwat Innovation Fund, investigated the feasibility of using existing optical fiber networks for leak detection. The system uses a laser to detect noise along 'dark fibers' (unused fiber optic cables) that can be deployed inside or alongside a water main. But rather than deploying new fiber, which would be expensive, they found that a significant fiber network already exists, close to water industry assets, for systems such as broadband and railway signaling. The experiments proved that the currently dormant 'dark fiber' could be used for water leakage detection, lowering the cost and environmental impact. It demonstrated that leak noise could be detected by a fiber up to 5m away [14].

3.2.6 Data Lineage:

Data lineage refers to recording the life-cycle of data, including its origins, movements, transformations, and interactions with different processes and systems. This type of documentation enables users to observe and trace different touch points along the data journey, helping ensure data quality and meet regulatory requirements.

For utilities, a Data lineage tool can help ensure compliance by keeping a clear record of how sensitive customer data from Smart Meters or other CRM systems are handled. This will help them to comply with data privacy laws such as GDPR and CCPA and avoid fines. As per report from global law firm DLA Piper, Data protection supervisory authorities across Europe have issued a total of EUR 1.64bn in fines since 28 January 2022[8]. Additionally, Data lineage tools can help maintain data quality by providing a clear traceability of data movement across various systems. For example, if there's an error in the billing process, utilities can use the data lineage tool to trace back the billing data to its source, identify where the error occurred, and correct it.

3.2.7 Data Access Management:

Data Access management enables organizations to define ownership of data, implement access controls and secure confidential information. It is a crucial element of overall Data Governance as well managed data access significantly helps users in collaborating on data assets while ensuring compliance.

Let's understand how utilities are implementing a comprehensive Data Access Management framework to achieve compliance. For efficient network planning and to transition to a low carbon economy, utilities need granular level access to smart meter data. But as per standard license condition 10A (SLC10A), there are prohibitions on processing electricity consumption data which relates to a period of less than one month [10]. To comply with these conditions, Scottish and Southern Electricity Networks (SSEN) has implemented a comprehensive data access management framework. Their system is controlled using role-based access controls mechanism. Only authorized users can access the system and have the ability to request customer consumption data. They have implemented an automatic aggregation logic that prevents a user requesting a more granular level consumption data than what has been set by the system administrators to avoid unintentional individual consumers consumption data being requested [11].

3.2.8 Business Glossary:

A business glossary provides definitions for common terms and metrics used within the organization or more generally within an industry. It ensures everyone in the organization or the industry uses data consistently and understands it the same way. A business glossary helps in reducing misunderstandings and promotes data literacy.

A business glossary can facilitate regulatory compliance in utilities by ensuring consistent understanding and usage of regulatory terms. For example, utilities are required to report on 'Embodied Carbon' as per OFGEM environmental reporting guidelines [12]. Embodied Carbon refers to the carbon dioxide emitted during the manufacture,

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com

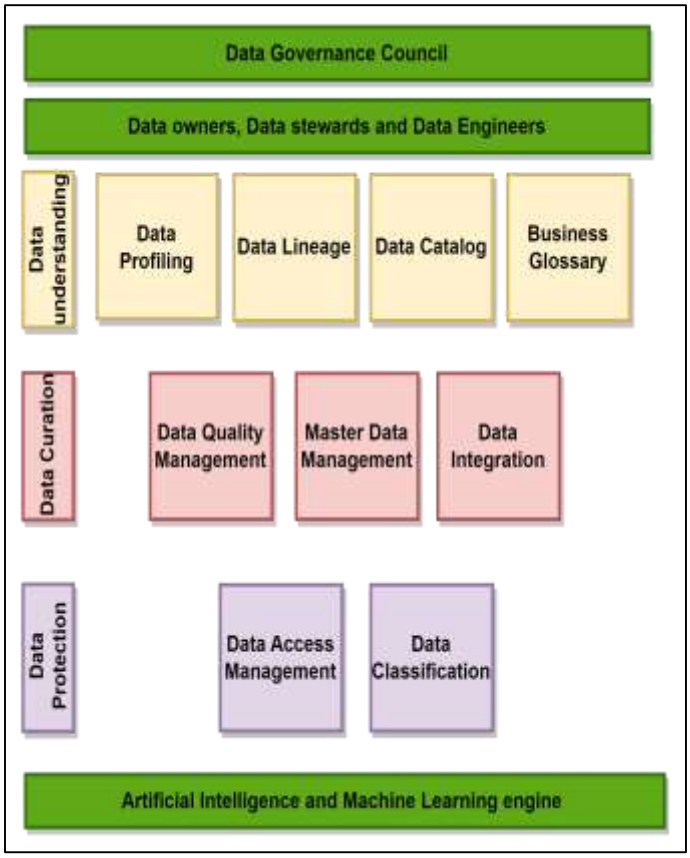
transport, and construction of building materials, together with end-of-life emissions. By following the official definition, utilities can ensure consistent and accurate reporting. This can help avoid miscommunication, misreporting, and potential non-compliance penalties.

3.2.9 Data Classification:

As per IBM study, Utilities CDOs (Chief Digital Officer) position greater responsibility, importance, and confidence around data security-oriented topics, reflecting the strong focus on security for the industry [13]. Data Classification is crucial for information security and compliance as it regulates access to sensitive data such as PII or data from Critical National Infrastructure assets. Data objects are classified based on criteria like privacy, confidentiality, sensitivity, and compliance regulations, ensuring secure access control.

Utilities can break up the data into categories such as customer, employee, financial, and operational data from the grid. Once data has been classified, they can use data classification tools and procedures to classify data under each category such as 'confidential', 'sensitive' and 'unclassified', making it easier to tackle the regulatory requirements associated with each category individually. For example, they can use operational data from the grid to report on 'Electricity Distribution Losses' as per the regulatory guidelines while customer data can be used to show compliance with regulations such as GDPR and CCPA.

To summarize, below is an AI augmented capability map of data governance to serve the needs of different business personas. The capabilities are categorized under three important pillars: data understanding, data curation and data protection.



4 Data Governance solutions from Vendors

4.1 Collibra:

Collibra is a leader as per The Forrester Wave Data Governance Solutions 2023 report [15]. Collibra's unified offering 'Collibra Data Intelligence Platform' provides a single system of engagement for data that covers a wide range of business use cases and personas. This single system enables organizations to find, access, understand and trust data. Collibra Data Intelligence Platform provides an integrated suite of key data governance capabilities such as Data Catalog, Data Lineage, Data Quality, Data Privacy and AI Governance, and covers both Governance and Analytics use cases. Collibra will suit customers looking for a one-stop-shop data governance solution that supports complex data discovery to data product journey use cases.

4.1.1 Strengths:

1. A single platform providing integrated set of capabilities covering wide range of business use cases and serving needs of various business personas
2. Strong capabilities in policy management & compliance and workflow automation

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com

3. Strong partner ecosystem with integrations with leading cloud providers such as AWS, Google Cloud, Microsoft, Snowflake etc
4. Presence across multiple industries such as 'Energy and Utilities', 'Financial Services', 'Healthcare', 'Insurance', 'Life sciences', and 'Retail'.
5. Support for both Governance and Analytics use cases.

4.2 Alation:

Alation is a leader as per The Forrester Wave Data Governance Solutions 2023 report [15]. Rather than having a single product with an integrated suite of capabilities, Alation provides multiple products such as Data Catalog, Data Lineage, Data Governance and Analytics. Alation focuses on AI-fueled automation of policy management, enhanced workflow, and analytics capabilities. It has strong capabilities in the field of machine learning (ML) and intelligent asset classification capabilities with a deep focus on collaboration tools and data valuation models. Additionally, it has strong training programs and dedication to build and maintain client communities to enhance data literacy. So, in our opinion, customers looking for strong data discovery and cataloging capabilities accompanied by strong lineage and collaboration tools should consider Alation.

4.2.1 Strengths:

1. Industry leading Machine Learning capabilities
2. Extensive training programs to enhance data literacy.
3. Industry leading Data Cataloguing capabilities powered by sophisticated Machine learning
4. Support for multiple cloud deployment options.
5. Strong UX and Usability
6. Strong Analytics use cases

4.3 Informatica:

Informatica is a longtime player in data management and used widely across industries such as energy & utilities,education, financial services, healthcare, life sciences, manufacturing, retail and telecoms. Informatica's Data Management Cloud platform provides an integrated suite of capabilities such as data catalog, data integration, data quality, MDM, privacy and governance etc. The platform is modular i.e. you can start with a single capability and grow at your own pace. The platform offers capabilities that enable data management across various sources, including on-premises, cloud, multi-cloud and multi-hybrid environments. The data management platform's core capabilities are augmented by their AI engine CLAIRE (AI-powered metadata and intelligent automation engine) which sits horizontally. CLAIRE augments the capabilities of a broad range of users. For example, it helps data analysts by making it easier for them to locate and prepare the data, and helps business users quickly identify data that should be subject to compliance controls. Informatica's data governance solution will suit large enterprises looking for a proven track record and for enterprises who are already using its products and looking to extend capabilities.

4.3.1 Strengths:

1. Vast library of data integrations including leading cloud providers such as AWS, Microsoft Azure, Google Cloud, Salesforce and Oracle etc.
2. Capabilities to cover a range of use cases from large scale enterprise deployment to customized projects such as unstructured data discovery
3. Strong AI powered automation capabilities to support various use cases and business personas
4. Established player with proven record across a range of industries including Energy and Utilities.

4.4 IBM:

IBM is a longtime player in Data Governance with a rich history of serving enterprise clients. 'IBM Cloud Pak for Data' is a platform with an integrated suite of services to provide a comprehensive data management solution covering the needs of various business personas. The key service of the platform is IBM knowledge catalog. IBM Knowledge Catalog provides capabilities for data discovery, data quality management, data lineage, data cataloging, and data protection across a distributed data landscape. The platform is modular and provides additional services across different areas to augment the capabilities of IBM Knowledge Catalog, some of which are included with the platform and others can be purchased separately. For example, 'IBM Cloud Pak for Data' includes AI services such as 'IBM Match 360' which helps to build a 360-degree view of the customers by matching data from different sources. The platform also provides options to include other AI services for advanced use cases. For example, 'watsonx.ai' can be integrated to build generative AI solutions but this service needs to be purchased separately. Moreover, the platform includes a knowledge accelerator for energy and utilities 'IBM Knowledge Accelerator E&U'. The IBM® Knowledge Accelerator for E&U is an industry-specific glossary for E&U organizations that is imported to IBM Watson Knowledge Catalog. The knowledge accelerator provides alignment and lineage to CIM 61968 and 61970 industry standards which helps in accelerating auto classification of various data sources [16].

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com

Strengths:

1. Modular architecture to build a data governance solution based on specific needs and use cases
2. Supports various deployment options to suit different business needs such as on-prem and public cloud including IBM, AWS, Azure and GCP
3. Strong industry alignment with E&U industry through various case studies and knowledge accelerators.
4. Extensive roadmap to enhance data governance life-cycle through WatsonX AI solutions.

5 Recommendations on evaluating a data governance solution for utilities:

After analyzing solutions from different vendors, it is quite evident that most of the vendors interpret data governance differently and there is no common standard/ set of capabilities for assessing data governance solution maturity across vendors. Moreover, many vendors offer a range of products with overlapping sets of capabilities, making it difficult to evaluate and choose the solution to suit a specific business need. Furthermore, in our opinion, the majority of vendors have generic data governance solutions which might require a lot of customization to suit the needs of the energy and utilities industry. Lastly, in our opinion, it is not very clear how current solutions address data sovereignty and country/ region specific. compliance needs. Hence we have recommended key capabilities for a comprehensive data governance solution for E&U in section 3 above. Moreover in our opinion below additional points should be considered by E&U leaders before investing in a data governance solution:

1. Consider tools with specific accelerators and solutions for utilities such as alignment and to CIM 61968 and 61970 industry standards and adherence to country/region specific data protection rules.

2. Prepare a checklist of specific needs and evaluate tool capabilities and track record to meet those specific needs. Since utilities are required to share data with wide range of market participants to meet regulatory and market needs while protecting sensitive data, the tool should have strong data cataloging and data protection capabilities

3. Evaluate the solution keeping in mind the needs and capabilities of various business personas, the tool should provide an intuitive user interface and strong workflow automation capabilities.

4. As utilities are behind industry peers in using AI to automate decision making, look for tools with AI powered automation across various data governance capabilities such as data cataloging, classification etc.

References:

1. A strategy for a Modern Digitised Energy System, Energy Data Taskforce report
2. The electric grid is about to be transformed, The Economist
3. Definition of Data Governance- Gartner
4. What is Data Governance? AWS
5. Building the Common Data Hub on AWS, Accelerates Zero-Carbon Transition | ENGIE Case Study, AWS
6. Demand Flexibility Service: Live Events | ESO
7. Australian Renewable Energy Mapping Infrastructure (AREMI) | CSIRO Research
8. European data regulators issued over EUR 1 billion in GDPR fines: Ireland now ranked second highest, DLA Piper survey reports
9. Data Best Practice Guidance, OFGEM
10. Electricity Distribution Consolidated Standard Licence Conditions, OFGEM
11. Smart Meter Data Privacy Plan, SSEN
12. RIIO-ED2 Environmental Reporting Guidance, OFGEM
13. Data Governance in the Energy & Utilities Industry, IBM
14. Dark Fibre - Ofwat Innovation Fund, OFWAT
15. Forrester
16. Common Information Model (CIM), ENTSO
17. Data Governance Market Size, Share | Trends Analysis [2030]

Correspondence: M K Gupta, Department of Electrical Engineering,, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: mkgupta72@gmail.com