1. **Data Acquisition (DA):**
    - **Electronic Reporting of Energy-Related Data:** Facilitating the digital sharing of crucial energy data from verified sources [1] [2] [3] [4].
    - **Effective Integration of Heterogeneous Data:** Integrating various types of energy data from different sources (RFID, GID, GPS, camera, sensors, etc.), seamlessly into a cohesive system [5] [6] [7] [8] [2] [9] [4] [10] [11] [12].
    - **Effective Device Communication Protocols:** Ensuring seamless data exchange and quick response times through machine-to-machine and human-to-machine communication protocols [13] [8].
    - **Effective Collection and Transmission of Large Volumes of Data**: Facilitating the gathering and distribution of substantial data volumes across various energy systems with different data characteristics (like file formats, size, and growth rate) [14] [15] [16] [11].
    - **Support for DOIs:** Facilitating the use of Digital Object Identifiers (DOIs) for uniquely identifying data products [17].
    - **Easy Integration of Third-Party Services and Co-existence of Multi-Vendor Technologies:** This requirement is about ensuring interoperability across various platforms and technologies [4].

2. **Data Processing (DP):**
    - **Real-Time Processing of Data:** Ensuring immediate (high data rates) and simultaneous processing of energy data for timely and efficient decision-making [14] [18] [15] [19] [7] [20] [21] [22] [23] [11] [12] [24].
    - **Homogenizing Information:** Standardizing data models and incorporating data analysis to ensure consistency and relevance of data [6] [15] [25] [7] [10].
    - **Processing Large Volume of Data:** Processing a large volume of data generated by the different energy systems [19] [25], this include employing batch processing techniques [15] [20] [26] [8] [22] [9] [4].
    - **Development of reliable and fault-tolerant data management solutions:** Ensuring data is not lost or corrupted in the event of hardware failures, network issues, or other storage-related problems [27].
    - **Cloud Computing for Data Processing:** Leveraging the processing power and scalability of cloud computing for handling and analyzing energy data [27].

And for integrating heterogeneous data from stationary devices, UAV cameras, and online map services [22] [4].

- **Pre-processing of diverse data:** It involves identifying the types of data, filtering irrelevant data, classifying data for further processing, and optimizing it for analysis. This is crucial for managing the diverse data types in an energy ecosystems [22].
- **Data Cleaning and Automatic Correction:** Implementing functions for data cleaning and automatic data correction. This involves removing inaccuracies, inconsistencies, and redundancies from data sets, as well as applying algorithms to automatically correct errors [9].
- **Preprocessing services that identifies anomalies:** Once data is acquired, identifying and dealing with anomalies such as outliers, outliers, duplicates, and inconsistent or incomplete data or missing values is essential for maintaining data quality [3] [23] [4] [12] [24].
- **Adoption of Ad Hoc Data Science Solutions:** Having an ad Hoc data science solutions will help exploiting the full potential of available data [4].

3. **Data Storage (DS):**
   - **Data Reduction and Compression:** Optimizing energy data by reducing redundancy and compressing information without losing essential details [14] [28] [10] [24].
   - **Key-Value Databases for Efficient Data Storage:** Utilization of key-value databases for efficient data storage [27].
   - **Use of cloud computing:** Utilizing cloud infrastructure for storing large volumes of energy-related data [20].
   - **Appropriate Structuring and Archiving of Collected Data:** This involves organizing data in a structured manner that facilitates easy access and analysis, and implementing archiving strategies to ensure long-term preservation and retrievability of data [21] [4].
   - **Flexible Data Storage Options:** Providing the option to store data either in cloud-based storage solutions or on-premises (local storage) [2].

- **Robustness and Disaster Recovery in Database Management Systems:** a disaster recovery plan is needed for database management systems to handle the increasing data size and ensure data integrity [10] [24].
- **Ensuring Datasets are Stored in Readily Consumable Formats:** Ensuring that datasets are stored in a readily consumable format, such as generic formats like images and tables, to facilitate data consumption and transferability [29]

4. **Data Analysis (DAS):**

- **Data Visualization and Interpretation:** Developing tools and methods for clear and insightful visualization and interpretation of energy data. This includes creating user interfaces (GUIs) for complex searches and quality reporting [14] [17] [15] [30] [3] [31].
- **Distributed search indexes:** Building distributed search indexes to facilitate efficient data analysis [17].
- **Data-Driven Knowledge Extraction:** Focusing on extracting valuable insights and knowledge from (raw) data, turning it into a reusable resource [18] [30] [21].
- **Big Data Processing and Interaction:** Enabling the processing of large datasets and facilitating human interaction with these datasets. This involves implementing download-oriented approaches with tools for cataloging, publishing, and visualizing big data, ensuring that both the data processing and the user experience are optimized for large-scale data sets [17].
- **Implementation of Advanced Analytics Techniques:** Advanced analytics techniques such as predictive analysis and classification are key components of data analysis, focusing on extracting insights and making predictions based on the available data [6] [27] [16] [10]..
- **Insights into Data Quality:** Monitoring and providing detailed analysis regarding the quantity and quality of data/products [17] [32] [15] [25] [33] [2] [3].

5. **Data Lineage (DL):**

- **Centralized Data Management:** Streamlining energy data into a unified, secure, and transparent system for better oversight and control (solve data distribution challenge) [1] [12].

6. **Data Traceability (DT):**

   - **Data Integration During Query Processing:** Ensuring that the integration of data during query processing is traceable. This means keeping track of how data from different sources is combined and used in queries (e.g., using federated query engines) [34].
   - **Controlled and Secured Data Exchange:** Focusing on secure and traceable data exchange. This involves implementing measures to control data sharing and ensure that data transfers are both secure and traceable [34].

7. **Metadata Management (MDM):**

   - **Comprehensive Metadata for interpretation and reproducibility:** Providing extensive contextual information with each dataset for enhanced understanding and application, by using an established language for metadata based on standards and data models [1] [17] [32] [18] [34] [4] [29].
   - **Data Integrity Mechanisms:** Maintaining the accuracy and reliability of energy data through rigorous integrity checks [13] and using peer validation big data [23].
   - **Scalability:** Focusing on the ability of energy data systems to adapt and grow efficiently to meet increasing demands [35] [14] [28] [6] [26] [16] [34] [4] [10].
   - **Incorporation of FAIR Principles:** Integrating Findable, Accessible, Interoperable, and Reusable (FAIR) principles into the system [17] [30].
   - **Semantic Web Technologies:** Utilizing semantic web technologies for enhanced data management and monitoring, making it easier for both technical and non-technical stakeholders to engage with the data [18] [34].
   - **Comprehensive Tagging System for Big Data:** Effectively tagging large volumes of data for easy identification, sorting, and retrieval [29].

8. **Data Security (DSC):**

   - **Data Security and Privacy Measures:** Implementing security and privacy protocols, including the use of access control, data anonymization, unique identification systems, authentication, and encryption, to safeguard sensitive energy information [13] [14] [6] [25] [7] [30] [20] [8] [16] [34] [31] [36] [11] [24].

- **Privacy-Preserving Techniques and Tools**: Addressing privacy concerns associated with big data and IoT devices through specialized techniques and tools [13] [27].
- **GDPR Compliance for Personal Data:** Mandating prior consent for processing personal data, in line with GDPR regulations [28] and international law [24]. This includes implementing data scrubbing mechanisms to remove personally identifiable information (PII), ensuring that privacy is maintained while still providing efficient access to data for different types of data consumers [19]
- **Scope of Data Collection:** Ensuring the collection of personal data does not exceed the scope of given consent [28].
- **Open APIs for Data Publishers and Consumers:** Open APIs can be crucial for secure and efficient data exchange between different systems and stakeholders, ensuring interoperability while maintaining security and compliance [4].
- **Identifying and Storing Attack Paths:** It involves the proactive identification and documentation of potential cybersecurity threats or attack vectors in energy-related operations [36].

9. **Data Governance and Compliance**:
   - **Supervision and Auditing:** Conducting regular and impromptu audits to ensure compliance, quality and accuracy in energy data practices [1] [25].
   - **Penalties for Data Falsification:** Enforcing strict repercussions for any manipulation or falsification of energy data [1].
10. **Human Skills and Expertise in Data Management:** (I added this part because it was mentioned in more than 3 articles (I am still in the first 20 articles)
    - **Collaboration for Interoperability:** Emphasizing the necessity of collaboration between energy sector stakeholders and technology providers to achieve interoperability. [8].
    - **Need for Expertise in DM:** Highlighting the requirement for specialized expertise in data management. This includes understanding the nuances of data collection, processing, storage, and analysis, and staying updated with the latest DM technologies and practices [5] [26] [8].

Bibliography

[1] Bo Zhang et al. "Big data challenges in overcoming China's water and air pollution: relevant data and indicators". In: *SN Applied Sciences* 3.4 (2021). Cited by: 6; All Open Access, Gold Open Access, Green Open Access. DOI: 10.1007/s42452-021-04448-0.

[2] Tibor Horak et al. "Data Integration from Heterogeneous Control Levels for the Purposes of Analysis within Industry 4.0 Concept". In: *Sensors* 22.24 (2022). Cited by: 0; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/s22249860.

[3] Giovanni Gravito De Carvalho Chrysostomo et al. "A framework for big data analytical process and mapping—baprom: Description of an application in an industrial environment". In: *Energies* 13.22 (2020). Cited by: 5; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/en13226014.

[4] Marco Pau et al. "MATRYCS—A Big Data Architecture for Advanced Services in the Building Domain". In: *Energies* 15.7 (2022). Cited by: 8; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/en15072568.

[5] Dominik Kozjek et al. "Advancing manufacturing systems with big-data analytics: A conceptual framework". In: *International Journal of Computer Integrated Manufacturing* 33.2 (2020). Cited by: 28; All Open Access, Green Open Access, Hybrid Gold Open Access, pp. 169–188. DOI: 10.1080/0951192X.2020.1718765.

[6] José L. Hernández et al. "Interoperable open specifications framework for the implementation of standardized urban platforms". In: *Sensors (Switzerland)* 20.8 (2020). Cited by: 17; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/s20082402.

[7] Nazmus Sakib, Eklas Hossain, and Sheikh Iqbal Ahamed. "A Qualitative Study on the United States Internet of Energy: A Step towards Computational Sustainability". In: *IEEE Access* 8 (2020). Cited by: 19; All Open Access, Gold Open Access, Green Open Access, pp. 69003–69037. DOI: 10.1109/ACCESS.2020.2986317.

[8] Ibrahim Mashal et al. “The determinants of reliable smart grid from experts’ perspective”. In: *Energy Informatics* 6.1 (2023). Cited by: 0; All Open Access, Gold Open Access. DOI: 10.1186/s42162-023-00266-3.

[9] Haijun Yu et al. “Big-Data-Based Power Battery Recycling for New Energy Vehicles: Information Sharing Platform and Intelligent Transportation Optimization”. In: *IEEE Access* 8 (2020). Cited by: 21; All Open Access, Gold Open Access, pp. 99605–99623. DOI: 10.1109/ACCESS.2020.2998178.

[10] Ameema Zainab et al. “Big Data Management in Smart Grids: Technologies and Challenges”. In: *IEEE Access* 9 (2021). Cited by: 17; All Open Access, Gold Open Access, pp. 73046–73059. DOI: 10.1109/ACCESS.2021.3080433.

[11] Bokolo Anthony Jnr. “Smart city data architecture for energy prosumption in municipalities: concepts, requirements, and future directions”. In: *International Journal of Green Energy* 17.13 (2020). Cited by: 26; All Open Access, Hybrid Gold Open Access, pp. 827–845. DOI: 10.1080/15435075.2020.1791878.

[12] Sangil Park and Jun-Ho Huh. “A Study on Big Data Collecting and Utilizing Smart Factory Based Grid Networking Big Data Using Apache Kafka”. In: *IEEE Access* 11 (2023). Cited by: 0; All Open Access, Gold Open Access, pp. 96131–96142. DOI: 10.1109/ACCESS.2023.3305586.

[13] Mandeep Kumar and Amritpal Singh. “Probabilistic data structures in smart city: Survey, applications, challenges, and research directions”. In: *Journal of Ambient Intelligence and Smart Environments* 14.4 (2022). Cited by: 4; All Open Access, Bronze Open Access, pp. 229–284. DOI: 10.3233/AIS-220101.

[14] Dabeeruddin Syed et al. “Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications”. In: *IEEE Access* 9 (2021). Cited by: 66; All Open Access, Gold Open Access, Green Open Access, pp. 59564–59585. DOI: 10.1109/ACCESS.2020.3041178.

[15] Ali Davoudian and Mengchi Liu. “Big Data Systems: A Software Engineering Perspective”. In: *ACM Computing Surveys* 53.5 (2020). Cited by: 33; All Open Access, Bronze Open Access. DOI: 10.1145/3408314.

[16] Grischa Beier, Julian Kiefer, and Jutta Knopf. “Potentials of big data for corporate environmental management: A case study from the German automotive industry”. In: *Journal of Industrial Ecology* 26.1 (2022). Cited by: 16; All Open Access, Hybrid Gold Open Access, pp. 336–349. DOI: 10.1111/jiec.13062.

[17] Rosa Virginia Encinas Quille et al. "Architecture of a Data Portal for Publishing and Delivering Open Data for Atmospheric Measurement". In: *International Journal of Environmental Research and Public Health* 20.7 (2023). Cited by: 1; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/ijerph20075374.

[18] Fabio Parisi, Maria Pia Fanti, and Agostino Marcello Mangini. "Information and communication technologies applied to intelligent buildings: A review". In: *Journal of Information Technology in Construction* 26 (2021). Cited by: 5; All Open Access, Gold Open Access, pp. 458–488. DOI: 10.36680/j.itcon.2021.025.

[19] Jenniffer Sidney Guerrero-Prado et al. "The power of big data and data analytics for AMI data: A case study". In: *Sensors (Switzerland)* 20.11 (2020). Cited by: 18; All Open Access, Gold Open Access, Green Open Access, pp. 1–27. DOI: 10.3390/s20113289.

[20] Martin MacAk, Mouzhi Ge, and Barbora Buhnova. "A Cross-Domain Comparative Study of Big Data Architectures". In: *International Journal of Cooperative Information Systems* 29.4 (2020). Cited by: 2; All Open Access, Hybrid Gold Open Access. DOI: 10.1142/S0218843020300016.

[21] Samuel Janík et al. "Effective Data Utilization in the Context of Industry 4.0 Technology Integration". In: *Applied Sciences (Switzerland)* 12.20 (2022). Cited by: 1; All Open Access, Gold Open Access. DOI: 10.3390/app122010517.

[22] Tomasz Dudek and Artur Kujawski. "The Concept of Big Data Management with Various Transportation Systems Sources as a Key Role in Smart Cities Development". In: *Energies* 15.24 (2022). Cited by: 1; All Open Access, Gold Open Access. DOI: 10.3390/en15249506.

[23] Mazen Juma, Fuad Alattar, and Basim Touqan. "Securing Big Data Integrity for Industrial IoT in Smart Manufacturing Based on the Trusted Consortium Blockchain (TCB)". In: *IoT* 4.1 (2023). Cited by: 5; All Open Access, Gold Open Access, pp. 27–55. DOI: 10.3390/iot4010002.

[24] Noor Ahmed Qarabsh, Sana Sabah Sabry, and Haneen Ahmed Qarabash. "Smart grid in the context of industry 4.0: An overview of communications technologies and challenges". In: *Indonesian Journal of Electrical Engineering and Computer Science* 18.2 (2020). Cited by: 29; All Open Access, Gold Open Access, Green Open Access, pp. 656–665. DOI: 10.11591/ijeecs.v18.i2.pp656-665.

[25] Ming Zeng et al. “Sustainable Insights for Energy Big Data Governance in China: Full Life Cycle Curation from the Ecosystem Perspective”. In: *Sustainability (Switzerland)* 14.10 (2022). Cited by: 1; All Open Access, Gold Open Access. DOI: 10.3390/su14106013.

[26] Roberto Tardio, Alejandro Mate, and Juan Trujillo. “An Iterative Methodology for Defining Big Data Analytics Architectures”. In: *IEEE Access* 8 (2020). Cited by: 8; All Open Access, Gold Open Access, Green Open Access, pp. 210597–210616. DOI: 10.1109/ACCESS.2020.3039455.

[27] Arezou Naghib et al. “A comprehensive and systematic literature review on the big data management techniques in the internet of things”. In: *Wireless Networks* 29.3 (2023). Cited by: 3; All Open Access, Bronze Open Access, pp. 1085–1144. DOI: 10.1007/s11276-022-03177-5.

[28] Philip Matesanz et al. “Demand-driven data acquisition for large scale fleets”. In: *Sensors* 21.21 (2021). Cited by: 1; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/s21217190.

[29] Chunpeng Wang et al. “Deploying the Big Data Science Center at the Shanghai Synchrotron Radiation Facility: The first superfacility platform in China”. In: *Machine Learning: Science and Technology* 2.3 (2021). Cited by: 8; All Open Access, Gold Open Access. DOI: 10.1088/2632-2153/abe193.

[30] Jenniffer S. Guerrero-Prado, Wilfredo Alfonso-Morales, and Eduardo F. Caicedo-Bravo. “A data analytics/big data framework for advanced metering infrastructure data”. In: *Sensors* 21.16 (2021). Cited by: 10; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/s21165650.

[31] Saman Fattahi, Sharifu Ura, and Muhammad Noor-E-alam. “Decision-Making Using Big Data Relevant to Sustainable Development Goals (SDGs)”. In: *Big Data and Cognitive Computing* 6.2 (2022). Cited by: 3; All Open Access, Gold Open Access. DOI: 10.3390/bdcc6020064.

[32] Ashwini Konanahalli, Marina Marinelli, and Lukumon Oyedele. “Drivers and Challenges Associated With the Implementation of Big Data Within U.K. Facilities Management Sector: An Exploratory Factor Analysis Approach”. In: *IEEE Transactions on Engineering Management* 69.4 (2022). Cited by: 12; All Open Access, Green Open Access, pp. 916–929. DOI: 10.1109/TEM.2019.2959914.

[33] Hyeontaek Oh et al. "Competitive Data Trading Model With Privacy Valuation for Multiple Stakeholders in IoT Data Markets". In: *IEEE Internet of Things Journal* 7.4 (2020). Cited by: 39; All Open Access, Green Open Access, pp. 3623–3639. DOI: 10.1109/JIOT.2020.2973662.

[34] Valentina Janev et al. "Responsible Knowledge Management in Energy Data Ecosystems". In: *Energies* 15.11 (2022). Cited by: 5; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3390/en15113973.

[35] Luis Mendoza-Pitti et al. "Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review". In: *IEEE Access* 9 (2021). Cited by: 10; All Open Access, Gold Open Access, Green Open Access, pp. 26119–26137. DOI: 10.1109/ACCESS.2021.3057543.

[36] Naureen Naqvi, Sabih ur Rehman, and Md Zahidul Islam. "A Hyperconnected Smart City Framework: Digital Resources Using Enhanced Pedagogical Techniques". In: *Australasian Journal of Information Systems* 19.1 (2020). Cited by: 6; All Open Access, Gold Open Access, pp. 1–42. DOI: 10.3127/ajis.v24i0.2531.