Data Interoperability

edited by

Andreas Winter
Carl von Ossietzky University, Oldenburg

Christian Schönberg
Carl von Ossietzky University, Oldenburg

Oldenburg Lecture Notes on Software Engineering

OLNSE Number 8/2024
September 2024

Data Interoperability

edited by

Andreas Winter
Carl von Ossietzky University, Oldenburg

Christian Schönberg
Carl von Ossietzky University, Oldenburg

Oldenburg Lecture Notes on Software Engineering

OLNSE Number 8/2024
September 2024

Oldenburg Lecture Notes
on Software Engineering (OLNSE)

Carl von Ossietzky University Oldenburg
Department for Computer Science
Software Engineering Group
26111 Oldenburg, Germany

Content

DATA INTEROPERABILITY

Im Interreg North Sea Projekt „Data for All" (D4A)[1] werden innovative Daten-getriebene Ansätze zur Schaffung und Verbesserung von digitalen Diensten im Bereich von Smart Cities und Smart Regions erforscht und praktisch erprobt. Smarte Systeme – Smart Cities, Smart Health Applications, Smart Agriculture Systems, usw. – produzieren große Mengen an Daten. Diese Daten sind oft sehr unterschiedlich aufgebaut, selbst wenn sie ähnliche Informationen enthalten. Sie folgen oft keinem standardisierten Schema, oder interpretieren Standards auf unterschiedliche Weise. Sie verwenden unterschiedliche Formate, Formatierungen, Kodierungen, Maßeinheiten, Intervalle, Konventionen und Annahmen.

Beispielsweise kann ein Anbieter von E-Scootern seine Daten über die Gefährte nach deren Aufenthaltsort strukturieren, während ein anderer Anbieter die Daten nach dem genauen Typs des Scooters strukturiert. Dadurch ist eine gemeinsame Nutzung beider Datensätze deutlich erschwert. Aber nicht nur strukturell unterscheiden sich Daten Smarter Systeme, auch im Detail gibt es große Unterschiede: So speichert ein Scooter-Anbieter alle zwei Minuten den Aufenthaltsort der Geräte, ein anderer nur alle drei Minuten. Will man die Zeitreihen zusammenlegen, kommt es ggf. zu Inkonsistenzen. Auch unterschiedliche Maßeinheiten (mm/cm/in, usw.) oder andere technische Formate (JSON/XML/SQL/NoSQL) erschweren die gemeinsame interoperable Verwendung der Daten.

Im folgenden Paper wird zunächst der Begriff Dateninteroperabilität definiert. Im Anschluss werden verschiedene Herausforderungen der Dateninteroperabilität im rechtlichen, ethischen, organisatorischen und technischen Kontext herausgearbeitet und erklärt.

In the Interreg North Sea project "Data for All" (D4A)[1], innovative data-driven approaches to create and improve digital services in the field of Smart Cities and Smart Regions are researched and practically tested. Smart systems – smart cities, smart health applications, smart agriculture systems, etc. – produce large amounts of data. These data are often structured very differently, even if they contain similar information. They often do not follow a standardized schema, or they interpret standards in different ways. They use different formats, formatting, coding, units of measurement, intervals, conventions, and assumptions.

For example, one e-scooter provider may structure its vehicle data according to their location, while another provider may structure the data according to the exact type of scooter. This makes it much more difficult to use both sets of data together. But it is not only structurally that data of smart systems differ, there are also major differences in detail: for example, one scooter provider stores the whereabouts of the vehicles every two minutes, another only every three minutes. If one wants to combine the time series, inconsistencies may occur. Different units of measurement (mm/cm/in, etc.) or different technical formats (JSON/XML/SQL/NoSQL) also make it difficult to use the data in an interoperable manner.

The following paper first defines the term data interoperability. It then proceeds to explain challenges of data interoperability in legal, ethical, organizational, and technical contexts.

[1] https://www.interregnorthsea.eu/dataforall

Data Interoperability

Azadeh Jalilian, Christian Schönberg, Andreas Winter
Carl von Ossietzky Universität Oldenburg
Software Engineering Group
Oldenburg, Germany
(azadeh.jalilian) | (christian.schoenberg) | (andreas.winter) @uol.de

***Abstract*—This paper explores the impact and challenges of data interoperability in smart cities and regions, with a focus on the "Data for All" (D4A) project. It examines how advanced digitalization can optimize urban efficiency and quality of life by integrating diverse systems and services. Key dimensions of interoperability addressed include organizational aspects such as stakeholder coordination and governance, regulatory and ethical issues like data privacy and ownership, and technical difficulties involving disparate data formats and schemas. The paper evaluates current solutions like metadata standards and data contracts and briefly discusses strategies to address their challenges. It emphasizes the essential role of effective data integration in supporting sustainable urban development and improving public services. Additionally, the research aims to identify new, more efficient mechanisms to enhance data interoperability, focusing on the three key dimensions of D4A.**

1. Introduction

Smart cities and smart regions are transforming urban and rural areas using advanced digitalization technologies and data analytics to improve quality of life and increase efficiency. They employ information and communication technology (ICT) to create sustainable environments, enhance community operations, and foster economic growth through data driven decision making [1]. In today's data-driven landscape, the ability to seamlessly exchange, integrate, and utilize data from various sources and domains has become a critical enabler for innovation, collaboration, and effective decision-making across diverse sectors. Data interoperability in smart cities is crucial, as it ensures that diverse systems and services can work together efficiently to enhance urban living.

In 2024 the soccer team Holstein Kiel qualified for the Bundesliga. Holstein-Stadion in the port city of Kiel in northern Germany is the home of Holstein Kiel. The upcoming Bundesliga matches require a new mobility concept for football fans in Kiel, which should be based on smart region concepts. When planning a journey to stadium, travelers often have options like taking a bicycle or a boat or car. A map of mobility in Kiel is shown in [2]. Traditionally, one might choose a single mode of transportation for the entire trip. However, using data from various sources (such as boat timetables and bicycle stations) and making it interoperable (so that data from both sources can be used together), it is possible to identify a more efficient route that combines different modes of transportation. For example, one could travel from point A to point B via ferry and use a bicycle from point B to the stadium. This combined approach can offer a more convenient, flexible and potentially faster journey compared to using just one mode of transportation. This small example, taken from the German D4A pilot, is used throughout the paper to illustrate interoperability challenges and solutions and is referred to as the "football example."

Data interoperability refers to the ability of different systems, platforms, organizations, and stakeholders to seamlessly exchange and use data in a standardized and efficient manner [3]. The integration of heterogeneous data from different sources requires a common interoperability objective to be achieved, which restricts used data and their capabilities. Data interoperability is closely intertwined with the broader concept of system interoperability, which focuses on the features and infrastructures that facilitate the development, establishment, advertisement, distribution, and collaborative use of interoperability-enabling concepts and technologies [4].

This holistic approach recognizes that achieving true data interoperability requires technical solutions and organizational and governance frame works that promote collaboration, oversight, and accountability in data management [5]. It enables seamless integration and interaction between different data sources, making accessing and utilizing diverse data sets possible. The importance of data interoperability has been widely recognized and emphasized in various smart city and region domains, including healthcare [6], transportation [7], and agriculture [8], but it must also be considered across domains.

The interoperability objective to improve fan mobility for football matches requires the effective use of data integration. By analyzing mobility data, the movement of fans heading to the stadium can be anticipated. Integrating event schedules with transportation and parking information allows bus drivers to modify routes to avoid congested areas. Additionally, car users can park near the stadium and switch to bicycles, scooters, or buses for the remainder of their trip. A coordinated strategy helps alleviate traffic congestion during major events. The lack of data interoperability often impedes the effective utilization of data. In the football example, this issue can lead to significant traffic congestion during football matches.

Interoperability in the "Data for All" (D4A) project will be viewed at three levels, including organizational, regulatory/ethical, and technical dimensions.

Organizational interoperability addresses the policy, governance, regulatory, and cultural aspects that enable effective data sharing and collaboration between various organizations [9]. Addressing the regulatory/ethical dimension of interoperability ensures a concerted effort to create frameworks that can accommodate diverse ethical standards while this involves ongoing dialogue and collaboration among stakeholders to develop adaptable and robust ethical guidelines that can manage the complexities of interoperable systems [10]. At the technical level, interoperability ensures that systems can communicate and exchange data using common protocols and procedures. Addressing data interoperability challenges requires a multifaceted approach encompassing the development of common data, models, and exchange protocols [5].

While [11] views regulatory issues as part of organizational interoperability, D4A separates legal and ethical issues from organizational Interoperability. Like technical interoperability, legal and ethical interoperability restrict and realize features required by organizational interoperability.

This paper will explore the impact of data interoperability in smart regions, focusing on the D4A project. It will examine the challenges across the organizational, regulatory/ethical, and technical dimensions. It will also review current solutions and their limitations to address these challenges.

2. Challenges

In the three dimensions mentioned above, it is essential to consider several roles are necessary for understanding the challenges that follow:

Data creator/editor: data creators and editors are responsible for generating and maintaining high-quality data. Their main tasks include creating and inputting accurate data, either manually or through automated methods, and ensuring adherence to quality standards [12].

Data owner: Bhansali [13] emphasizes that a data owner as an individual or an entity is responsible for the management and oversight of data assets. This role involves ensuring data quality, security, compliance, and accessibility, determining access permissions, and enforcing data-related policies within an organization.

Data provider: In smart regions, data providers hold critical responsibility for the delivery and maintenance of data, ensuring that it meets the needs of various stakeholders, as highlighted by Jobst et al. [14]. They must address complex issues like data privacy and security to ensure sensitive information remains protected while facilitating effective data sharing and collaboration. That is a delicate balance, navigating these technical and ethical considerations while maintaining data integrity and accuracy. Meeting these challenges demands continuous teamwork, innovation, and investment. Furthermore, data providers are tasked with preserving data integrity as it traverses through different systems [15], [16].

While data owners focus on the governance and policy aspects of data management, data providers handle the practical aspects of data delivery and maintenance, including addressing privacy, security, and integrity issues. The data provider is authorized to enter into data usage contracts, contingent upon holding the requisite level of authority within the organization. It is essential to ensure strict adherence to all relevant legal and regulatory requirements.

In the football example, the roles are as follows: Kiel-Region/Addix are viewed as data provider. Scooter and bike rental companies, bus and ferry providers are data creators. The city of Kiel acts as data owner which also includes ownership of the data provided by its contractors.

2.1. Organizational Challenges

Organizational aspects of interoperability involve defining business goals, aligning and coordinating business processes, and enhancing collaboration among organizations with diverse internal structures and procedures that aim to exchange information. The primary objective of organizational interoperability is to meet user needs by ensuring services are available, easily identifiable, accessible, and user-centric. Essentially, it enables business organizations to offer services to each other and users, customers, or the broader public [17].

2.1.1. Coordination and Collaboration. Achieving interoperability within and between organizations requires effective coordination and collaboration. Diverse goals, priorities, and workflows can complicate alignment, while institutional inertia and cultural differences may cause resistance and conflicts. A critical aspect of this process is the role of data owners, who ensure that data is accurate, accessible, and used appropriately across different systems and organizations. Aligning stakeholders across government and private sector is essential but often contentious and time consuming. Successful interoperability depends on effective coordination, clear data management practices, and fostering a collaborative culture [18].

To effectively host a football match, it is crucial for various organizations such as football event organizers, transportation providers (including buses, boats, and bicycles), and parking facilities to collaborate and share information. This coordinated effort helps alleviate traffic congestion and reduce overcrowding in parking areas, ultimately enhancing the overall experience for fans.

2.1.2. Governance and Management. Establishing clear roles and responsibilities is crucial for defining accountability within a project. Challenges that commonly arise include ensuring adherence to data protection regulations, and integrating data from disparate systems. These issues can lead to inconsistencies in data, vulnerabilities in security, and operational inefficiencies, which ultimately impact decision-making and performance [18].

For instance, consider the scenario of a football match where transportation services such as buses, ferries, bicycles, and scooters are provided by different entities using separate systems with incompatible data formats. Integrating these systems poses significant difficulties. Data inconsistencies may result in inaccurate real-time updates about transportation options or schedules for fans, leading to confusion and negatively affecting their overall match experience.

2.2. Regulatory/ Ethical Challenges

Regulatory/Ethical considerations in data interoperability involve issues such as privacy, consent, and data ownership. Ensuring compliance with regulations (e.g., General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) while enabling data exchange requires robust mechanisms for identifying, encrypting, and securing data handling practices [19]. Lee A. Bygrave [20] critiques the EU's current data protection framework, particularly its limitations regarding the ethical management of data interoperability. Bygrave argues that the EU GDPR, while robust in safeguarding personal data, does not adequately address the complexities and ethical challenges posed by data interoperability in algorithmic systems.

2.2.1. Data Ownership. In smart regions, data ownership involves managing and controlling data from transportation modes like buses, ships, and e-scooters, with challenges including privacy, security, equitable access, and clear legal rights. For example, improper management of e-scooter data could lead to unauthorized access or misuse of user information [21]. To address these issues, it is essential to protect user privacy and ensure that data is used responsibly. When analyzing data to place e-scooters in high-demand areas, consideration should be given to maintaining user privacy and maintaining fair access to data [22].

Another challenge is determining data ownership after data has been combined. In the football example, ownership is straightforward since all the data belongs to the Kiel region. However, in other situations, ownership is not as clear.

2.2.2. The purpose of data. In [23] Weinhardt identifies significant ethical challenges related to combining data sets, such as privacy violations and concerns over informed consent. The paper emphasizes that using data collected for one purpose in new contexts can infringe on individuals' rights and raises critical issues about data ownership and the ethical use of information.

The municipality is granted access to the e-scooter location data solely to identify potential new station sites. The use of this data for any other activities or objectives is strictly prohibited. Additionally, the application of data analysis techniques that could expose or reconstruct individual user routes is not allowed to safeguard user privacy and security.

2.3. Technical Challenges

Achieving data interoperability at a technical level involves overcoming various obstacles such as disparate data formats, incompatible schemas, and differing data standards across systems. Integration efforts often face difficulties in mapping and transforming data from one format to another seamlessly. Additionally, issues like consistency, and versioning further complicate the interoperability process [24]. Data providers ensure that the data they supply is accessible, consistent, and compatible with the systems that consume the data.

The focus is on addressing technical challenges related to data interoperability. The aim is to enhance seamless data exchange across various platforms and systems, ensuring efficiency and reliability in information sharing.

2.3.1. Data storage. When it comes to data storage, the data necessitate scalable solutions beyond traditional relational databases. The discrepancy between data sets, where one utilizes the XML format and the other employs the JSON format, results in significant complications during data aggregation. SQL databases, known for their structured data models and strong consistency, face challenges with scalability traditionally scale vertically, and are best suited for applications requiring complex queries and transactions. NoSQL databases like MongoDB and Cassandra offer the required flexibility and excel in horizontal scaling but introduce complexities in terms of data consistency and query capabilities [25].

2.3.2. Data Model. Data models employed by different databases, such as relational, hierarchical, and network models are very diverse. This leads to significant conflicts in how data is represented and manipulated [26]. Suppose a bus provider might save their data in a relational database using tables to represent the data, while a ferry provider could use a hierarchical database to organize their data in a tree-like structure. These fundamental differences complicate the integration process, requiring sophisticated methods to translate and reconcile the disparate data models.

2.3.3. Query Language. The challenge of query language integration arises from the diverse query languages used across different databases, such as SQL, and GraphQL [27] each with its unique syntax and semantics. This variation complicates the formulation of unified queries executable across multiple data bases. Effective integration thus demands the development of query translation mechanisms capable of converting query between languages without losing their intended meaning [28]. Additionally, distributed query processing involves advanced techniques to minimize data movement and optimize resource utilization. Query rewriting, which entails translating user queries into source specific queries that adhere to each source's semantics and constraints, is a complex task requiring sophisticated algorithms [29].

2.3.4. Schema. In databases and data management, a schema defines the organization of data and the relationships between different data elements [30]. In [31], one database might store customer information in a single table, while another might distribute this information across multiple tables.

XML databases store and manage data in XML format, which is hierarchical and supports nested structures. NoSQL databases encompass various types, including key-value stores, document stores, column-family stores, and graph databases [28]. Integrating schema requires identifying correspondences between schema elements and resolving conflicts, such as differing data types and naming conventions.

2.3.5. Data Quality. Data quality refers specifically to the quality of data values or instances. Data quality issues are often termed errors, anomalies, or data "dirtiness" and can include problems such as missing attribute values, incorrect attribute values, or inconsistent representations of the same data. It's not unusual for operational databases to contain 60% to 90% inaccurate or poor-quality data [32].

Data integration can either enhance quality by consolidating sources or degrade it through inconsistencies and inaccuracies, depending on how well the process is managed.Ensuring high data quality in data integration is the responsibility of the data provider. This involves data cleaning to remove noise, errors, and inconsistencies, which are common in datasets. High-quality data is essential for reliable integration, as inaccuracies can lead to misleading results when integrating diverse data sources [33].

2.3.6. Security and Authorization. Based on [34], security and authorization in data interoperability present both technical and organizational challenges. Technically, it involves encryption, authentication, and data integrity, while organizationally, it requires policy development, regulatory compliance, and coordinated efforts across different entities. According to [26] integrating security policies from multiple autonomous data bases is complex due to differing authentication, authorization, and encryption mechanisms.

3. Solutions

Several critical issues regarding data interoperability within smart ecosystems have been identified. Interoperability among diverse systems and devices often faces significant hurdles due to varying protocols, standards, and data formats. This lack of uniformity hinders seamless communication and integration, complicating the development and maintenance of cohesive smart ecosystems. Additionally, the scale and complexity of data generated and exchanged within these systems further exacerbate interoperability challenges, necessitating sophisticated solutions to ensure efficient data management and utilization [35]. To address these challenges, a range of current solutions have been proposed and implemented.

3.1. Metadata

Metadata is data that provides information about other data, helping to organize, find, and understand it by detailing its characteristics like origin, format, and relationships [36]. Metadata is central to any information organization and management function, serving as a tool to search, navigate, and explore information. It enriches, links, opens, and filters resources, driving their visibility, discoverability, access, and use [37]. According to [36] the problem of metadata often includes inconsistency across different systems, leading to difficulties in data integration and interoperability. Additionally, incomplete or inaccurate metadata can hinder effective data discovery and management, while managing metadata can be resource-intensive and complex.

3.2. Standards

In [38], a significant impediment to seamless data interoperability is the lack of standardization in data formats, communication protocols, and metadata representation. The absence of universally accepted standards hinders the smooth exchange and integration of data. Encouraging the adoption of existing standards is crucial but challenging, as it involves persuading diverse stakeholders to conform to uniform practices and frameworks. Additionally, many standards are often vague, overly general, or inflexible, leading to varying interpretations and implementations.

3.3. Contracting and Licensing

Data contracts define data structure and usage, facilitate connections between data producers and consumers, and are central to decentralized data mesh architectures. Some tools, like "Data Contract Studio" help manage data contracts by organizing sections on servers, terms, conditions, data models, and data quality rules. Compliance with data protection laws (e.g., GDPR, HIPAA) is essential, and data contracts are also used in APIs, web services, and block chain applications [39].

3.4. Data Warehouse

Gardner in [40], explained that a data warehouse is a centralized system that stores, manages, and analyzes large volumes of data from various sources, integrating information into a unified view to support business intelligence activities like querying and reporting. Optimized for read-heavy operations, it uses data extraction, transformation, and loading (ETL) processes to ensure data consistency and quality, facilitating complex queries and data mining. However, data warehouses can be costly and complex to manage, with challenges in integrating diverse data sources, maintaining performance, and ensuring data freshness as volumes increase.

4. Conclusion

Data interoperability in smart cities and regions is crucial for improving efficiency, sustainability, and quality of life. Integrating information and communication technologies (ICT) with advanced data analytics can significantly enhance urban services and transportation systems. Ensuring seamless data exchange is vital for fostering innovation, making informed decisions, and enabling effective collaboration among diverse stakeholders. However, achieving interoperability poses numerous challenges, including organizational, regulatory, ethical, and technical barriers. To overcome these obstacles, aligning processes, protecting data privacy and security, and managing data diversity and quality is essential. Several strategies are being employed to tackle these issues, such as using metadata, implementing standardized protocols, and establishing secure data contracts. Continued efforts in governance development, stakeholder collaboration, and technological investment are crucial. The "Data for All" (D4A) project highlights the importance of these efforts. By addressing interoperability challenges, smart cities can unlock their full potential, promoting sustainable growth and improving living standards for their residents.

References

[1] J. S. Gracias, G. S. Parnell, E. Specking, E. A. Pohl, and R. Buchanan, “Smart cities-a structured literature review,” *Smart Cities*, vol. 6, no. 4, pp. 1719–1743, 2023.

[2] “Datenplattform: Mobility live access.” https://www.kielregion.de/mobilitaetsregion/digitale-mobilitaet/datenplattform/.

[3] D. Nüst, C. Granell, B. Hofer, M. Konkol, F. O. Ostermann, R. Sileryte, and V. Cerutti, “Reproducible research and giscience: an evaluation using agile conference papers,” *PeerJ*, vol. 6, p. e5072, 2018.

[4] L. Y. Pang, R. Y. Zhong, J. Fang, and G. Q. Huang, “Data-source interoperability service for heterogeneous information integration in ubiquitous enterprises,” *Advanced Engineering Informatics*, vol. 29, no. 3, pp. 549–561, 2015.

[5] R. D. Kush, D. Warzel, M. A. Kush, A. Sherman, E. A. Navarro, R. Fitzmartin, F. Pétavy, J. Galvez, L. B. Becnel, F. Zhou, *et al.*, “Fair data sharing: the roles of common data elements and harmonization,” *Journal of biomedical informatics*, vol. 107, p. 103421, 2020.

[6] R. L. Richesson and J. Krischer, “Data standards in clinical research: gaps, overlaps, challenges and future directions,” *Journal of the American Medical Informatics Association*, vol. 14, no. 6, pp. 687–696, 2007.

[7] N. Čuš-Babič, S. F. Guerra De Oliveira, and A. Tibaut, “Interoperability of infrastructure and transportation information models: A public transport case study,” *Applied Sciences*, vol. 12, no. 12, p. 6234, 2022.

[8] I. Roussaki, K. Doolin, A. Skarmeta, G. Routis, J. A. Lopez-Morales, E. Claffey, M. Mora, and J. A. Martinez, “Building an interoperable space for smart agriculture,” *Digital Communications and Networks*, vol. 9, no. 1, pp. 183–193, 2023.

[9] Z. Boukhers, C. Lange, and O. Beyan, “Enhancing data space semantic interoperability through machine learning: a visionary perspective,” in *Companion Proceedings of the ACM Web Conference 2023*, pp. 1462–1467, 2023.

[10] D. Danks and D. Trusilo, “The challenge of ethical interoperability,” *Digital Society*, vol. 1, no. 2, p. 11, 2022.

[11] H. Van Der Veer and A. Wiles, “Achieving technical interoperability,” *European telecommunications standards institute*, 2008.

[12] “Datenplattform: Defining principles, elements, and roles and responsibilities in a data governance policy.” https://www.dataversity.net/.

[13] N. Bhansali, *Data governance: Creating value from information assets*. CRC Press, 2013.

[14] M. Jobst and T. Fischer, “Approaching a common conscious dataspace from a data provider perspective–requirements and perspectives,” in *International Conference on Computational Science and Its Applications*, pp. 333–343, Springer, 2022.

[15] “Datenplattform: Data to the people.” https://www.datagalaxy.com/en/.

[16] “Datenplattform:the importance of data ownership in data governance.” https://dainstudios.com/.

[17] F. B. Vernadat, “Technical, semantic and organizational issues of enterprise interoperability and networking,” *Annual Reviews in Control*, vol. 34, no. 1, pp. 139–144, 2010.

[18] H. Kubicek and R. Cimander, “Three dimensions of organizational interoperability,” *European Journal of ePractice*, vol. 6, pp. 1–12, 2009.

[19] A. Oakley, “Hipaa, hippa, or hippo: What really is the heath insurance portability and accountability act?,” *Biotechnology Law Report*, vol. 42, no. 6, pp. 306–318, 2023.

[20] L. A. Bygrave, “Eu data protection law falls short as desirable model for algorithmic regulation,” *Algorithmic Regulation*, p. 31, 2017.

[21] S. Severengiz, S. Finke, N. Schelte, and H. Forrister, “Assessing the environmental impact of novel mobility services using shared electric scooters as an example,” *Procedia Manufacturing*, vol. 43, pp. 80–87, 2020.

[22] A. Ciociola, M. Cocca, D. Giordano, L. Vassio, and M. Mellia, “E-scooter sharing: Leveraging open data for system design,” in *2020 IEEE/ACM 24th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, pp. 1–8, IEEE, 2020.

[23] M. Weinhardt, “Ethical issues in the use of big data for social research,” *Historical Social Research/Historische Sozialforschung*, vol. 45, no. 3, pp. 342–368, 2020.

[24] W. Tan, X. Shi, X. Li, and et al., “Data interoperability

in heterogeneous big data analytics," *Journal of Parallel and Distributed Computing*, vol. 138, pp. 38–49, 2020.

[25] R. Cattell, "Scalable sql and nosql data stores," *Acm Sigmod Record*, vol. 39, no. 4, pp. 12–27, 2011.

[26] A. P. Sheth and J. A. Larson, "Federated database systems for managing distributed, heterogeneous, and autonomous databases," *ACM Computing Surveys (CSUR)*, vol. 22, no. 3, pp. 183–236, 1990.

[27] D. Kastowo, E. Utami, and A. H. Muhammad, "Fhir, bigchaindb, and graphql approach for interoperability between heterogeneous health information system," in *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, pp. 272–277, IEEE, 2022.

[28] S. K. Gajendran, "A survey on nosql databases," *University of Illinois*, 2012.

[29] A. Halevy, A. Rajaraman, and J. Ordille, "Data integration: The teenage years," in *Proceedings of the 32nd international conference on Very large data bases*, pp. 9–16, 2006.

[30] H. Kondylakis, G. Flouris, and D. Plexousakis, "Ontology and schema evolution in data integration: review and assessment," in *On the Move to Meaningful Internet Systems: OTM 2009: Confederated International Conferences, CoopIS, DOA, IS, and ODBASE 2009, Vilamoura, Portugal, November 1-6, 2009, Proceedings, Part II*, pp. 932–947, Springer, 2009.

[31] A. Doan, P. Domingos, and A. Y. Halevy, "Reconciling schemas of disparate data sources: A machine-learning approach," in *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, pp. 509–520, 2001.

[32] P. Oliveira, F. Rodrigues, and P. R. Henriques, "A formal definition of data quality problems.," in *ICIQ*, 2005.

[33] M. Lenzerini, "Data integration: A theoretical perspective," in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 233–246, 2002.

[34] M. Elkhodr, S. Shahrestani, and H. Cheung, "The internet of things: New interoperability, management and security challenges," *arXiv preprint arXiv:1604.04824*, 2016.

[35] N. N. Hurrah, E. Khan, and S. A. Parah, "Smart ecosystems for sustainable development: Opportunities, challenges, and solutions," *Intelligent Multimedia Signal Processing for Smart Ecosystems*, pp. 3–28, 2023.

[36] J. Riley, "Understanding metadata," *Washington DC, United States: National Information Standards Organization (http://www. niso. org/publications/press/UnderstandingMetadata. pdf)*, vol. 23, pp. 7–10, 2017.

[37] B. Haslhofer and W. Klas, "A survey of techniques for achieving metadata interoperability," *ACM Computing Surveys (CSUR)*, vol. 42, no. 2, pp. 1–37, 2010.

[38] M. S. Gal and D. L. Rubinfeld, "Data standardization," *NYUL Rev.*, vol. 94, p. 737, 2019.

[39] S. J. Preis, "Synergizing data contracts and bpm to improve process interoperability: An analysis of gaps and opportunities of data exchange agreements in bpm models,"

[40] S. R. Gardner, "Building the data warehouse," *Communications of the ACM*, vol. 41, no. 9, pp. 52–60, 1998.