UNIVERSIDAD
NACIONAL
DE COLOMBIA

Multimodal Data Exchange Systems for Enhanced Interaction and Management

Ph.D. Thesis Proposal

Yeison Nolberto Cardona-Álvarez

Advisor:
Prof. Andrés Marino Álvarez-Meza, Ph.D.

Co-Advisor:
Prof. César Germán Castellanos-Dominguez, Ph.D.

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y de Computación
Doctorado en Ingeniería - Automática
Manizales, Colombia
2023

1 Abstract

This doctoral thesis proposal outlines a framework designed to enhance multimodal data exchange systems, emphasizing incremental improvements in scalability, efficiency, and data type management. Beginning with an evaluative review of extant data exchange technologies, the study identifies areas ripe for development. The proposed architecture aims to refine Application Programming Interfaces (APIs), strengthen messaging systems, and streamline Extract, Transform, Load (ETL) processes within a secure context.

The necessity for advanced data exchange mechanisms stems from the growing complexity and volume of digital information, particularly in healthcare and research settings. Efficient and secure data exchange is critical for timely decision-making and effective management of resources. This thesis, therefore, aims to address these challenges by proposing a robust and scalable data exchange framework.

Central to this proposal is the problem of enhancing data exchange in high-stakes environments, such as Neonatal Intensive Care Units (NICUs) and large-scale, unstructured data projects. The proposed framework is designed to improve data accessibility and integration, enhancing decision-making processes and operational efficiency in these critical areas.

The research objectives are set forth with a clear vision, intending to make a well-defined contribution to the analysis of Electroencephalography (EEG) data and the management of NICU data. A pragmatic, iterative approach to development, underscored by thorough testing, forms the core of the proposed methodology.

Expectations for the research outcomes include the development of a refined data exchange model applicable in the NICU of 'Hospital Universitario de Caldas - SES HUC' and the organization of unstructured data in a specific project 'Prototipo funcional de Lengua electrónica para identificación de sabores en cacao fino de origen colombiano'. These outcomes will be achieved while acknowledging the inherent limitations and challenges of the environments, including data privacy concerns and the need for high reliability in critical care settings.

Integral to this proposal is Table 1, which encapsulates the roles and challenges pertinent to various data exchange methods. This table provides a structured comparison of APIs, messaging systems, ETL processes, and file transfer protocols. It outlines their essential functions in the context of data exchange—such as facilitating interoperability, managing real-time processing, and ensuring efficient data integration—and acknowledges the associated challenges, including standardization, scalability, and performance. This table (see Table 1: Roles and Challenges in Data Exchange Methods) is crucial for understanding the current landscape and the potential impact of the pro-

posed enhancements.

The schedule for development is meticulously planned to ensure a structured progression.

Method	**Role in Data Exchange**	**Challenges**
APIs	Facilitate interoperability and data format conversion between different platforms and environments.	Standardization, developer-centric security, scalability, heterogeneous API management, and Artificial Intelligence (AI) system reliability.
Messaging Systems	Asynchronous communication for distributed components, managing time-series data and real-time processing.	Performance overhead, Peer to Peer (P2P) scalability improvements, real-time data logging, and high-volume real-time traffic management.
ETL Processes	Batch processing and warehousing with efficient end-to-end integration for temporal datasets.	Integration with big data, fault tolerance, sharding for scalability, and consensus protocols for security.
File Transfer	Bulk movement of files with support for various protocols, ensuring efficient transfers across systems.	Bandwidth control, protocol optimization, industry 4.0 support, and digitalization in Small Medium Enterprises (SMEs).

Table 1 Roles and Challenges in Data Exchange Methods

2 Motivation

The exchange of information, commonly termed Data Exchange, is a backbone of the infrastructure of any research project that requires collaboration between multiple systems and platforms [1]. Since research often involves the use of a diversity of information sources and computational methods, multimodality becomes a key aspect [2, 3, 4, 5, 6]. This scenario requires not only the ability to transmit information across different protocols and formats but also the ability to efficiently handle large volumes of information in real time [7, 8, 9]. A stable and well-designed architecture is essential in this context [10, 11]. It allows not only smooth interaction between different platforms but also greater scalability and adaptability to future needs [12]. Given that the research field is inherently dynamic, having a scalable architecture can be the difference between a project that can evolve and adapt to new research questions and one that is constrained by its technical limitations [13, 14].

Data Exchange, a critical component within the framework of contemporary information systems architecture, significantly influences a spectrum of fields from systems engineering and computer science to biomedicine, meteorology, and social sciences

[15]. This multidisciplinary and multimodal facet of Data Exchange highlights its essential role in fostering innovation and facilitating collaboration in research and development [16]. As technological landscapes have evolved, so have the capabilities of Data Exchange systems, transcending historical barriers like information silos and system incompatibility [17]. Modern systems are equipped with database-backed architectures, end-to-end integration, and proficient handling of time-series information—critical for real-time processing of voluminous datasets [18]. The event-driven architecture and multi-protocol support are quintessential for the multimodal nature of current information streams, ensuring seamless navigation across varied information types and communication protocols [19] Essential features such as secure information encryption, scalable infrastructure, fault-tolerant designs, and low-latency operations, along with rigorous monitoring and auditing, are not just features but are the bedrock that bolsters the robustness and efficacy of Data Exchange systems [20]. These advancements have enabled concurrent operations and cross-platform compatibility, meeting the high demands of global research and industry applications [21].

Despite the advancements in Data Exchange, significant technological challenges and knowledge gaps persist that impede optimal system performance and hinder the realization of its full potential [22]. Issues such as ensuring integrity during high-velocity exchanges, achieving interoperability among disparate systems, and maintaining security and privacy in an era of increasing cyber threats remain formidable [23]. Moreover, the surge in the volume, velocity, and variety of Big Data introduces complexities in management and necessitates the development of more sophisticated methods for efficient processing and analysis [24]. While existing systems offer a degree of fault tolerance and scalability, the evolving landscape of digital technologies demands continuous research to bridge these gaps, particularly in areas such as real-time analytics, distributed computing, and adaptive security protocols [25]. The need to streamline exchange processes for multimodal datasets across various domains underscores the urgency to address these technological constraints and to extend the knowledge frontier in this critical field.

The ramifications of enhancing Data Exchange systems extend far beyond the technical sphere, directly impacting society and industry at large [26]. In healthcare, for instance, streamlined information sharing can lead to better patient outcomes through the facilitation of real-time medical data analysis and communication [27]. In the financial sector, robust mechanisms for data transfer are crucial for real-time transaction processing, affecting global markets and economies [28]. Furthermore, in the realm of environmental sciences, accurate and timely information dissemination is pivotal for climate modeling and disaster response, potentially saving lives and resources [29]. The industry implications are equally profound, where manufacturing, logistics, and supply chain management rely heavily on the efficient transfer of information to optimize production and distribution [30]. Advanced systems for data management and exchange thus act as enablers of smart cities and Industry 4.0, fostering innovation and driving economic growth [31]. The societal and industrial transformation predicated on effective data utilization underscores the profound importance of addressing the current challenges and pushing the boundaries of existing technologies.

As we look toward the future, the landscape of Data Exchange is poised for transformative innovations, driven by emerging technologies and evolving market needs [32]. Blockchain technology, with its decentralized and secure nature, is expected to revolutionize how information is exchanged, particularly in ensuring integrity and traceability [33]. The increasing adoption of the Internet Of Thingss (IoTs) promises a surge in real-time, sensor-generated datasets, necessitating more efficient and scalable Data Exchange solutions [8]. AI and Machine Learning (ML) are set to play a pivotal role in automating and optimizing information processing and analytics, enabling more intelligent and predictive Data Exchange systems [34]. Edge computing is emerging as a crucial technology for managing the deluge of digital content at the source, reducing latency and bandwidth use in information transfer [35]. Additionally, the integration of quantum computing could lead to breakthroughs in information encryption and processing speeds, redefining the paradigms of secure and efficient Data Exchange [36]. These trends and innovations are not just incremental changes; they represent a potential leap forward in capabilities, opening new horizons for how data is shared and utilized across various sectors.

In today's rapidly evolving commercial landscape, information sharing serves as a vital cogwheel that ensures seamless communication, real-time decision-making, and process automation across diverse systems and applications [37]. One of the noteworthy capabilities of contemporary data interchange tools is multimodality, which allows a single tool to manage multiple types of information and protocols [38]. This feature is particularly significant as businesses today are burdened with an increasing variety of data formats and communication protocols, requiring a more versatile and robust solution for their integration needs [39]. Multimodality empowers organizations to harmonize information transfers across various domains without requiring multiple tools or complex configurations, thereby reducing both operational overhead and total cost of ownership [40]. In this complex milieu, understanding the scalability—whether vertical or horizontal—of information exchange tools is crucial [41], as is their ability to natively integrate with existing platforms and comply with pertinent regulations such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) [42]. As such, selecting the right data interchange solution necessitates a multi-faceted evaluation, encompassing technical capabilities, regulatory compliance, and even the level of programming expertise required for effective utilization [43].

In a local context, the Signal Processing and Recognition Group (SPRG) at the Universidad Nacional de Colombia has been engaged in the analysis of neurophysiological information to propose and develop machine learning methodologies for assisted diagnosis of mental conditions [44], automated human activity recognition [45], and biomedical dataset analysis [46]. Having achieved the capability to work with proprietary databases for research [47], the group is paving new avenues for collaboration and application of their research through a growing interest in the domain of information exchange. Specifically, there is a focus on the management and analysis of multimodal datasets to maximize the efficacy and applicability of research solutions.

After examining the applications and contexts both globally and locally, it is crucial to advance research in data exchange to support tasks particularly in neuroscience, machine learning for medical diagnostics, and analysis of biomedical datasets [48]. These areas, closely tied to the existing research focus of the SPRG at the Universidad Nacional de Colombia, represent critical fields where enhanced data exchange can significantly contribute to the advancement of knowledge and the development of innovative solutions. The integration of robust data exchange systems in these areas is key to unlocking new possibilities in research and application, particularly in the realms of neurophysiology and healthcare technologies [49]. The integration of data exchange technologies in these sectors promises not only to streamline operational processes but also to facilitate significant advancements in predictive analytics and decision-making. In healthcare, for instance, efficient data exchange can revolutionize patient care by enabling real-time analysis and sharing of critical health data. Therefore, focusing research efforts in these domains is vital to harness the full potential of data exchange technologies, ultimately leading to societal benefits and technological innovation [50].

3 Problem Statement

Adopting type-specific data exchange systems in scientific research offers discernable technical advantages that can significantly elevate the quality and speed of investigations [51]. One of the most salient benefits is the optimization of processing. Algorithms tailored to particular types can substantially reduce computational bottlenecks, thereby accelerating the throughput and efficiency of pipelines [52]. Similarly, the use of specialized encryption algorithms and validation processes adds an additional layer of security that is closely aligned with the unique attributes and vulnerabilities associated with the specific types at hand [53]. On the practical side, the benefits are equally compelling. Specialized exchange systems improve quality through rigorous integrity checks, cleansing, and validation processes that are intricately designed to suit the specific type [54]. This yields a high level of analytical quality, made possible through the application of specialized statistical models and machine learning algorithms that can better capture the nuances and complexities of the subject matter [55]. While initial implementation may come with overhead costs, the long-term savings in both time and resources can be considerable, effectively justifying the initial investment [56]. In addition, the user experience is substantially enhanced by the deployment of intuitive user interfaces and management tools that are customized for the type in question, ultimately making these specialized systems more accessible and user-friendly for researchers [57].

Furthermore, it's crucial to highlight the role of various primary methods utilized in Information Exchange—namely (i) APIs, (ii) Messaging, (iii) ETLs, and (iv) File Transfer—as each method serves specific needs and addresses unique challenges in the field. These techniques often function as the building blocks of a sophisticated Information Exchange architecture, providing specialized solutions for information integration, real-time communication, data transformation, and secure file transfer, respectively.

3.1 APIs in Data Exchange: Navigating Integration Challenges

APIs serve as a critical component in modern Information Exchange architectures [58], providing standardization in software communication. These interfaces are responsible for defining rules, protocols, and tools that facilitate efficient information handling between various systems [59]. As a result, they enable interoperability, simplifying the sharing of information across different platforms and environments [60]. In scenarios involving cloud-based repositories, on-premises databases, or IoT devices, these frameworks ensure a consistent access method, making sure timely and accurate information transmission occurs [61]. This capability not only improves information accessibility but also its usability, as these systems can convert datasets into formats easily interpreted by diverse systems [62]. Hence, they are vital in the structure of any data interchange strategy, enhancing information fluidity and the integration of multi-system architectures [63].

The role of APIs in Data Exchange is highlighted by several critical features that bolster system robustness and functionality [64]. End-to-End Integration enables seamless communication between systems, negating the need for manual data mapping [65]. The Time-Series Data Handling feature is crucial for applications where tracking temporal data sequences is essential [66]. Real-Time Processing supports dynamic interactions, offering immediate feedback or activation of triggers [67]. Secure Data Encryption is vital for protecting the integrity and confidentiality of information during transfer between systems [68]. Scalability allows the API infrastructure to accommodate varying demands, while Fault-Tolerant Design ensures system resilience against component failures [69]. Monitoring and Auditing Capabilities offer transparency and accountability in real-time data transaction tracking [70]. Cross-Platform Compatibility enables APIs to function across diverse operating systems and hardware configurations [71]. Lastly, the capacity for handling Concurrent Operations maximizes system resource utilization, boosting overall performance [72].

In the field of information exchange architectures, several key challenges have been identified in using APIs. A focus on developer-centric approaches is highlighted for enhancing security, suggesting a collaborative stance with developers for more robust protections [73]. The need for stringent standardization in data sharing processes is emphasized to guarantee effective implementation [74]. Challenges in managing large volumes of experimental data and deriving actionable insights are also noted [75]. Scalability issues in cloud applications are linked to programming practices, indicating a need for a change in developer habits [76]. The task of interfacing with varied APIs across different data stores presents complexities in maintaining system coherence [77]. The necessity for customization in APIs to adapt to unique or proprietary data sources is also recognized [78]. Security and scalability challenges in microservices architectures call for a modernized approach in API design [79]. The importance of API performance in fault-tolerance and real-time processing in AI systems is highlighted, underlining their role in system reliability [80]. For distributed cyber-physical systems, scalability and fault-tolerance are essential design considerations [81]. A blockchain framework for auditing collaborative clinical trials is proposed as a novel solution to security issues in data interchange [82]. Security models for socio-cyber-physical systems stress the need to assess the security level of critical business processes, closely linked to API security [83]. Lastly, an innovative multi-key, partially homomorphic encryption scheme for low-end devices is proposed to address privacy and security concerns in APIs used in IoT solutions [84]. These challenges underscore the intricate relationship between API design, security, and system functionality in Data Exchange.

3.2 Messaging Protocols: Streamlining Asynchronous Data Flows

Messaging systems play an integral role in Information Transfer architectures [85] by facilitating asynchronous communication between distributed components [86]. Specialized for Database-Backed Architecture, these systems offer a structured approach to manage and retrieve digital content [87]. They are particularly robust when it comes to Time-Series Data Handling, allowing for seamless management of temporal information sequences [88]. Real-Time Processing capabilities further make messaging systems invaluable for applications requiring immediate information transactions. Additionally, these systems excel in High-Volume Dataset Management, ensuring scalability when managing large datasets [89]. Overall, messaging systems act as a versatile conduit for information flow, accommodating a wide range of information types and operational scales [90].

The utility of messaging systems in Data Exchange is emphasized by a range of features. Event-Driven Architecture allows for responsive system behavior based on specific conditions or triggers [91]. Scalable Architecture ensures the system can adapt as data volumes grow [92], while Fault-Tolerant Design mechanisms offer resilience, allowing continued function even in the presence of component failures [93]. Monitoring and Auditing Capabilities afford system administrators visibility into information flow and transaction history, ensuring data integrity [94]. Low-Latency Operation is crucial for time-sensitive applications, speeding up information transactions [95]. Concurrent Operations capability enables multiple information exchanges to occur simultaneously, optimizing system throughput [96]. These characteristics collectively contribute to the robustness, flexibility, and efficiency of messaging systems in Data Exchange architectures [97].

Another method employed in the realm of Data Exchange is Messaging, which encapsulates different facets of communication in distributed systems. Secure messaging, particularly in healthcare, raises patient safety concerns that necessitate rethinking system designs to maintain confidentiality and integrity [98]. P2P architectures in messaging systems can vastly improve scalability, offering reductions in message transmissions while preserving timely delivery, which is crucial in advanced infrastructure systems [99]. The integration of blockchain technology in vehicular network messaging emphasizes the need for secure, real-time, and distributed data logging to enhance fault tolerance and scalability [100]. Large-scale graph processing systems often suffer from significant performance overhead and extended recovery times; the 'Imitator' concept has been introduced to address replication-based fault tolerance [101]. Real-time traffic monitoring systems pose the challenge of developing scalable and fault-tolerant systems; employing event-based microservices with Apache Kafka Streams has been suggested as a viable solution to handle such high-throughput, real-time data [102]. Furthermore, the integration of workload balancing and fault tolerance is pivotal in distributed stream processing systems, ensuring that these systems can handle dynamic workloads efficiently without compromising system availability or performance [103]. These diverse aspects of Messaging within Data Exchange highlight the ongoing evolution of communication technologies to address the growing demands for performance, reliability, and security in distributed systems.

3.3 ETL Optimization: Enhancing Data Processing Efficiency

ETL processes are specialized mechanisms crucial for Information Exchange [104], predominantly in scenarios that involve batch processing and information warehousing [105]. These processes interface directly with Database-Backed Architectures to pull datasets reliably from source systems into data warehouses [106]. ETL's forte lies in End-to-End Integration, streamlining the complete information flow from its origin to the ultimate repository [107]. Designed with Time-Series Data Handling in mind, ETL processes offer efficient techniques for capturing, transforming, and storing temporal datasets [108].

In terms of system features, ETL excels in High-Volume Dataset Management [109], leveraging its robustness to process extensive datasets without sacrificing performance. This performance scalability is attributed to its Scalable Architecture [110], ensuring adaptability to meet evolving information requirements. Furthermore, a Fault-Tolerant Design has been incorporated into ETL systems to offer resilience, particularly when the information exchange process encounters errors or system failures [111]. Monitoring and Auditing Capabilities provide an additional layer of control and visibility [112], allowing for real-time oversight over the various dataset transformation stages. Finally, the ability to conduct Concurrent Operations means that ETL systems can process multiple information streams in parallel [113], thereby optimizing the overall efficiency of the information exchange process [114].

ETL method is central to data processing and management, presenting various challenges and advancements. The relational data model, despite its growth, faces integration difficulties with big data technologies while preserving Atomicity, Consistency, Isolation, and Durability (ACID) properties, which is essential for the reliability of ETL processes [115]. Stream processing in large-scale wireless networks is another critical aspect of ETL, with specific challenges associated with data analysis within these networks [116]. Fault tolerance is a recurring theme, where an optimized Byzantine Fault Tolerance (BFT) algorithm is presented as pertinent to the fault-tolerant design, which is a crucial component of ETL systems [117]. Sharding databases are proposed to tackle the issues of scaling databases, emphasizing its importance for the ETL process in terms of fault tolerance and scalability [118]. The efficiency, scalability, and security in consensus protocols are also addressed, which, while indirectly related, impact the ETL procedures [119]. Lastly, a Byzantine Fault Tolerance based consensus algorithm is proposed to enhance throughput scalability, a key consideration in the design and optimization of ETL systems [120]. These discussions reflect the ongoing efforts to refine ETL methodologies to meet the demands of robustness and efficiency in information management.

3.4 File Transfer Techniques: Innovating in Data Movement Strategies

File Transfer serves as a fundamental method in Information Exchange architectures [121], specifically designed for the bulk transfer of files between systems. The model excels in High-Volume File Handling [122], offering efficient mechanisms for transferring large or numerous files. Multi-Protocol Support ensures compatibility with a range of transfer protocols, from File Transfer Protocol (FTP) and Secure File Transfer Protocol (SFTP) to Hypertext Transfer Protocol (HTTP), granting it flexibility in various applications. The Plug-and-Play Capability of file transfer mechanisms facilitates their integration into existing systems [123], eliminating the need for complex setup processes. In essence, File Transfer methods offer a simple yet robust approach to file movement, serving as a viable option for many data exchange scenarios [124].

The strengths of File Transfer lie in its key features that contribute to its overall robustness and functionality [125]. Secure File Encryption guarantees the integrity and confidentiality of transferred files [126]. A Scalable Architecture permits adaptation to increasing file sizes and transfer volumes [127]. Monitoring and Auditing Capabilities enable real-time tracking and validation, ensuring file integrity throughout the transfer process [128]. The method's Cross-Platform Compatibility permits its use across various system architectures, making it especially useful in multi-vendor settings [129]. Finally, Concurrent Operations support allows for simultaneous file transfers, thereby optimizing the overall system performance and efficiency [130].

File transfer methods are pivotal in various domains, reflecting the necessity for efficient and secure file exchange. The importance of protocol alternatives for file-intensive applications is underscored, as they are fundamental to optimizing file transfer processes [131]. In the realm of Grid FTP, it is argued that concurrency can be a more effective strategy than parallelism for controlling bandwidth, which is crucial for managing large-scale file transfers. In the agricultural sector, big data challenges necessitate secure file encryption and scalable architecture within blockchain systems to ensure secure and reliable information exchange [132]. The advancements of Industry 4.0 also bring new paradigms to the factories of the future, including the need for robust file transfer mechanisms to support emerging technologies [133]. In media, the integrity of content is paramount; thus, a multi-stakeholder media provenance management system is proposed to counter synthetic media risks, emphasizing the significance of a secure file transfer [134]. The exchange of sensitive health records across jurisdictions, such as HIV surveillance data, highlights the importance of secure file transfer protocols and data integration to maintain data privacy and accuracy [135]. Lastly, the digitalization challenges faced by manufacturing SMEs, including those related to file transfer and scalability, further illustrate the need for efficient file transfer solutions [136]. These facets collectively highlight the evolving landscape of file transfer methodologies and their critical role in the secure and efficient exchange of files across various sectors.

3.5 Summary and research question

In the domain of data exchange architectures, a unified strategy for orchestrating APIs, messaging systems, ETL workflows, and file transfer protocols is critical [137]. The selected challenges encapsulate a crucial convergence where the integrity of digital content, operational efficiency, and system reactivity converge [138], as summarized in Table 2. Proficient API management underpins the fluid transition of datasets across disparate platforms, guaranteeing reliable access and transactional manipulation [139]. This is complemented by robust messaging infrastructures, engineered to facilitate the swift movement of digital information, which is indispensable for dynamic data interaction and timely responses [140]. These messaging frameworks seamlessly integrate with ETL routines that are instrumental in the orderly conversion and consolidation of datasets, preparing them for comprehensive analysis and action-driven insights [141]. Furthermore, the refinement of file transfer mechanisms to optimize bandwidth utilization is imperative, ensuring the expeditious movement of substantial file payloads that modern information ecosystems demand [142]. Addressing these intertwined concerns not only bolsters the efficacy of each individual component but also reinforces the collective information exchange network, culminating in a scalable, robust architecture poised to meet the expanding requirements of contemporary analytical ventures [143]. The resolution of these interconnected issues is crucial, forming the foundation for sophisticated analytics, catalyzing data-informed strategic decisions, and reinforcing the foundational efficiency within organizations reliant on digital content [144].

In response to the distinct data challenges of the 'Sistema de integración de EEG, ECG y SpO2 para seguimiento de neonatos en unidad de cuidados intensivos del Hospital Universitario de Caldas - SES HUC' and the 'Prototipo funcional de Lengua electrónica para identificación de sabores en cacao fino de origen colombiano', a multifaceted data exchange architecture is being proposed. This architecture is designed to manage Casa Luker's unstructured data repositories and address the real-time data acquisition needs of the hospital's NICU. Through the utilization of APIs, the architecture will facilitate seamless access and manipulation of diverse data streams, ensuring coherent data integration for Casa Luker's analytic needs. Concurrently, messaging systems will be established to enable the prompt delivery of critical health metrics, ensuring a responsive data ecosystem for neonatal care. ETL processes will transform

Area of Focus	Key Challenge	Leading Solutions/Technologies	Impact on the Ecosystem/Metrics
APIs	Management of heterogeneous APIs across various platforms, ensuring compatibility and seamless interaction.	Unified API management platforms (e.g., Swagger, Postman)	Facilitates interoperability; metrics may include reduced integration time and error rates.
Messaging	Handling high-volume data efficiently in real-time, asynchronous communication setups.	Advanced message queuing protocols (e.g., Apache Kafka, RabbitMQ)	Essential for scalability; metrics may include throughput and latency measures.
ETL	Efficient data integration and transformation, especially in batch processing and warehousing scenarios.	Modern ETL tools (e.g., Talend, Informatica)	Influences data quality and accessibility; metrics could be processing time and data integrity rates.
File Transfer	Optimizing bandwidth usage for the transfer of large files or datasets, particularly in bandwidth-intensive scenarios.	High-speed transfer protocols (e.g., Aspera, GridFTP)	Critical for performance; metrics might include transfer speed and system utilization rates.

Table 2 Comparative Overview of Data Exchange Domains: Challenges, Solutions, and Impacts

unstructured data into structured insights, while file transfer protocols will handle the secure and efficient movement of large datasets. This combined approach is tailored to support the rigorous demands of both real-time medical care and in-depth agri-food research, reflecting the architecture's versatility and readiness to adapt to a spectrum of data integration and processing requirements.

Amid the interrelated yet distinct challenges inherent to the four key Data Exchange methodologies – APIs, Messaging, ETL, and File Transfer – the significance of integrating multimodal information systems becomes apparent. A pressing research inquiry arises: How can we realize the full potential of multimodal data exchange architectures through the synergistic refinement of API governance, the scaling capacity of messaging systems, the streamlining of ETL operations, and the strategic optimization of file transfer bandwidth?; such advancements are vital to cater to the intricate and heterogeneous requisites of advanced data-driven applications in this era of unprecedented information growth.

4 State of the art

The following exposition presents a detailed landscape of the current state of data exchange technologies, offering a discerning evaluation of the capabilities, design, and integration challenges of various commercial tools and systems. It elucidates the multifaceted criteria essential for robust data exchange, including API management, system scalability, and compliance with regulatory standards. In the wake of rapidly evolving data-driven environments, this assessment serves as a critical resource for understanding the nuances of existing solutions and their alignment with emerging technological trends and industry best practices.

4.1 Evaluation of Commercial Tools for Data Exchange

The Table 3 provides a comparative assessment of commercial data exchange tools, highlighting their proficiency in managing APIs, messaging systems, ETL operations, and file transfers. Tools such as MuleSoft Anypoint and TIBCO stand out for their multimodal approach, offering versatility across various data types and protocols [145]. This multimodality is pivotal in environments where diverse data sources and formats are prevalent, ensuring seamless integration and data flow [146]. Microsoft Azure Data Factory and Apache Kafka are noted for their data-agnostic nature, enabling universal data management capabilities without the need for specific configurations [147]. The evaluation underscores the importance of scalability, native integration, protocol support, and regulatory compliance in system architecture decision-making [148]. These dimensions are critical in ensuring that the chosen tools not only align with the current technological requirements but also adhere to evolving data protection and privacy regulations [149].

While current commercial tools like MuleSoft Anypoint and TIBCO excel in managing multimodal data, they often fall short in dynamic, real-time environments, which are crucial in sectors such as healthcare monitoring. This shortfall is particularly pronounced in scenarios characterized by unpredictable data velocity and volume, challenging these tools to maintain optimal performance without manual intervention or custom development [150]. A detailed comparison of these tools, along with others, can be found in Table ??. This research proposal aims to address these gaps, proposing innovative solutions to enhance the adaptability and efficiency of data exchange in these complex, high-demand environments. The goal is to ensure efficient, secure, and regulation-compliant data handling in critical fields, thereby contributing significantly to the improvement of current data exchange technologies and methodologies [151].

Database Technology	Multimodal	Data-Agnostic	Scalability	Native Integration	Protocol Support	Regulatory Compliance	Required Programming Level
MuleSoft Anypoint	×	-	Vertical/Horiz	Salesforce, SAP	HTTP, REST, SOAP	GDPR, HIPAA	Medium
TIBCO	×	-	Vertical	Salesforce	HTTP, REST, JMS	GDPR	High
IBM Integration Bus	×	-	Vertical	IBM Cloud	HTTP, REST, SOAP, MQTT	GDPR, HIPAA	Medium
Microsoft Azure Data Factory		×	Horizontal	Azure services	HTTP, REST	GDPR, Azure Policy	Low
AWS Data Pipeline	-	-	Horizontal	AWS services	AWS SDK	GDPR, HIPAA	Low
Apache Kafka	-	×	Horizontal	Hadoop, Spark	Kafka Protocol	-	High
Dell Boomi	-	-	Vertical	Salesforce, SAP	HTTP, REST	GDPR, HIPAA	Medium
SAP Data Services	-	-	Vertical	SAP	HTTP, REST, SOAP	GDPR, SAP Policy	High

Table 3 Comparative Analysis of Database Integration Technologies: Multimodality, Scalability, and Compliance Features

Tool/Technology	Key Features	Strengths	Limitations	Applicability
MuleSoft Anypoint	Multimodal, API management	Versatility in data types and protocols	Challenges in dynamic real-time environments	Various industries, except high-velocity environments
TIBCO	Multimodal, versatility in data and protocols	Integration with various data sources	May require custom development for optimal performance	Businesses with diverse integration needs
Microsoft Azure Data Factory	Data-agnostic, universal data management	Easy integration with Azure ecosystems	Specific configurations needed for certain data types	Projects integrated with Microsoft services
Apache Kafka	Real-time data processing, data-agnostic	High performance in large data environments	Complex management for beginners	Systems requiring processing of large volumes of data in real-time

Table 4 Comparative Table of Data Exchange Tools

4.2 Architecture and Design of Data Exchange Systems

Table 5 provides a comprehensive overview of the architecture and design of data exchange systems, categorizing key components such as API Gateways, messaging queues, ETL pipelines, and file transfer services [152]. These components are integral to the system's architecture, interacting with relevant technologies and protocols like REST, GraphQL, RabbitMQ, Apache Kafka, and FTP [153]. They play critical roles in ensuring the performance and stability of the data exchange ecosystem. API Gateways, for instance, serve as pivotal conduits for managing and routing data requests, while also providing protocol translation capabilities [154]. Messaging queues like RabbitMQ and Apache Kafka are essential for handling asynchronous data transfers, thereby decoupling system components for enhanced scalability and fault tolerance [155]. ETL pipelines, facilitated by tools such as Apache NiFi, are crucial for the extraction, transformation, and loading of data, ensuring that data is accurately processed and stored [156]. File transfer services, supporting protocols like FTP and SFTP, are vital for the secure and efficient movement of large data sets, particularly in environments dealing with big data and IoT [157]. The table further underscores the significance of a well-architected system, encompassing data storage solutions, processing engines, monitoring and auditing mechanisms, and security layers [158]. These components collectively contribute to efficient and secure data management, highlighting the need for a holistic approach in system design that addresses not only technical requirements but also operational and compliance aspects [159].

In the architecture and design of data exchange systems, challenges include the complexity of integrating diverse components like API Gateways and messaging queues, especially in large-scale systems [160]. Ensuring consistent performance and stability across components such as ETL pipelines and file transfer services under varying loads is another challenge. Additionally, balancing security with functionality amidst multiple components and protocols presents significant risks, requiring adherence to strict data protection regulations [161]. Lastly, managing real-time data effectively, particularly in high-velocity IoT scenarios, poses challenges in latency and timely processing [162]. Addressing these issues is essential for efficient and secure data management.

4.3 Integration of Technologies in Data Exchange

A clear demonstration of how different technologies converge in data exchange is observed through the synergy between commercial tools and architectural components [163]. Solutions like IBM Integration Bus and Dell Boomi, which provide native integration capabilities and specialized protocol support, are augmented by architectural

Component	Description	Relevant Technologies/Protocols
API Gateway	Serve as the entry point for all clients. Facilitate request routing, composition, and protocol translation.	REST, GraphQL, OpenAPI
Messaging Queue	Handle asynchronous data transfer. Decouple system components for better scalability and fault tolerance.	RabbitMQ, Apache Kafka, MQTT
ETL Pipeline	Extract data from various sources. Transform data into a suitable format. Load data into a data warehouse or database.	Apache NiFi, Talend, Informatica
File Transfer Service	Manage bulk data transfer. Support different protocols like FTP, SFTP, and HTTP for compatibility.	FTP, SFTP, SCP, HTTP
Data Storage	Databases for structured data. Data lakes for unstructured or semi-structured data.	SQL, NoSQL, Hadoop, Amazon S3
Processing Engines	Real-time processing for immediate insights. Batch processing for large datasets.	Apache Spark, Apache Flink, Hadoop MapReduce
Monitoring and Auditing	Track system health, performance metrics, and data integrity. Log activities for compliance and troubleshooting.	Prometheus, Grafana, ELK Stack
Security Layer	Implement encryption, authentication, and authorization.	OAuth, JWT, TLS/SSL
Scalability and Load Management	Use containerization and orchestration tools like Docker and Kubernetes. Implement auto-scaling and load balancing.	Docker, Kubernetes, AWS ECS

Table 5 High-Level Overview of Architecture Design with Relevant Technologies

elements such as messaging queues and file transfer services [164]. These elements are pivotal in facilitating asynchronous data transfer and managing substantial data volumes, a necessity in modern data-driven landscapes [165]. IBM Integration Bus excels in integrating disparate systems and applications, streamlining data flow across an organization. Dell Boomi offers a comprehensive platform for connecting applications and data sources, both in cloud and on-premises environments, enhancing data consistency and accessibility [164]. The incorporation of messaging queues like RabbitMQ or Apache Kafka with these tools enables efficient data processing and dissemination, critical for real-time data handling and analytics [166]. Additionally, file transfer services, supporting protocols like FTP and SFTP, are vital for the secure and efficient movement of large datasets, particularly in environments dealing with big data and IoT [167]. This synergy between tools and architectural components is crucial in addressing scalability, performance, and regulatory compliance challenges in intricate data exchange environments [168]. It highlights the necessity of a cohesive approach where each technology complements the others, forming a robust, adaptable data exchange ecosystem capable of evolving with business needs and technological advancements [169].

In the realm of data exchange technology integration, challenges arise from the need to seamlessly integrate varied systems and applications. Tools like IBM Integration

Bus and Dell Boomi demonstrate this complexity, necessitating advanced solutions for streamlined data flow [1]. Additionally, real-time data processing, particularly with the rise of IoT, requires more efficient and secure methods for handling large datasets [2]. Addressing these challenges in your doctoral thesis could involve developing more cohesive and adaptable strategies, ensuring that each technology component effectively complements others, thus creating a robust, scalable data exchange ecosystem that aligns with evolving business and technological landscapes [3].

4.4 Challenges and Future Trends in Data Exchange

Looking towards the future, the challenges in data exchange encompass the ongoing integration of emerging technologies, managing the increasing complexity of data, and ensuring regulatory compliance. Emerging trends, such as the growing adoption of cloud computing and the IoT, are driving the need for more flexible and secure data exchange systems [170]. The evolution of commercial tools and architectural components must focus on enhancing interoperability, data security, and operational efficiency to meet the demands of an ever-changing data exchange landscape [169]. The integration of cloud-based solutions, like Azure Data Factory and AWS Data Pipeline, is becoming increasingly crucial, offering scalable and versatile platforms for data management [171]. These cloud services facilitate the handling of vast data volumes generated by IoT devices, while also providing robust security measures to protect sensitive information [172]. Furthermore, the advancement in machine learning and artificial intelligence technologies is expected to play a significant role in automating and optimizing data exchange processes, enabling more intelligent and efficient systems [173]. The challenge lies in seamlessly integrating these advanced technologies while maintaining compliance with stringent data protection regulations like GDPR and HIPAA [174]. As the data exchange field continues to evolve, staying abreast of these trends and challenges will be essential for developing systems that are not only technologically advanced but also aligned with legal and ethical standards [175].

The future challenges in data exchange focus on integrating emerging technologies, managing complex data, and ensuring regulatory compliance. With cloud computing and IoT gaining prominence, there's a push for more adaptable, secure systems. Enhancing interoperability, data security, and efficiency is crucial, with cloud solutions like Azure Data Factory and AWS Data Pipeline becoming key players [176]. The role of machine learning and AI in automating and optimizing data exchange is also significant [177]. A major challenge is integrating these technologies while adhering to strict data protection laws like GDPR and HIPAA, ensuring systems are technologically advanced yet legally and ethically compliant [178].

4.5 Summary

In synthesizing the current landscape of data exchange technologies, several critical challenges have emerged, underscoring areas in need of further research and development. Commercial tools like MuleSoft Anypoint and TIBCO, while adept in handling multimodal data, struggle in dynamic, real-time environments, especially in sectors with high data variability such as healthcare monitoring. Architectural and design complexities become apparent in integrating diverse components like API Gateways and messaging queues, particularly in large-scale systems, where ensuring consistent performance and stability across various elements remains a significant hurdle. The integration of advanced technologies within data exchange systems introduces additional challenges, especially in real-time processing scenarios. Furthermore, the impending necessity to balance technological advancements with stringent regulatory compliance, such as GDPR and HIPAA, adds another layer of complexity. These challenges are concisely summarized in Table ??, which highlights the critical need for innovative solutions that enhance adaptability, efficiency, and security in data exchange systems, ensuring they are equipped to handle the evolving demands of modern, data-driven environments.

Challenge Category	Brief Description
Performance in Dynamic Environments	Tools such as MuleSoft Anypoint and TIBCO struggle to adapt in dynamic environments, especially in the healthcare sector.
Integration of Diverse Components	Challenges in integrating API Gateways and message queues in large-scale systems while maintaining performance and stability.
Efficient Real-Time Data Handling	Issues in effectively processing large volumes of data in scenarios with high variability and demand.
Balancing Technological Advancements and Regulatory Compliance	The need to integrate advanced technologies while adhering to strict regulations like GDPR and HIPAA.

Table 6 Challenges in Data Exchange Technologies

5 Aims

5.1 General Aim

Develop an integrated architecture for multimodal data exchange that ensures scalability, operational efficiency, and effective data transfer management to meet the complex demands of contemporary data-driven applications.

5.2 Specific Aims

- **Develop a Unified API Gateway** that facilitates secure and efficient integration between disparate systems, effectively managing authentication, authorization, and request routing.
- **Implement a Message System for Multimodal Data** that supports asynchronous communication and high-volume data processing, maintaining service integrity and availability under high demand.
- **Develop a Data Transfer and Storage Framework** that streamlines the processing, integration, and accessibility of data, while optimizing bandwidth for diverse data handling needs.

6 Architecture Design for a Data Exchange System

This section outlines the architecture design for the proposed data exchange system, highlighting its key components and their roles in facilitating efficient and secure data transfer.

API Gateway

- **Function**: Serves as the single entry point for all clients.
- **Responsibilities**: Routing requests, composition, protocol translation.
- **Benefits**: Ensures a seamless interface for client interactions.

Messaging Queue

- **Role**: Manages asynchronous data transfers.
- **Advantages**: Enhances scalability, decouples system components, and improves fault tolerance.

ETL Pipeline

- **Purpose**: Handles data extraction, transformation, and loading.
- **Integration**: Connects with various data sources and storage solutions.

File Transfer Service

- **Functionality**: Facilitates bulk data transfer.
- **Support**: Compatible with protocols like FTP, SFTP, and HTTP.

Data Storage

- **Types**: Databases for structured data; Data lakes for unstructured/semi-structured data.
- **Utility**: Accommodates diverse data types within the system.

Processing Engines

- **Division**: Split into real-time processing and batch processing units.
- **Capabilities**: Immediate insights and efficient handling of large datasets.

Monitoring and Auditing

- **Scope**: Includes tracking of system health, performance metrics, and data integrity.
- **Tools**: Employ logging for compliance and troubleshooting.

Security Layer

- **Measures**: Encryption, authentication, and authorization.
- **Objective**: Ensures data security throughout the exchange process.

Scalability and Load Management

- **Technologies**: Utilizes Docker and Kubernetes for containerization and orchestration.
- **Strategies**: Implements auto-scaling and load balancing.

In summary, the proposed architecture, as illustrated in figure 1, provides a comprehensive solution for a robust and efficient data exchange system. It is designed to be scalable, secure, and adaptable to future data demands and technological advancements.

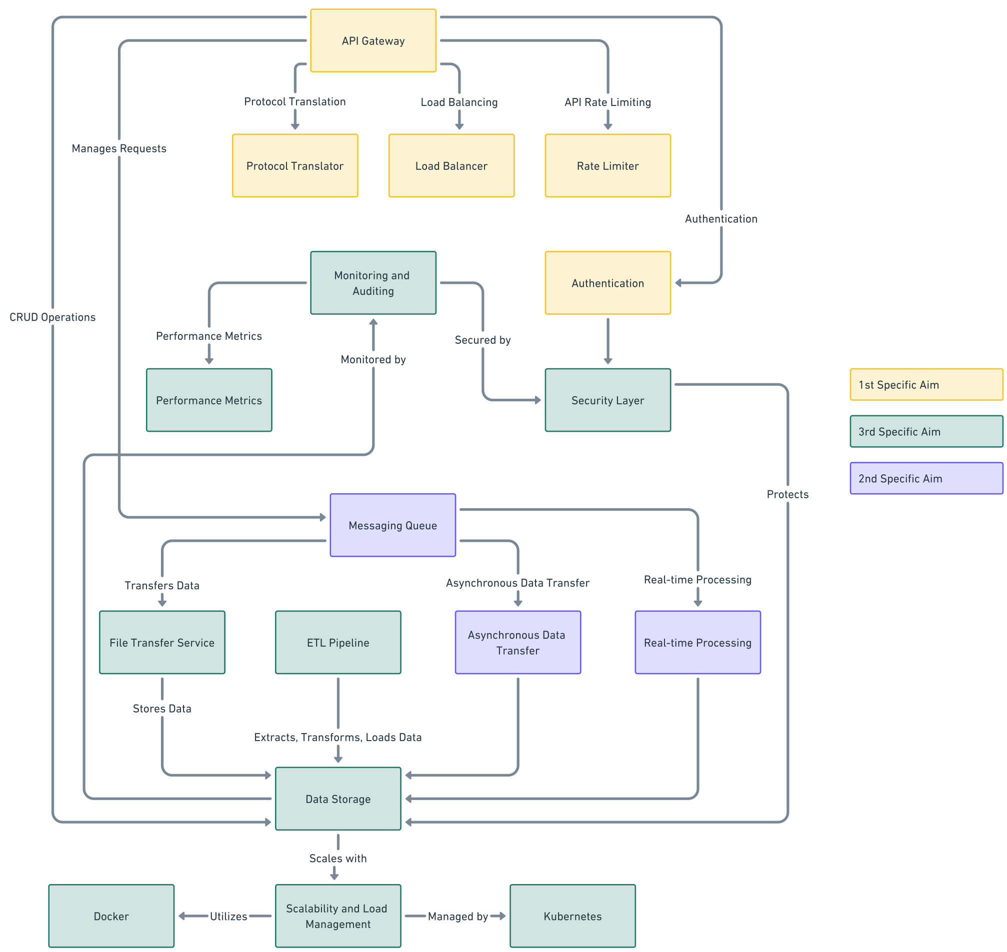

Figure 1 High-Level Architecture Diagram: Integration and Data Flow in Processing and Storage Systems

7 Methodology for System Development and Implementation

The development of a sophisticated data exchange system is governed by a methodology that emphasizes systematic planning, iterative development, and continuous improvement. This methodology is structured into distinct phases to ensure a thorough approach from initial conceptualization to deployment and beyond.

7.1 Objective 1: Develop a Unified API Gateway

- **Phase 1 (Requirement Analysis):** Define the specific requirements for the API Gateway, including security protocols, types of systems to be integrated, and routing mechanisms.
- **Phase 2 (System Design):** Design the architecture of the API Gateway, detailing how it will handle requests, manage authentication, and ensure authorization.
- **Phase 4 (Development and Testing):** Develop the API Gateway, followed by rigorous testing for security, functionality, and performance.
- **Phase 6 (Deployment and Scaling):** Deploy the API Gateway in a controlled environment, initially with limited integrations to monitor its performance and scalability.
- **Phase 7 (Monitoring and Optimization):** Continuously monitor the API Gateway for performance and make necessary optimizations for efficiency.

7.2 Objective 2: Implement a Message System for Multimodal Data

- **Phase 3 (Prototyping and Iteration):** Develop a prototype of the message system to test its feasibility for handling multimodal data and high-volume processing.
- **Phase 4 (Development and Testing):** Full-scale development of the message system, ensuring it maintains integrity and availability under high demand.
- **Phase 5 (Data Governance and Compliance):** Implement data governance practices and compliance checks within the message system.
- **Phase 7 (Monitoring and Optimization):** Monitor the system post-deployment and optimize for improved performance and service integrity.

7.3 Objective 3: Develop a Data Transfer and Storage Framework

- **Phase 1 (Requirement Analysis):** Analyze and document the requirements for data transfer and storage, including bandwidth optimization and data handling needs.
- **Phase 2 (System Design):** Design the framework, focusing on how it will handle different data types, optimize bandwidth, and ensure data accessibility.
- **Phase 4 (Development and Testing):** Develop the framework with emphasis on efficient data processing and integration.
- **Phase 6 (Deployment and Scaling):** Deploy the framework, starting with smaller datasets and scale as needed to handle larger data volumes.
- **Phase 8 (Documentation and Training):** Prepare documentation for the framework and conduct training sessions for users and administrators.

8 Realized Outcomes: Advancements in Data Integration and Analytical Techniques

This section delineates the significant milestones achieved in the realm of neurophysiological data acquisition and analysis. Through rigorous research and development, a suite of tools and frameworks has been devised to address the nuances of data management in diverse settings, ranging from the intricacies of EEG-based neurophysiological experiments to the critical demands of neonatal intensive care. Each of the following subsections elaborates on the advancements within their respective domains, highlighting the integration of innovative technologies and methodologies that collectively push the boundaries of precision, efficiency, and scalability in data processing and analysis.

8.1 Enhancing EEG Research: Advancements in OpenBCI Framework for Neurophysiological Studies

The article ”A Novel OpenBCI Framework for EEG-Based Neurophysiological Experiments” [179] introduces a new framework designed to improve the efficiency and flexibility of conducting EEG experiments using the OpenBCI platform, particularly with the Cyton board and ADS1299 hardware. This system supports distributed computing, various electrode configurations, and real-time feedback, addressing current Brain-computer interface (BCI) limitations in communication and configuration for specific neurophysiological protocols. It provides a scalable and adaptable solution for real-time data processing in BCI applications, including user-friendly interfaces for stimuli delivery and motor imagery, demonstrating its potential for advancing EEG research and BCI technology.

8.2 TimeScaleDB App: Innovating Time-Series Data Management with Multimedia Integration

TimeScaleDB [180] App is currently under development as a web application built on Django, specifically crafted for the management and querying of time-series data. It harnesses the capabilities of RealTimeDB, a dedicated time-series database that builds upon the reliable PostgreSQL platform, ensuring efficient data storage and advanced analysis of time-based datasets. The application presently offers API endpoints for seamless interaction with various models, such as Source, Measure, Channel, and Time-Series. It also incorporates custom pagination classes and viewsets to bolster user experience and flexibility. Looking ahead, there are plans to expand its functionality by integrating multimedia support, allowing users to associate images and video with their time-series data, thereby enriching the analytical context and user interactivity of the platform.

8.3 The Foundation Framework: Integrating Microservices and Real-Time Processing for Time-Series Analysis

The Foundation framework [181] is a microservices-based, Docker-integrated system designed to provide a consistent and scalable environment for time-series data management within my doctoral research. Built with Python, it leverages Django and the Django REST framework to create a flexible API that seamlessly connects with TimescaleDB, establishing a solid base for comprehensive time-series data manipulation. With Apache Kafka incorporated, it enables meticulous real-time data streaming and processing, while real-time clock synchronization maintains strict timing accuracy. This framework is set to enhance data analytical capabilities significantly and introduce multimedia data integration, offering a richer, more detailed data analysis experience.

8.4 Synchronizing Neonatal Care: ESP32 Integration for Data Centralization in NICU

The development and implementation of the ESP32-based centralization system in the NICU is currently underway, poised to revolutionize the way patient data is collected and monitored. These microcontrollers are adeptly configured to passively connect to various patient monitors and devices, ensuring a non-intrusive yet effective data acquisition process. The paramount challenge in this endeavor is the meticulous synchronization of signals, which is critical for maintaining the integrity and coherence of data from disparate sources. Adding to the complexity is the need to manage a high volume of data traffic to the servers, necessitating a robust infrastructure that can support the continuous influx without compromising performance. As the system is being honed for deployment, addressing these dual challenges of synchronization and data flow management is essential for providing a seamless and reliable monitoring solution in the delicate environment of the NICU.

8.5 Deep Learning in MI Paradigms: A Leap Forward in Real-Time BCI Processing

Presented at the "III Congreso Latinoamericano de Investigación, Innovación y Emprendimiento Educativo", the paper "Real-Time Processing for BCI-Based MI Paradigms Using Deep Learning Models" marks a significant milestone in the evolution of BCI technologies. This work delves into the integration of advanced deep learning models for the interpretation and real-time processing of Motor imagery (MI) paradigms within BCIs. The focus lies in enhancing the speed and accuracy of signal processing in BCI systems, leveraging state-of-the-art neural network architectures. The paper discusses the challenges and solutions in implementing these models for real-time applications, shedding light on the potential of deep learning to transform the landscape of neurophysiological research and applications. The effectiveness of the proposed approach is demonstrated through rigorous experimental setups, showcasing notable improvements in response times and overall system performance, thereby setting a new benchmark for real-time BCI processing. This study not only contributes to the academic field but also paves the way for practical, user-friendly BCI applications in various sectors, including healthcare and assistive technologies.

8.6 Advances in Marker Synchronization for Brain-Computer Interface Systems

This subsection highlights the advancements achieved through the development of a patented method and system for marker synchronization in BCI systems. This innovative approach, as detailed in the patent, addresses the challenges in the synchronization of markers associated with BCI systems, contributing significantly to the accuracy and reliability of signal acquisition in neurophysiological research and applications. This method enhances the precision of marker synchronization, which is crucial for the correct interpretation and analysis of EEG signals in BCI experiments, thereby advancing the field of neurophysiology and the practical applications of BCIs in various domains.

8.7 Advancements in Marker Synchronization for Brain-Computer Interface Systems

The system, titled ”MÉTODO Y SISTEMA PARA LA SINCRONIZACIÓN DE MARCADORES ASOCIADOS A SISTEMAS DE INTERFAZ CEREBRO-COMPUTADOR,” was submitted for a patentability search process under the ”Crearlo no es suficiente” initiative. The Universidad Nacional de Colombia sede Manizales is listed as the main beneficiary. This submission, with the postulation ID 343 and Application number NC2022/0007405, was made on May 28, 2022. This patent application represents a significant advancement in the field of BCI systems, specifically focusing on the synchronization of markers within these systems. This development is crucial for enhancing the accuracy and reliability of signal acquisition in neurophysiological research and BCI applications.

9 Implementation schedule

The successful realization of any complex, multifaceted project hinges critically on the development and adherence to a comprehensive implementation schedule. This section delineates crafted timeline for the various phases of our project, encompassing the development, testing, deployment, and optimization of the proposed systems and methodologies.

	Activities									
	Aim 1			Aim 2			Aim 3			Scientific text
Months	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3	Writing
1-2	×	×								
2-4		×								
4-6										
6-8										
8-10										
10-12										
12-14										
14-16										
16-18										

References

[1] Emmanuel Helm, Andreas Schuler, and Herwig Mayr. Cross-enterprise communication and data exchange in radiology in austria: Technology and use cases. *eHealth*, 248:64–71, 2018. (page 3)

[2] Tiina Tuominen, Catalina Jiménez Hurtado, and Anne Ketola. Why methods matter: Approaching multimodality in translation research. *Linguistica Antverpiensia: New Series–Themes in Translation Studies*, 17:1–21, 2018. (page 3)

[3] Yang Qiu, Yang Zhang, Yifan Deng, Shichao Liu, and Wen Zhang. A comprehensive review of computational methods for drug-drug interaction detection. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(4):1968–1985, 2021. (page 3)

[4] Betül Güvenç Paltun, Samuel Kaski, and Hiroshi Mamitsuka. Diverse: Bayesian data integrative learning for precise drug response prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(4):2197–2207, 2021. (page 3)

[5] Aniruddha Adiga, Gursharn Kaur, Benjamin Hurt, Lijing Wang, Przemyslaw Porebski, Srinivasan Venkatramanan, Bryan Lewis, and Madhav Marathe. Enhancing covid-19 ensemble forecasting model performance using auxiliary data sources. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1594–1603. IEEE, 2022. (page 3)

[6] Vitalii I Yesin, Mikolaj Karpinski, Maryna V Yesina, Vladyslav V Vilihura, Olga Veselska, and Lukasz Wieclaw. Approach to managing data from diverse sources. In *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, volume 1, pages 1–6. IEEE, 2019. (page 3)

[7] Saurabh Shukla, Mohd Fadzil Hassan, Duc Chung Tran, Rehan Akbar, Irving Vitra Paputungan, and Muhammad Khalid Khan. Improving latency in internet-of-things and cloud computing for real-time data transmission: a systematic literature review (slr). *Cluster Computing*, pages 1–24, 2021. (page 3)

[8] Jin Du et al. Real-time information exchange strategy for large data volumes based on iot. *Computational Intelligence and Neuroscience*, 2022, 2022. (pages 3 and 5)

[9] Jie Yan, Yigong Huang, and Lei Song. Research on real-time data transmission and feature analysis technology of smart tool holder. In *Journal of Physics: Conference Series*, volume 2419, page 012102. IOP Publishing, 2023. (page 3)

[10] Sharif Islam, Alex Hardisty, Wouter Addink, Claus Weiland, and Falko Glöckler. Incorporating rda outputs in the design of a european research infrastructure for natural science collections. *Data Science Journal*, 19:50–50, 2020. (page 3)

[11] Nelson Tavares de Sousa, Wilhelm Hasselbring, Tobias Weber, and Dieter Kranzlmüller. Designing a generic research data infrastructure architecture with continuous software engineering. 2018. (page 3)

[12] Ivana Podnar Žarko, Joaquin Iranzo, Christoph Ruggenthaler, Jose Antonio Sanchez Murillo, João Garcia, Pavle Skočir, and Sergios Soursos. Collaboration mechanisms for iot platform federations fostering organizational interoperability. In *2018 Global Internet of Things Summit (GIoTS)*, pages 1–6. IEEE, 2018. (page 3)

[13] Samodha Pallewatta, Vassilis Kostakos, and Rajkumar Buyya. Microservices-based iot applications scheduling in edge and fog computing: A taxonomy and future directions. *arXiv preprint arXiv:2207.05399*, 2022. (page 3)

[14] Abhijeet Thakare and Young-Gab Kim. Secure and efficient authentication scheme in iot environments. *Applied Sciences*, 11(3):1260, 2021. (page 3)

[15] Cinzia Cappiello, Avigdor Gal, Matthias Jarke, and Jakob Rehof. Data ecosystems: sovereign data exchange among organizations (dagstuhl seminar 19391). In *Dagstuhl Reports*, volume 9. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2020. (page 4)

[16] Orlando Troisi and Mara Grimaldi. Guest editorial: Data-driven orientation and open innovation: the role of resilience in the (co-) development of social changes. *Transforming Government: People, Process and Policy*, 16(2):165–171, 2022. (page 4)

[17] Nader Alagha and Lars Løge. Ijsc&n special issue "opportunities and challenges of maritime vhf data exchange systems": Guest editorial message. *International Journal of Satellite Communications and Networking*, 41(2):99–101, 2023. (page 4)

[18] Alexander S Suleykin and Peter B Panfilov. Designing data-intensive application system for production plans data processing and near real-time analytics. In *2022 8th international conference on control, decision and information technologies (CoDIT)*, volume 1, pages 1495–1500. IEEE, 2022. (page 4)

[19] Nader Trabelsi, Cristiano Politowski, and Ghizlane El Boussaidi. Event driven architecture: An exploratory study on the gap between academia and industry. In *2023 IEEE/ACM 5th International Workshop on Software Engineering Research and Practices for the IoT (SERP4IoT)*, pages 25–32. IEEE, 2023. (page 4)

[20] Xinxin Zhang, Li Xu, and Aihua Li. Fault-tolerant secure routing of bh_n-based data center networks. *Security and Communication Networks*, 2021:1–10, 2021. (page 4)

[21] J Christian Attiogbé, Flavio Ferrarotti, and Sofian Maabout. Advances and challenges for model and data engineering. *J. Univers. Comput. Sci.*, 27(7):646–649, 2021. (page 4)

[22] Daniel M Walker, Valerie A Yeager, John Lawrence, and Ann Scheck Mcalearney. Identifying opportunities to strengthen the public health informatics infrastructure: exploring hospitals' challenges with data exchange. *The Milbank Quarterly*, 99(2):393–425, 2021. (page 4)

[23] Puyuan Zheng, Tong Wu, and Jingyang Dong. Edge-based dynamic data integrity checking scheme for iot. In *2022 10th International Conference on Information Systems and Computing Technology (ISCTech)*, pages 434–441. IEEE, 2022. (page 4)

[24] Loris Belcastro, Riccardo Cantini, Fabrizio Marozzo, Alessio Orsino, Domenico Talia, and Paolo Trunfio. Programming big data analysis: principles and solutions. *Journal of Big Data*, 9(1):1–50, 2022. (page 4)

[25] Alex Groh, Trevor Olson, Carter Nenninger, Anil Godumagadda, and Evan Corne. Rules engine exchange: Driving operational excellence through an advanced real-time alerting system. In *IADC/SPE International Drilling Conference and Exhibition*. OnePetro, 2022. (page 4)

[26] Calogero Carletto. Better data, higher impact: improving agricultural data systems for societal change. *European Review of Agricultural Economics*, 48(4):719–740, 2021. (page 4)

[27] Mounir El Khatib, Samer Hamidi, Ishaq Al Ameeri, Hamad Al Zaabi, and Rehab Al Marqab. Digital disruption and big data in healthcare-opportunities and challenges. *ClinicoEconomics and Outcomes Research*, pages 563–574, 2022. (page 4)

[28] Lokanath Mishra and Vaibhav Kaushik. Application of blockchain in dealing with sustainability issues and challenges of financial sector. *Journal of Sustainable Finance & Investment*, 13(3):1318–1333, 2023. (page 4)

[29] Guido Luzi, José Antonio Navarro, Anna Barra, Oriol Monserrat, and Michele Crosetto. Heimdall: a h2020 project aimed at developing a multi-hazard cooperative management, data exchange, response planning and scenario building tool: the landslides case. In *EGU General Assembly Conference Abstracts*, pages EGU21–9994, 2021. (page 4)

[30] Hazar DORDUNCU and Zemzem Selin ORUÇ. Industry 4.0, digitalization, and big data: perspective of logistics and supply chain management. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 30(3):170–180, 2021. (page 5)

[31] Marieh Talebkhah, Aduwati Sali, Meisam Gordan, Shaiful Jahari Hashim, and Fakhrul Zaman Rokhani. Comprehensive review on development of smart cities using industry 4.0 technologies. *IEEE Access*, 2023. (page 5)

[32] Tanweer Alam. Blockchain-based internet of things: Review, current trends, applications, and future challenges. *Computers*, 12(1):6, 2022. (page 5)

[33] Sabri Barbaria, Marco Casassa Mont, Essam Ghadafi, Halima Mahjoubi Machraoui, and Hanene Boussi Rahmouni. Leveraging patient information sharing using blockchain-based distributed networks. *IEEE Access*, 10:106334–106351, 2022. (page 5)

[34] Georgios Lampropoulos, Kerstin Siakas, Julio Viana, and Olaf Reinhold. Artificial intelligence, blockchain, big data analytics, machine learning and data mining in traditional crm and social crm: A critical review. In *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 504–510. IEEE, 2022. (page 5)

[35] Zhanyu Liu and Xishuan Zhang. Multimedia and multimodal sensing with edge computing for personalized healthcare supply chain system data optimization. *Personal and Ubiquitous Computing*, 27(3):955–972, 2023. (page 5)

[36] M Riedel, M Book, H Neukirchen, G Cavallaro, and A Lintermann. Practice and experience using high performance computing and quantum computing to speed-up data science methods in scientific applications. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, pages 281–286. IEEE, 2022. (page 5)

[37] Mbongowo Mbuh, Peter Metzger, Peter Brandt, Kelli Fika, and Monica Slinkey. Application of real-time gis analytics to support spatial intelligent decision-making in the era of big data for smart cities. *EAI Endorsed Transactions on Smart Cities*, 4(9), 2019. (page 5)

[38] Mohammad Mansour, Amal Gamal, Ahmed I Ahmed, Lobna A Said, Abdelmoniem Elbaz, Norbert Herencsar, and Ahmed Soltan. Internet of things: A comprehensive overview on protocols, architectures, technologies, simulation tools, and future directions. *Energies*, 16(8):3465, 2023. (page 5)

[39] Inna Kouper and Kimberly Cook. Challenges in curating interdisciplinary data in the biodiversity research community. *Biodiversity Information Science and Standards*, 5:e79084, 2021. (page 5)

[40] Danilo Bojović, Nebojša Obradović, and Nebojša Kurjakov. Improving operational efficiency and reducing costs in distribution utility with the use of iec 61850 communication protocol. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–5. IEEE, 2022. (page 5)

[41] Ahmed Hussein Ali. A survey on vertical and horizontal scaling platforms for big data analytics. *International Journal of Integrated Engineering*, 11(6):138–150, 2019. (page 5)

[42] W Liu, EK Park, U Krieger, and SS Zhu. Smart e-health security and safety monitoring with machine learning services. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–6. IEEE, 2020. (page 5)

[43] Marwa Mahmoudi, Abdelwaheb Aydi, and Hatem Ibrahim. Site selection for artificial recharge with treated wastewater with the integration of multi-criteria evaluation and electre iii. *Environmental Science and Pollution Research*, pages 1–16, 2021. (page 6)

[44] Mehdi Ordikhani-Seyedlar and Mikhail A Lebedev. Augmenting attention with brain–computer interfaces. In *Brain–Computer Interfaces Handbook*, pages 549–560. CRC Press, 2018. (page 6)

[45] John LaRocco, Minh Dong Le, and Dong-Guk Paeng. A systemic review of available low-cost eeg headsets used for drowsiness detection. *Frontiers in neuroinformatics*, page 42, 2020. (page 6)

[46] Victoria Peterson, Catalina Galván, Hugo Hernández, and Ruben Spies. A feasibility study of a complete low-cost consumer-grade brain-computer interface system. *Heliyon*, 6(3), 2020. (page 6)

[47] Yessica Alejandra Gomez Rivera. *Estrategia de procesamiento de señales EEG en sistemas BCI utilizando aprendizaje profundo y medidas de conectividad*. PhD thesis, Universidad Nacional de Colombia. (page 6)

[48] Jarrett F. Lebov, Kelly Watson, Trevor Tyson, Benjamin C. Sparklin, and David A. Rasko. Analytical methods, bioinformatic tools and pipelines a highly generalizable semi-supervised deepsvdd methodology for detecting anomalous metagenomic samples. 2022. (page 6)

[49] Mohammad S Al-Kahtani, Faheem Khan, and Whangbo Taekeun. Application of internet of things and sensors in healthcare. *Sensors*, 22(15):5738, 2022. (page 6)

[50] Bouthaina Dammak, Mariem Turki, Saoussen Cheikhrouhou, Mouna Baklouti, Rawya Mars, and Afef Dhahbi. Lorachaincare: An iot architecture integrating blockchain and lora network for personal health care data monitoring. *Sensors*, 22(4):1497, 2022. (page 6)

[51] Nibras Talib Mohammed, Enas Ali Mohammed, and Zinah A Abutiheen. The scientific research impediments according to the viewpoint of karbala university academics. In *Journal of Physics: Conference Series*, volume 1818, page 012051. IOP Publishing, 2021. (page 6)

[52] Nico Curti. Implementation and optimization of algorithms in biomedical big data analytics. 2020. (page 6)

[53] Sharad Salunke, Bharti Ahuja, Mohammad Farukh Hashmi, Venkatadri Marriboyina, and Neeraj Dhanraj Bokde. 5d gauss map perspective to image encryption with transfer learning validation. *Applied Sciences*, 12(11):5321, 2022. (page 6)

[54] Ionut Iosifescu Enescu, Gian-Kasper Plattner, Lucia Espona Pernas, Dominik Haas-Artho, and Rebecca Buchholz. Improved fair data publication quality in specialized environmental data portals. Technical report, Copernicus Meetings, 2021. (page 7)

[55] Mélina Côté, Mazid Abiodoun Osseni, Didier Brassard, Élise Carbonneau, Julie Robitaille, Marie-Claude Vohl, Simone Lemieux, François Laviolette, and Benoît Lamarche. Are machine learning algorithms more accurate in predicting vegetable and fruit consumption than traditional statistical models? an exploratory analysis. *Frontiers in Nutrition*, 9:740898, 2022. (page 7)

[56] AV Eder and OV Ivanov. Improving the efficiency of food industry enterprises as a result of the modern it solutions implementation. *Proceedings of the Voronezh State University of Engineering Technologies*, 81(3):364–367, 2019. (page 7)

[57] Jennifer Dickman Portz, Elizabeth A Bayliss, Sheana Bull, Rebecca S Boxer, David B Bekelman, Kathy Gleason, and Sara Czaja. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. *Journal of medical Internet research*, 21(4):e11604, 2019. (page 7)

[58] Athanasios Kiourtis, Argyro Mavrogiorgou, Dimosthenis Kyriazis, Alessio Graziani, and Francesco Torelli. Improving health information exchange through wireless communication protocols. In *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 32–39. IEEE, 2020. (page 7)

[59] David Bell and Jun Hong. *Flexible and Efficient Information Handling: 23rd British National Conference on Databases, BNCOD 23, Belfast, Northern Ireland, UK, July 18-20, 2006, Proceedings*, volume 4042. Springer, 2006. (page 7)

[60] Akshay Ayyanchira and Elias Mahfoud. Cross-platform immersive visualization and navigation with augmented reality. *Journal of visualization*, 2022. (page 7)

[61] Dhruv Garg, Prathik Shirolkar, Anshu Shukla, and Yogesh Simmhan. Torquedb: Distributed querying of time-series data from edge-local storage. In *Euro-Par 2020: Parallel Processing: 26th International Conference on Parallel and Distributed Computing, Warsaw, Poland, August 24–28, 2020, Proceedings 26*, pages 281–295. Springer, 2020. (page 7)

[62] Giuseppe Desogus, Emanuela Quaquero, Giulia Rubiu, Gianluca Gatto, and Cristian Perra. Bim and iot sensors integration: A framework for consumption and indoor conditions data monitoring of existing buildings. *Sustainability*, 13(8):4496, 2021. (page 7)

[63] Qingsu He, Yuping Liu, Lihua Jiang, Zhiqiang Zhang, Muqing Wu, and Min Zhao. Data sharing mechanism and strategy for multi-service integration for smart grid. *Energies*, 16(14):5294, 2023. (page 7)

[64] ⊠ Yu Doroshenko and BV Bodak. Designing restful api for the e-procurement system in private sector. *PROBLEMS IN PROGRAMMING*, (1):3–15, 2021. (page 7)

[65] Tobias Zeimetz and Ralf Schenkel. Sample driven data mapping for linked data and web apis. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 3481–3484, 2020. (page 7)

[66] Shreshth Tuli, Giuliano Casale, and Nicholas R Jennings. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *arXiv preprint arXiv:2201.07284*, 2022. (page 7)

[67] Youn J Kang, Hany M Arafa, Jae-Young Yoo, Cagla Kantarcigil, Jin-Tae Kim, Hyoyoung Jeong, Seonggwang Yoo, Seyong Oh, Joohee Kim, Changsheng Wu, et al. Soft skin-interfaced mechano-acoustic sensors for real-time monitoring and patient feedback on respiratory and swallowing biomechanics. *NPJ digital medicine*, 5(1):147, 2022. (page 7)

[68] Dinesh Kumar Anguraj. Advanced encryption standard based secure iot data transfer model for cloud analytics applications. *Journal of Information Technology and Digital World*, 4(2):114–124, 2022. (page 8)

[69] Ceasar E Eko, Idongesit Eteng, and Eyo E Essien. Design and implementation of a fault tolerant web-based examination system for developing countries. *Eastern-European Journal of Enterprise Technologies*, 1(2):115, 2022. (page 8)

[70] Md Raisul Hasan Shahrukh, Md Tabassinur Rahman, and Nafees Mansoor. Aid nexus: A blockchain based financial distribution system. *arXiv preprint arXiv:2311.08372*, 2023. (page 8)

[71] Andre Augusto Menegassi and Andre Takeshi Endo. Automated tests for cross-platform mobile apps in multiple configurations. *IET software*, 14(1):27–38, 2020. (page 8)

[72] Andrew Griesdorn, Quinn Hirt, Raed Salih, Michael R Clark, and Scott Brookes. Support emergency response in automatic identification system using an opportunistic resource utilization networks. In *Automatic Target Recognition XXXIII*, volume 12521, pages 187–191. SPIE, 2023. (page 8)

[73] M. Green and Matthew Smith. Developers are not the enemy!: The need for usable security apis. *IEEE Security Privacy*, 14:40–46, 2016. (page 8)

[74] O. Borgogno and G. Colangelo. Data sharing and interoperability: Fostering innovation and competition through apis. *Comput. Law Secur. Rev.*, 35:105314, 2019. (page 8)

[75] J. Vetter and M. O. McCracken. Statistical scalability analysis of communication operations in distributed applications. pages 123–132, 2001. (page 8)

[76] Justin Y. Shi, Moussa Taifi, A. Pradeep, Abdallah Khreishah, and Vivek Antony. Program scalability analysis for hpc cloud: Applying amdahl's law to nas benchmarks. *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, pages 1215–1225, 2012. (page 8)

[77] Rami Sellami, S. Bhiri, and Bruno Defude. Odbapi: A unified rest api for relational and nosql data stores. *2014 IEEE International Congress on Big Data*, pages 653–660, 2014. (page 8)

[78] Idowu Ayoola, M. Wetzels, Peter J. F. Peters, S. V. Berlo, and L. Feijs. Do change platform: A service-based architecture for secure aggregation and distribution of health and wellbeing data. *International journal of medical informatics*, 117:103–111, 2018. (page 8)

[79] Victor Velepucha and Pamela Flores. A survey on microservices architecture: Principles, patterns and migration challenges. *IEEE Access*, 2023. (page 8)

[80] Ioannis Kourouklides and Kleitos Alexandrou. An overview of the gut-ai foundation: Vision for an ecosystem of concepts and implementations. 2023. (page 8)

[81] Giovanni Cicceri, Giuseppe Tricomi, Luca D'Agati, Francesco Longo, Giovanni Merlino, and Antonio Puliafito. A deep learning-driven self-conscious distributed cyber-physical system for renewable energy communities. *Sensors*, 23(9):4549, 2023. (page 8)

[82] Nail Adeeb Ali Abdu and Zhaoshun Wang. Blockchain framework for collaborative clinical trials auditing. *Wireless Personal Communications*, 132(1):39–65, 2023. (page 8)

[83] Serhii Yevseiev, Yuliia Khokhlachova, Serhii Ostapov, Oleksandr Laptiev, Olha Korol, Stanislav Milevskyi, Oleksandr Milov, Serhii Pohasii, Yevgen Melenti, Vitalii Hrebeniuk, et al. Models of socio-cyber-physical systems security. 2023. (page 8)

[84] Saci Medileh, Abdelkader Laouid, Mohammad Hammoudeh, Mostefa Kara, Tarek Bejaoui, Amna Eleyan, and Mohammed Al-Khalidi. A multi-key with partially homomorphic encryption scheme for low-end devices ensuring data integrity. *Information*, 14(5):263, 2023. (page 8)

[85] Moirangthem Sailash Singh, Ramkrishna Pasumarthy, Umesh Vaidya, and Steffen Leonhardt. On quantification and maximization of information transfer in network dynamical systems. *Scientific Reports*, 13(1):5588, 2023. (page 8)

[86] Vishal Sawant, Debraj Chakraborty, and Debasattam Pal. Asynchronous distributed consensus with minimum communication. *arXiv preprint arXiv:2305.02448*, 2023. (page 8)

[87] Allan Vikiru, Mfadhili Muiruri, and Ismail Ateya. An overview on cloud distributed databases for business environments. *arXiv preprint arXiv:2301.10673*, 2023. (page 8)

[88] Ana Almeida, Susana Brás, Susana Sargento, and Filipe Cabral Pinto. Time series big data: a survey on data stream frameworks, analysis and algorithms. *Journal of Big Data*, 10(1):83, 2023. (page 8)

[89] Khalid Ayed Alharthi, Arshad Jhumka, Sheng Di, Lin Gui, Franck Cappello, and Simon McIntosh-Smith. Time machine: generative real-time model for failure (and lead time) prediction in hpc systems. *Proceedings of the DSN 2023*, 2023. (page 9)

[90] Pierre-Frédérick DENYS, Michel R Dagenais, and Martin Pepin. Advanced tracing methods for container messaging systems analysis. 2021. (page 9)

[91] Jens Strueker and Harald Weppner. A cloud-based messaging service for cross-enterprise data exchange with smart objects. 2012. (page 9)

[92] Gleb Polozhiy and Nikolay Boldyrikhin. Mail client with data transfer protected with end-to-end encryption. In *E3S Web of Conferences*, volume 363. EDP Sciences, 2022. (page 9)

[93] Alam Rahmatulloh, Fuji Nugraha, Rohmat Gunawan, and Irfan Darmawan. Event-driven architecture to improve performance and scalability in microservices-based systems. In *2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, pages 01–06. IEEE, 2022. (page 9)

[94] Alber Fleischmann, Werner Schmidt, and Christian Stary. Complex event processing in e-services. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, pages 251–259. IEEE, 2016. (page 9)

[95] Kalimullah Lone and Shabir Ahmad Sofi. Cost efficient task offloading for delay sensitive applications in fog computing system. *SN Computer Science*, 4(6):817, 2023. (page 9)

[96] Zhou Jiang and Guo Jin. Scenario construction of aging-appropriate smart home based on information fusion of multi-sensor devices. In *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 1127–1132. IEEE, 2023. (page 9)

[97] Theodoros Nestoridis, Chrysa Oikonomou, Anastasios Temperekidis, Fotios Gioulekas, and Panagiotis Katsaros. Scalable iot architecture for balancing performance and security in mobile crowdsensing systems. In *2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, pages 1–8. IEEE, 2020. (page 9)

[98] H. Lanham, Luci K. Leykum, and J. Pugh. Examining the complexity of patient-outpatient care team secure message communication: Qualitative analysis. *Journal of Medical Internet Research*, 20, 2018. (page 9)

[99] Guimin Zhang, Brian L. Smith, and J. Guo. Peer-to-peer-based publish/subscribe architecture for advanced infrastructure systems. *Journal of Computing in Civil Engineering*, 24:65–72, 2010. (page 9)

[100] Chi-Sheng Shih, Wei-Yu Hsieh, and Chia-Lung Kao. Traceability for vehicular network real-time messaging based on blockchain technology. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 10(4):1–21, 2019. (page 9)

[101] Peng Wang, Kaiyuan Zhang, Rong Chen, Haibo Chen, and Haibing Guan. Replication-based fault-tolerance for large-scale graph processing. In *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 562–573. IEEE, 2014. (page 9)

[102] Seda Kul, Isabek Tashiev, Ali Şentaş, and Ahmet Sayar. Event-based microservices with apache kafka streams: A real-time vehicle detection system based on type, color, and speed attributes. *IEEE Access*, 9:83137–83148, 2021. (page 9)

[103] Junhua Fang, Pingfu Chao, Rong Zhang, and Xiaofang Zhou. Integrating workload balancing and fault tolerance in distributed stream processing system. *World Wide Web*, 22:2471–2496, 2019. (page 9)

[104] Ka Yung Cheng, Santiago Pazmino, and Björn Schreiweis. Etl processes for integrating healthcare data–tools and architecture patterns. In *pHealth 2022*, pages 151–156. IOS Press, 2022. (page 10)

[105] Neha Gupta and Sakshi Jolly. Enhancing data quality at etl stage of data warehousing. *International Journal of Data Warehousing and Mining (IJDWM)*, 17(1):74–91, 2021. (page 10)

[106] Suyeon Lee and Sungyong Park. Performance analysis of big data etl process over cpu-gpu heterogeneous architectures. In *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*, pages 42–47. IEEE, 2021. (page 10)

[107] Katja Seeliger, Luca Ambrogioni, Yagmur Güçlütürk, Leonieke M van den Bulk, Umut Güçlü, and Marcel AJ van Gerven. End-to-end neural system identification with neural information flow. *PLoS computational biology*, 17(2):e1008558, 2021. (page 10)

[108] Nevo Itzhak, Shahar Tal, Hadas Cohen, Osher Daniel, Roze Kopylov, and Robert Moskovitch. Classification of univariate time series via temporal abstraction and deep learning. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1260–1265. IEEE, 2022. (page 10)

[109] Ola Benderius, Christian Berger, and Krister Blanch. Are we ready for beyond-application high-volume data? the reeds robot perception benchmark dataset. *arXiv preprint arXiv:2109.08250*, 2021. (page 10)

[110] Gennadiy Sheptalin, Boris Ivanenko, Anton Sheptalin, and Yuliya Petrichenko. Digitalization of enterprise architecture as a means of ensuring the flexibility and adaptability of management information system. In *Informatics and Cybernetics in Intelligent Systems: Proceedings of 10th Computer Science On-line Conference 2021, Vol. 3*, pages 441–448. Springer, 2021. (page 10)

[111] Cong Minh Ho and Kyoung Kwan Ahn. Design of an adaptive fuzzy observer-based fault tolerant controller for pneumatic active suspension with displacement constraint. *IEEE Access*, 9:136346–136359, 2021. (page 10)

[112] Helmut Spengler, Ingrid Gatz, Florian Kohlmayer, Klaus A Kuhn, and Fabian Prasser. Improving data quality in medical research: a monitoring architecture for clinical and translational data warehouses. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 415–420. IEEE, 2020. (page 10)

[113] Azhar Talha Syed and Shikharesh Majumdar. Parallel processing techniques for analyzing large video files: a deep learning based approach. In *2022 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pages 270–279. IEEE, 2022. (page 10)

[114] Zakia Asad, Mohammad Asad Rehman Chaudhry, and David Malone. Greener data exchange in the cloud: A coding-based optimization for big data processing. *IEEE Journal on Selected Areas in Communications*, 34(5):1360–1377, 2016. (page 10)

[115] Sucharitha Shetty, B Dinesh Rao, and Srikanth Prabhu. Growth of relational model: Interdependence and complementary to big data. *International Journal of Electrical and Computer Engineering*, 11(2):1780, 2021. (page 10)

[116] Dianne SV Medeiros, Helio N Cunha Neto, Martin Andreoni Lopez, Luiz Claudio S. Magalhães, Natalia C Fernandes, Alex B Vieira, Edelberto F Silva, and Diogo M F. Mattos. A sur-

vey on data analysis on large-scale wireless networks: online stream processing, trends, and challenges. *Journal of Internet Services and Applications*, 11:1–48, 2020. (page 10)

[117] Yuxi Li, Liang Qiao, and Zhihan Lv. An optimized byzantine fault tolerance algorithm for consortium blockchain. *Peer-to-Peer Networking and Applications*, 14:2826–2839, 2021. (page 10)

[118] Bahaa Mahmoud Abdelhafiz and Mourad Elhadef. Sharding database for fault tolerance and scalability of data. In *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pages 17–24. IEEE, 2021. (page 10)

[119] Xixi Wang and Yepeng Guan. A hierarchy byzantine fault tolerance consensus protocol based on node reputation. *Sensors*, 22(15):5887, 2022. (page 10)

[120] Soohyeong Kim, Sejong Lee, Chiyoung Jeong, and Sunghyun Cho. Byzantine fault tolerance based multi-block consensus algorithm for throughput scalability. In *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–3. IEEE, 2020. (page 10)

[121] Johannes Mohr, Claudia Kleinschrodt, Stephan Tremmel, and Frank Rieg. Compatibility improvement of interrelated items in exchange files—a general method for supporting the data integrity of digital twins. *Applied Sciences*, 12(16):8099, 2022. (page 11)

[122] Miaochao Hu, Wei Guo, and Weisheng Hu. Dynamic scheduling algorithms for large file transfer on multi-user optical grid network based on efficiency and fairness. In *2009 Fifth International Conference on Networking and Services*, pages 493–498. IEEE, 2009. (page 11)

[123] Peipei Xu, Lianxiang Jiang, Bingui Xu, Mingxiang Li, and Fei Wang. Highly integrated modular avionics from platform to payload for micro-satellites. In *Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022)*, volume 12610, pages 78–83. SPIE, 2023. (page 11)

[124] Rohmat Tulloh, Jafaruddin Gusti Amri Ginting, Asep Mulyana, and Muhammad Lutfi. Performance comparison of file transfer protocol service between link state and distance vector routing protocol in software defined network. In *IOP Conference Series: Materials Science and Engineering*, volume 982, page 012026. IOP Publishing, 2020. (page 11)

[125] K Jaspin, Shirley Selvan, S Sahana, and G Thanmai. Efficient and secure file transfer in cloud through double encryption using aes and rsa algorithm. In *2021 international conference on emerging smart computing and informatics (ESCI)*, pages 791–796. IEEE, 2021. (page 11)

[126] Md Arifuzzaman and Engin Arslan. Use only what you need: Judicious parallelism for file transfers in high performance networks. In *Proceedings of the 37th International Conference on Supercomputing*, pages 122–132, 2023. (page 11)

[127] Dewei Yi, Petar Baltov, Yining Hua, Sam Philip, and Pradip Kumar Sharma. Compound scaling encoder-decoder (cosed) network for diabetic retinopathy related bio-marker detection. *IEEE journal of biomedical and health informatics*, 2023. (page 11)

[128] Hirokuni Kitahara, Kugamoorthy Gajananan, and Yuji Watanabe. Real-time container integrity monitoring for large-scale kubernetes cluster. *Journal of Information Processing*, 29:505–514, 2021. (page 11)

[129] Yunhao Mao. Skybridge: a cross-cloud storage system for sky computing. In *Proceedings of the 23rd International Middleware Conference Doctoral Symposium*, pages 15–17, 2022. (page 11)

[130] Jason P Sermeno and John Jowil D Orquia. Integrating fuzzy logic and dynamic programming in multithreaded concurrent file transfer schemes. In *2021 Second International Conference on Innovative Technology Convergence* (CITC), pages 56–62. IEEE, 2021. (page 11)

[131] K. Khurshid, Imdad Ullah, Z. Shah, Najm Hassan, and T. Ahanger. Protocols for transferring bulk data over internet: Current solutions and future challenges. *IEEE Access*, 9:95228–95249, 2021. (page 11)

[132] Javier Ordóñez, Angelos Alexopoulos, Konstantinos Koutras, Athanasios Kalogeras, Kyriakos Stefanidis, and Vanessa Martos. Blockchain in agriculture: A pestels analysis. *IEEE Access*, 2023. (page 11)

[133] Noble Anumbe, Clint Saidy, and Ramy Harik. A primer on the factories of the future. *Sensors*, 22(15):5834, 2022. (page 11)

[134] J Aythora, R Burke-Ag”uero, A Chamayou, S Clebsch, M Costa, J Deutscher, N Earnshaw, L Ellis, P England, C Fournet, et al. Multi-stakeholder media provenance management to counter synthetic media risks in news publishing. In *Proc. Intl. Broadcasting Convention (IBC)*, volume 1, page 8, 2020. (page 11)

[135] Auntre D Hamp, Rupali K Doshi, Garret R Lum, and Adam Allston. Cross-jurisdictional data exchange impact on the estimation of the hiv population living in the district of columbia: evaluation study. *JMIR Public Health and Surveillance*, 4(3):e9800, 2018. (page 11)

[136] Tamas Kiss. A cloud/hpc platform and marketplace for manufacturing smes. In *11th International Workshop on Science Gateways, IWSG 2019*, 2019. (page 12)

[137] Luca M Ghiringhelli, Christian Carbogno, Sergey Levchenko, Fawzi Mohamed, Georg Huhs, Martin Lüders, Micael Oliveira, and Matthias Scheffler. Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. *npj computational materials*, 3(1):46, 2017. (page 12)

[138] Martin Shapiro, Sondra Renly, Ali Maiorano, Jerry Young, Eli Medina, Aaron Neinstein, and Anobel Y Odisho. Digital health at enterprise scale: Evaluation framework for selecting patient-facing software in a digital-first health system. *JMIR Formative Research*, 7:e43009, 2023. (page 12)

[139] Kejie Wu. Design of intelligent management and control software for gas drainage in the whole mine based on cross platform architecture. In *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*, volume 7, pages 1606–1615. IEEE, 2023. (page 12)

[140] Mubashar Amjad, Gareth Taylor, Chun Sing Lai, Zhengwen Huang, and Maozhen Li. Scalability and reliability analysis of a novel cloud platform for tso-dso information and data exchange. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–5. IEEE, 2022. (page 12)

[141] Ramesh Venkatakrishnan. Design, implementation, and assessment of innovative data warehousing; extract, transformation, and load (etl); and online analytical processing (olap) in bi. *International Journal of Database Management Systems (IJDMS)*, 12(3), 2020. (page 12)

[142] Md Arifuzzaman, Brian Bockelman, James Basney, and Engin Arslan. Falcon: Fair and efficient online file transfer optimization. *IEEE Transactions on Parallel and Distributed Systems*, 2023. (page 12)

[143] Lang Lang, Zhenlong Zhu, Xuanye Liu, Jianxin Zhao, Jixing Xu, and Minghui Shan. Architecture and operation adaptive network for online recommendations. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3139–3149, 2021. (page 12)

[144] Hamdan AlSaadi, Faisal Rashid, Paulinus Bimastianto, Shreepad Khambete, Lucian Toader, Fernando Landaeta Rivas, Erwan Couzigou, Adel Al-Marzouqi, Hassan El-Masri, and Wiliem Pausin. Unlocking value from data is key to successful digital transformation. In *SPE/IADC Middle East Drilling Technology Conference and Exhibition*, page D021S010R001. SPE, 2021. (page 12)

[145] Arul Christhuraj Alphonse, Alexandra Martinez, and Akshata Sawant. *MuleSoft for Salesforce Developers: A practitioner's guide to deploying MuleSoft APIs and integrations for Salesforce enterprise solutions*. Packt Publishing Ltd, 2022. (page 14)

[146] Chao Zhang, Zichao Yang, Xiaodong He, and Li Deng. Multimodal intelligence: Representation learning, information fusion, and applications. *IEEE Journal of Selected Topics in Signal Processing*, 14(3):478–493, 2020. (page 14)

[147] Sudhir Rawat, Abhishek Narain, Sudhir Rawat, and Abhishek Narain. Introduction to azure data factory. *Understanding Azure Data Factory: Operationalizing Big Data and Advanced Analytics Solutions*, pages 13–56, 2019. (page 14)

[148] Shanshan Li, He Zhang, Zijia Jia, Chenxing Zhong, Cheng Zhang, Zhihao Shan, Jinfeng Shen, and Muhammad Ali Babar. Understanding and addressing quality attributes of microservices architecture: A systematic literature review. *Information and software technology*, 131:106449, 2021. (page 14)

[149] Paulo Henrique Da Silva, Fabiane Benitti, and Michelle Wangham. Framework for the development of computational solutions for the support of requirements engineering with a focus on data protection. In *Proceedings of the XXXVI Brazilian Symposium on Software Engineering*, pages 419–424, 2022. (page 14)

[150] Matthias J Sax. Performance optimizations and operator semantics for streaming data flow programs. 2020. (page 14)

[151] Murat Bakirci. A novel swarm unmanned aerial vehicle system: Incorporating autonomous flight, real-time object detection, and coordinated intelligence for enhanced performance. *Traitement du Signal*, 40(5), 2023. (page 14)

[152] Srijith, Karan Bantia R, Govardhan N, and A. R. Inter-service communication among microservices using kafka connect. *2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS)*, pages 43–47, 2022. (page 16)

[153] Antonio Quiña-Mera, Pablo Fernandez, José María García, and Antonio Ruiz-Cortés. Graphql: A systematic mapping study. *ACM Computing Surveys*, 55(10):1–35, 2023. (page 16)

[154] Wenjin Yu, Tharam Dillon, Fahed Mostafa, Wenny Rahayu, and Yuehua Liu. A global manufacturing big data ecosystem for fault detection in predictive maintenance. *IEEE Transactions on Industrial Informatics*, 16(1):183–192, 2019. (page 16)

[155] Valeriu Manuel Ionescu. The analysis of the performance of rabbitmq and activemq. In *2015 14th RoEduNet International Conference-Networking in Education and Research (RoEduNet NER)*, pages 132–137. IEEE, 2015. (page 16)

[156] Abhinay Pandya, Panos Kostakos, Hassan Mehmood, Marta Cortes, Ekaterina Gilman, Mourad Oussalah, and Susanna Pirttikangas. Privacy preserving sentiment analysis on multiple edge data streams with apache nifi. In *2019 European Intelligence and Security Informatics Conference (EISIC)*, pages 130–133. IEEE, 2019. (page 16)

[157] Shiva Prasad Paudel and Frank Schindler. Evolution of managed file transfer in business to business. *International Journal of Information Technology Applications (ITA)*, 9(2):35, 2020. (page 16)

[158] Morey J Haber and Darran Rolls. *Identity Attack Vectors: Implementing an Effective Identity and Access Management Solution*. Apress, 2019. (page 16)

[159] Simone Fischer-Hübner, Cristina Alcaraz, Afonso Ferreira, Carmen Fernandez-Gago, Javier Lopez, Evangelos Markatos, Lejla Islami, and Mahdi Akil. Stakeholder perspectives and requirements on cybersecurity in europe. *Journal of information security and applications*, 61:102916, 2021. (page 16)

[160] David Portugal, José N Faria, Marco Domingues, and Luís Gaspar. Integration of a smart bed infrastructure with hospital information systems using fast health interoperability resources:* a case study of the wireless biomonitoring stickers and smart bed architecture: towards untethered patients (wow) r&d project. In *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE, 2023. (page 16)

[161] Jutta Buschbom, Breda Zimkus, Andrew Bentley, Mariko Kageyama, Christopher Lyal, Dirk Neumann, Andra Waagmeester, and Alex Hardisty. Participative decision making and the sharing of benefits: Laws, ethics, and data protection for building extended global communities. *Biodiversity Information Science and Standards*, 5:e75168, 2021. (page 16)

[162] R Raja Singh, Swapnil Banerjee, R Manikandan, Ketan Kotecha, V Indragandhi, and Subramaniyaswamy Vairavasundaram. Intelligent iot wind emulation system based on real-time data fetching approach. *IEEE Access*, 10:78253–78267, 2022. (page 16)

[163] Mert Onuralp Gökalp, Kerem Kayabay, Mohamed Zaki, Altan Koçyiğit, P Erhan Eren, and Andy Neely. Open-source big data analytics architecture for businesses. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, pages 1–6. IEEE, 2019. (page 16)

[164] Adeline SE Chun, Aiden Gallagher, Amar A Shah, Callum Jackson, Claudio Tagliabue, Iliya Dimitrov, James Blackburn, Joel Gomez, Kim Clark, Lee Gavin, et al. *Accelerating Modernization with Agile Integration*. IBM Redbooks, 2020. (page 17)

[165] Alberto Miranda, Adrian Jackson, Tommaso Tocci, Iakovos Panourgias, and Ramon Nou. Norns: extending slurm to support data-driven workflows through asynchronous data staging. In *2019 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 1–12. IEEE, 2019. (page 17)

[166] T Sharvari and K Sowmya Nag. A study on modern messaging systems-kafka, rabbitmq and nats streaming. *CoRR abs/1912.03715*, 2019. (page 17)

[167] K Varalakshmi, V Dharma Prakash, and Deepika H Noble Lourdhu. Multipurpose file transfer and file inquiry. *Journal of Science, Computing and Engineering Research*, 6(4):90–96, 2023. (page 17)

[168] David Hawig, Chao Zhou, Sebastian Fuhrhop, Andre S Fialho, and Navin Ramachandran. Designing a distributed ledger technology system for interoperable and general data protection regulation–compliant health data exchange: A use case in blood glucose data. *Journal of medical Internet research*, 21(6):e13665, 2019. (page 17)

[169] Hanane Allioui and Youssef Mourdi. Exploring the full potentials of iot for better financial growth and stability: A comprehensive survey. *Sensors*, 23(19):8015, 2023. (pages 17 and 18)

[170] Hua Deng, Zheng Qin, Letian Sha, and H. Yin. A flexible privacy-preserving data sharing scheme in cloud-assisted iot. *IEEE Internet of Things Journal*, 7:11601–11611, 2020. (page 18)

[171] Chris Schlueter Langdon and Riyaz Sikora. Creating a data factory for data products. In *Smart Business: Technology and Data Enabled Innovative Business Models and Practices: 18th Workshop on e-Business, WeB 2019, Munich, Germany, December 14, 2019, Revised Selected Papers 18*, pages 43–55. Springer, 2020. (page 18)

[172] L. Tawalbeh, Fadi Muheidat, Mais Tawalbeh, and Muhannad Quwaider. Iot privacy and security: Challenges and solutions. *Applied Sciences*, 10:4102, 2020. (page 18)

[173] Ravi Kumar, Vikky Kumar, and C. Kumar. Impact of artificial intelligence, robotics, and automation on employment. *YMER Digital*, 2022. (page 18)

[174] Intidhar Essefi, Hanen Boussi Rahmouni, Tony Solomonides, and Mohamed Fethi Ladeb. Hipaa controlled patient information exchange and traceability in clinical processes. In *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, pages 452–460. IEEE, 2022. (page 18)

[175] Kashif Ahmad, Majdi Maabreh, Mohamed Ghaly, Khalil Khan, Junaid Qadir, and Ala Al-Fuqaha. Developing future human-centered smart cities: Critical analysis of smart city security, data management, and ethical challenges. *Computer Science Review*, 43:100452, 2022. (page 18)

[176] Karen Rose, Scott Eldridge, and Lyman Chapin. The internet of things: An overview. *The internet society (ISOC)*, 80:1–50, 2015. (page 18)

[177] Di Chang, Zhiming Wang, and Xia Zhang. Machine-learning based transportation network sparsification for iot trucking automation and optimization. In *2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, pages 933–937. IEEE, 2022. (page 18)

[178] Mélanie Bourassa Forcier, Hortense Gallois, Siobhan Mullan, and Yann Joly. Integrating artificial intelligence into health care through data access: can the gdpr act as a beacon for policymakers? *Journal of Law and the Biosciences*, 6(1):317–335, 2019. (page 18)

[179] Yeison Nolberto Cardona-Álvarez, Andrés Marino Álvarez-Meza, David Augusto Cárdenas-Peña, Germán Albeiro Castaño-Duque, and German Castellanos-Dominguez. A novel openbci framework for eeg-based neurophysiological experiments. *Sensors*, 23(7):3763, 2023. (page 24)

[180] Timescaledb app django app realtimedb documentation. https://timescaledb-app.readthedocs.io/en/latest/. (undefined 25/11/2023 19:20). (page 25)

[181] dunderlabpythondunerlabfoundation. https://github.com/dunderlab/python-dunerlab.foundation. (undefined 25/11/2023 19:26). (page 25)