Universiteit Leiden

Master ICT in Business and the Public Sector

Connecting across China and Europe with privacy preservation and regulatory compliance

Name:	Jiaxin Wang
Student ID:	s3638618
Date:	14/01/1999

Specialisation: ICT in Business

1st supervisor: Dr Mirjam van Reisen
2nd supervisor: Dr Jong Kon Chin, S. (Simcha)
PhD supervisor: Putu Hadi Purnama Jati

Master’s Thesis in ICT in Business and the Public Sector

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden

Acknowledgements

I would like to express my deepest gratitude to my first supervisor, Dr. Mirjam van Reisen, for her invaluable guidance, insightful feedback, and unwavering support throughout the development of this thesis. Her expertise and encouragement have been instrumental in shaping my research and improving my analytical approach.

I am also sincerely grateful to my second supervisor, Dr. Jong Kon Chin, S. (Simcha), whose thoughtful advice and constructive criticism have helped refine my work. His keen insights and attention to detail have significantly contributed to the depth and clarity of my research.

Furthermore, I would like to extend my heartfelt appreciation to my PhD supervisor, Putu Hadi Purnama Jati, for his generous support, technical guidance, and continuous encouragement. His expertise has played a crucial role in helping me navigate the complexities of my research topic.

Finally, I would like to thank my family, friends, and colleagues for their unwavering support and encouragement throughout this journey. Their belief in me has been a constant source of motivation.

Abstract

This thesis investigates the application of the FAIR principles (Findability, Accessibility, Interoperability, Reusability) within a framework compliant with both China's Personal Information Protection Law (PIPL) and the European Union's General Data Protection Regulation (GDPR). By extending FAIR with the FAIR-OLR framework (Ownership, Localization, Regulatory Compliance), the study explores whether cross-border data visiting – a federated analysis paradigm enabling data query and utilization of distributed data stores without physical cross-border transfers – can reconcile PIPL's data localization mandates with GDPR's transfer requirements. Through the implementation of tools such as CEDAR templates, BioPortal ontologies, and AllegroGraph triple storage platform, the proposed architecture ensures machine-actionable metadata and semantic interoperability while maintaining complete compliance with both regulations. To validate the framework, key metrics based on FAIR-OLR principles were developed, including data residency compliance, query accuracy, and regulatory adherence. Testing results demonstrate the feasibility of secure and compliant cross-border data exchange, offering a technical pathway for transnational data collaboration in regulated sectors.

Key words: FAIR principle, FAIR-OLR, Data localisation, Data ownership, PIPL, GDPR, Cross-border data.

Contents

Table of Contents

1 Introduction

1.1. Problem Statement

With the global rise in smart device usage and the growing investment in digitalization since the late 2000s, data has increasingly become a vital element in the digital economy (Kravchenko et al., 2019). This shift is evident in the rise of new technologies like the Internet of Things and Artificial Intelligence (AI). Cross-border data flows are now essential to this modern economy, supporting 65% of the world's Gross Domestic Product (GDP) in 2022 (Bielialov et al., 2023). Additionally, cross-border e-commerce has grown dramatically, expanding 45 times in size over the 10 years leading up to 2023, reaching an estimated US$2.7 trillion (Bielialov et al., 2023).

On the one hand, the development of these technologies greatly facilitates People's Daily life, such as the application of machine learning and big data analysis, which can easily record a person's behavior habits on the Internet and use these data to improve people's experience on the Internet (Kravchenko et al., 2019). As a result, data has been gradually regarded as one of the valuable assets that fuel innovation, economic activity, and international collaboration (Ikram, 2024).

On the other hand, the global exchange of data across borders, driven by e-commerce and innovation, has raised significant concerns about national security and privacy, especially after cases of data misuse by governments for intelligence purposes (e.g. the Snowden case), and by tech companies for profit (e.g. the Facebook data scandal) in the 2010s (Bansal & Warkentin, 2021). Due to these concerns, more countries have started to regulate cross-border data transfers, where the number nearly doubled to 62 by 2021 over four years. For instance, countries like Germany and Japan now only permit data to be sent to nations with similar data protection standards. Besides, nations like India and Russia now have set requirements for local laws to conduct security assessments before data can be transferred out of the country. They also must have their data stored and processed on national soil. One should note that this concern becomes acutely pronounced when personal data flows across boundaries, suddenly operating in a different legal environment (Neto et al., 2021).

A key factor contributing to this issue is lack of a consistent standard on data protection laws among different countries. Such fragmentation creates uncertainty and potential risks. For example, the cross-border data sharing between China and EU countries, in the EU, the GDPR was established as the only legal framework, beginning from May 25, 2018. The aim of the GDPR is to establish protection of the privacy and personal data of individuals living in the EU; the regulations apply to any organization outside of the EU that processes data concerning EU citizens (Bakare et al., 2024). Widely known worldwide for its strict requirements extending to every aspect of data processing -from collection, storage, usage, and sharing- GDPR is seen as a great principle of law. Most importantly, personal data must be processed in a lawful, fair, and transparent manner (Politou et al., 2018). The GDPR provides individual rights on the data themselves, including the rights to access, rectify, and delete that data, as well as the right to data portability. In terms of non-compliance, for companies that do not abide by the GDPR could be subject to some serious fines-up to 4% of a company's global annual revenue or €20 million, whichever is higher. Moreover, countries influenced by these guidelines for data transfers across borders are beginning to change and develop their laws in order to continue exchanging data with the European Union. (Zaeem & Barber, 2020).

Because the GDPR applies not only to companies within the EU but also to those outside the EU that process the data of EU residents, many global companies are required to comply with its rules. As a result, the regulation has had a deep global impact, influencing data protection laws in other regions (Peloquin et al., 2020). Countries like Japan, South Korea, and China have introduced similar laws to align with these high standards. This trend is often referred to the "Brussels Effect" (Bradford, 2020). "Brussels Effect" refers to the large influence of the European Union's internal market, which pushes companies and countries around the world to adopt laws and practices that match EU standards (Christen et al., n.d.). Due to the "Brussels Effect", the GDPR also raises global awareness about privacy rights, pushing governments of different countries to improve their data protection laws.

As for China, China’s Personal Information Protection Law (PIPL) was established and became effective on November 1, 2021. The law applies to companies and organizations that collect, store, use, or share personal information of Chinese citizens, both within China and internationally (Torrisi, 2023). Here are some important principles of PIPL: PIPL requires that personal information must be collected and used only for specific and legal reasons. Companies need to get permission from people before using their personal information, except in some cases like emergencies or public interest (Greenleaf, 2021). Second, PIPL gives people rights over their personal information. People can ask to see their information, correct it if it’s wrong, or delete it if it’s no longer needed. They can also withdraw their permission at any time. Third, PIPL has strict rules for sending personal information outside of China. Companies must make sure that the information will be safe in the other country, or they might not be allowed to send it (Greenleaf, 2021).

According to the "Brussels Effect" and the rapid development of China's local Internet industry, PIPL has some similarities with the GDPR (Bradford, 2020). Meanwhile, the differences in culture, politics, development environment and other factors between China and Europe Union, have also led to some regulations and rules in PIPL that are more suitable for Chinese local environment, which are different from GDPR (Almada & Radu, 2024). The similarities and differences between The GDPR and China's PIPL illustrate the feasibilities, complexities and difficulties of information and cross-border data exchange in a globalized world (Christen et al., n.d.).

Yet, because of the differences between the two data protection regulations, organizations operating cross-border travel within the EU and China have very many hurdles to cross. The difference therefore brings the question: how can organizations comply not only with these competing legal standards, but also manage their data in an efficient, transparent, and equitable manner? To solve this problem, the current main methods used by companies are Standard Contractual Clauses (SCCs), data localization, and distributed storage.

Standard Contractual Clauses (SCCs) are legal tools given by the European Commission, designed to secure legality and safety while transferring data across borders, especially where data is sent from the European Union to countries with inadequate protection of data(Rzayeva, 2024). However, this approach can at best keep data leakage down to a minimum but anyway requires some type of oversight by the company.

Data Localization and Distributed Storage are strategies through which multinational companies store and process data to ensure different compliances. Data localization means that for data that is generated in a given territory, it remains strung within that territory. Such a design usually is accompanied by distributed storage, allowing data to be spread and orchestrated over several locations to control risk and, therefore, legal complexity (Selby, 2017, p. 5). However, traditional localization regimes, while attempting to enhance national security and data sovereignty, mean severe backlashes. These policies lead to a fragmented data governance since data kept within national borders inhibits cross-border collaboration and innovation. The mandate for local storage raises operational costs exponentially for multinational companies forced to set up and maintain expensive local data centers to comply with assorted legal and regulatory demands.

Given the existing solutions' limitations and problems, this article will address these very critical and convoluted issues offering some solutions aligned to the FAIR principles. The FAIR principles-Findability, Accessibility, Interoperability, and

Reusability-provide a strong conceptual base for storage, discovery, sharing, and reusability of data in modern data ecosystems.

In summation, the growing rate of adoption of smart devices and digitalization has rendered data a quintessential component of the global economy. Technologies such as the Internet of Things and Artificial Intelligence fruitfully rely on cross-border data flows, which became key determinants of global GDP growth. But the increasing importance of data ushered in a whole range of different challenges, particularly in terms of privacy and national security. The events of the Snowden revelations and the misuse of data by Facebook present to us the risks of data misuse. This has led several countries to clamp rules over cross-border data transfers. They may include countries like Germany and Japan, which allow data sharing only to those nations with equivalent levels of data privacy standards. At the other extreme, you have countries like China and Russia, which require previous approval from authorities before any data transfer and mandate that data be stored on their territory. These complex regulations pose serious challenges for international business, especially between the EU and China. Companies face the insidious task of juggling multiple disparate laws while ensuring effective data management. Such standard solutions as SCCs and Data Localization are often employed, although they are not without limitations. Therefore, this paper muses how the FAIR principles may provide a superior framework for the management of data in a legally convoluted milieu.

1.2. Research gap

Even if SCCs and Data Localization are seen as central mechanisms for the handling of cross-border data transfers, both have manifest limitations that have not been fully explored by current research. First and foremost, while SCCs do establish a mandate for data transfers once the parties enter into the contract, they require enormous compliance obligations to companies(Rzayeva, 2024). In short, it cannot be ignored that SCCs may not entirely take care of the risks of moving data to countries with lesser data privacy laws. This creates a crucial gap: SCCs confer certain legal protections but do not fully ensure nor mitigate the likelihood of a data breach or the misuse of information.

In this sense, Data Localization-the requirement for data across borders to be stored within the country of origin-is problematic here as well. While this helps in securing the data by limiting foreign entities' unauthorized access, it's also incredibly inefficient in terms of operations. It can increase the cost of doing business for multinational companies, forced to create their own data centers in each country that not only put a dent in the company's finances but also create very demanding technical challenges. Besides, Data Localization constrains data within national borders and a setback against cross-border data flow looms large as it restricts the scope of global data sharing and interoperability. Data compartmentalization hinders global scaling-up of transformative tech such as artificial intelligence and big data analytics thereby diluting the power of organizations to take advantage of big data and eroding their competitive advantage.

Unlike traditional methods that focus solely on legal compliance, we can see VODAN-Africa team has successfully implemented FAIR principles to facilitate federated data analysis across multiple jurisdictions, particularly in the context of healthcare data (Purnama Jati et al., 2022). However, the VODAN-Africa team has currently mainly explored some situations in Africa and the EU and has mainly focused on the healthcare data field. Some architectures provided by the VODAN-Africa team have offered a lot of inspiration for this article. This article shifts its focus from Africa and the EU to China and the EU, specifically to the framework established by China's PIPL and EU’s GDPR. Besides, this thesis also attempts to expand the data field to a wider range.

Existing research on the GDPR and PIPL has extensively covered the legal frameworks and compliance requirements of these regulations. However, there is still a gap in exploring how companies may effectively manage and utilize data across varying legal environments without sacrificing efficiency, transparency, or the potential for innovation. Current solutions such as SCCs and Data

Localization are more about compliance than about the actual engaging in effective data management and similar practices.

This evidence gaps in research, particularly with regards to subsequent proper action of data management on one hand and general comparison of PIPL with GDPR on the other hand. The FAIR principles appear as a suitable framework in this context. Unlike traditional methods that primarily focus on legal compliance, the FAIR principles provide a comprehensive approach that prioritizes the quality, usability, and sustainability of data. Meanwhile, the introduction of FAIR-OLR principles also provides a different solution for this situation. By adopting these principles, organizations can achieve a higher standard of data management that supports global collaboration in a legally complex world.

1.3. Research questions

In this paper, my main research question is:

How can the FAIR principles be applied to develop effective data management strategies that bridge the differences between China's Personal Information Protection Law (PIPL) and the European Union's General Data Protection Regulation (GDPR)?

The questions that we use to ultimately answer the research question and achieve the research objectives:

- How do PIPL and GDPR deal with the regulations of cross-border data transfers, particularly in balancing data privacy and national security?

  - What are the main differences between the PIPL and GDPR regulations regarding the handling and protection of sensitive data?

- How can Chinese Data Policies enable FAIR Principles and what are the impacts of FAIR Principles implemented in China?

  - How can the FAIR principles be adjusted to meet the challenges of PIPL's strict data localization rules and GDPR's cross-border data transfer requirements?
  - What strategies and technologies can be implemented within a FAIR-based framework to ensure compliance with both PIPL and GDPR?

- What are the key metrics to assess the architecture's performance?

1.4. The hypothesis

Applying the FAIR principles to data management can effectively bridge the regulatory gaps between China's PIPL and the EU's GDPR. By customizing these principles to deal with specific challenges like data localization and cross-border transfers, a FAIR-based framework can improve data accessibility, interoperability, and compliance in both regions. This is expected to be more efficient than traditional approaches such as SCCs and data localization while offering a more robust solution for the transfer of cross-border data across these complex legal frameworks.

1.5. Conceptual framework

In particular, challenges posed by cross-border data governance include fragmented legal frameworks, technical incompatibilities, and culture-driven differences in data management. In this context, the FAIR principles offer a strong conceptual lens through which to look at data-sharing challenges encountered internationally. Next to using just one standard for sharing a single access data set, FAIR will allow the easy discovery and access of datasets with vastly different legal contexts and cultures, simplifying compliance with multiple regulatory regimes-and interoperability supports data to be integrated into various systems and jurisdictions to bolster global cooperation and innovation(Lamprecht et al., 2020).

The FAIR principles provide a blend of high-level framework to set about a growing array of problems arising in data management and use in the digital era. First introduced at the Lorentz Centre in 2014, these principles were aimed at addressing the growing demand for machine-actionable data reuse and interoperability across various systems and platforms(Stocker et al., 2022).These principles were formulated in 2016 and have since gained worldwide acclaim for their effectiveness in guiding the organization, sharing, and sustainable reuse of digital resources.

FAIR principles can be found flexible guidance, as opposed to rigid standards. It can be adopted according to the needs of different stakeholders, including researchers, policymakers, and technology developers(Mons et al., 2020). Through the flexibility this paradigm presents, it becomes possible for FAIR to mitigate variability in data discoverability, access, and reuse amidst the vast spectrum of disciplines and jurisdictions. For example, within cross-border contexts, FAIR facilitates the assurance of compliance with such regulations as GDPR and PIPL by virtue of forming a common standard for metadata and improving interoperability(Jacobsen, de Miranda Azevedo, et al., 2020).

As data ecosystems become more intricate, the FAIR principles remain the bedrock for secure and efficient data sharing. Going forward, FAIR will be a clear anchor for trends in federated learning and decentralized networks to align itself with future technological and regulatory bottlenecks.

In the next sections, we shall further elaborate on the specific contents of the FAIR principles:

1.5.1. Findability

Discoverability is the basis for FAIR principles; this helps realize that the first step toward effective data management and reuse ought to be finding data. This idea was birthed to counter the increasing challenge of finding relevant datasets in the expanding digital ecosystem. As pointed out by Wilkinson et al. (2016), data should be accessible for both machine and human access, as machines provide increasing aid for data-driven research and automation. In order to do this, effective indexing, organization, and querying processes must be documented; hence the affordance of discoverability. Here, the use of machine-readable metadata and structured knowledge graphs defined by standardized ontologies allows easy navigation to locate any given dataset(Boeckhout et al., 2018). The use of persistent and globally unique identifiers ensures, aside from the others, that data remain reliably accessible through time, creating the basis for the subsequent principles of Access, Interoperability and Reusability.

For an even deeper understanding, to achieve findable data, Wilkinson et al. (2016) illustrated four principles:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

1.5.2. Accessibility

This is the second important pillar of the FAIR framework: Accessibility to guarantee that users may use the data with the caveat of complying with clearly articulated and well-defined conditions. Accessibility does not automatically imply openness; the mechanisms for requesting data must be clearly laid out, even behind restricted access, according to Wilkinson et al. (2016). This principle narrates the realization between discoverability and usability and emphasizes the need for standardized communication protocols that support human and machine access. Accessibility ensures secure retrieval of sensitive or proprietary data and upholds ethical, legal, and organizational obligations through well-structured authorization and authentication processes (Jacobsen, de Miranda

Azevedo, et al., 2020). The combination of these dual, opposite tenets of openness and control should strike a balance in which users can safely share data while adhering to a variety of regulations.

Basically, to be able to permit data accessibility, Wilkinson et al. (2016) identified four principles:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1. the protocol is open, free, and universally implementable

A1.2. the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

1.5.3. Interoperability

Interoperability is central to data integration as well as collaboration across disciplines, systems, and geographies. The FAIR principles emphasize the importance of data resources that allow for seamless communication and connection, promoting effective interaction of data with other datasets and computational tools(Mons et al., 2017). According to Wilkinson et al. (2016), for interoperability to take place, common vocabularies, formal ontologies, and standardized machine-readable and universally usable knowledge representation languages should be deployed. This allows different systems to be integrated and information processed seamlessly. The interoperability is very relevant in transnational data settings where there are different regulatory, cultural, and technical environments that force adoption of a common framework(Jacobsen, de Miranda Azevedo, et al., 2020).

To be more specific, to achieve data interoperability Wilkinson et al. (2016) has illustrated three principles:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

1.5.4. Reusability

Finally, reusability is at the end of the FAIR principles. This represents a desire to maximize data resources' values and impacts. Reusability goes beyond just being able to access data, as noted by Wilkinson and others (2016). It deals primarily with making sure that datasets are well documented, properly licensed, and equipped with rich metadata to be potentially integrated in different contexts. This principle especially serves as a foundation for reproducibility and accountability in scientific research, in that it ensures that datasets can then be presented with clarity so that they may again be used for new questions or new studies. Detailed provenance information, adherence to domain-specific community standards, and clarity in usage licenses are key to attaining reusability(Wilkinson et al., 2016).Reusability within the global data ecosystem promotes collaborative efforts while minimizing any duplication, enabling stakeholders to progress based on pre-existing resources rather than commencing from scratch(Zhang, Downs, & Li, 2019). Reusability reinvents data from a mere, static commodity to a progressive and dynamic catalyst in development and innovation spanning various arenas.

To be more specific, to achieve data Reusability Wilkinson et al. (2016) has illustrated three principles:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

1.6. Ethical and legal considerations

The research involves processing sensitive personal data and transferring it overseas, which calls for proper attention to ethical and legal considerations. Since this study will entail developing a data management framework that= may navigate differences in regulations between PIPL of China and GDPR of the EU, compliant with both legal frameworks.

Data Privacy and Protection

The scope of this study is mainly centered around the protection of personal data. It seeks to ensure that, under the GDPR, the subjects' fundamental rights are respected, such as the right of access, the right to correction, and the right to deletion. In the same vein, the PIPL requires personal data processing to prioritize privacy and civil liberties, especially during the transfer of data across national borders. The study will ensure all the data processing complies with the data protection principles found within both the GDPR and PIPL, with special attention being given to the transparency of data processing and obtaining informed consent wherever necessary.

Legitimate Data Processing

The processing of data in this study will be based on legal grounds clearly captured under both GDPR and PIPL. Such could include acquiring explicit consent from the data subject, processing data as may be needed for compliance with contract reasons, interests which don't hinder the rights and freedoms of individuals. Further, this research will have protective measures on sensitive data in order to handle it according to legal standards and ethical norms.

Cross-border Data Transfers

Given the international scope of this research, the ethical implications of cross-border data transfers are of new critical importance. In this research, questions related to data sovereignty, fair treatment of data subjects in various legal systems, and Review of transfer conditions subject to stringent controls under the GDPR and PIPL ensuring requisite security assessments and appropriate protective safeguards will be made.

1.7. Relevance

Academic Relevance

This study appeals to the academic sphere to fill a major gap in the current literature on cross-border data management and legal compliance. While there has been abundant research on both the EU GDPR and China's PIPL, less has been said so far on direct comparisons of these two schemes, or on the development of data management strategies for steering data through the differences. The application of the FAIR Principles permits this research to present an approach that will provide valuable instruction regarding data management strategies in different jurisdictions. By combining legal requirements with advanced data management frameworks, the research will contribute to closing the gap between the legal and data science communities with a new perspective into data governance.

Social Relevance

Since data are viewed as an asset, it assumes the highest level of importance to ensure proper management of data. Findings of this research are relevant to society since they provide guidance for companies that could ease their compliance with legislations such as GDPR and PIPL, strengthening international trade and collaboration. With the proposal of a data governance framework that guarantees legal compliance without compromising data accessibility and usability, the study presents a facility for multinational companies in potential overlapping legal jurisdictions. In addition to its preference of data protection, the study also contributes to the larger societal goal of protecting personal privacy and assures public confidence in digital technologies and global data flows.

1.8. Research approach

This research is pragmatic, focusing on the development and evaluation of a data management framework that closes the regulatory gap between the PIPL of China and the GDPR of the European Union. Pragmatism would fit this problem perfectly: focus on problem-solving and practical outcomes because this is aimed at resolving real-world difficulties with cross-border data management.

In the case of this research, pragmatism provides for iteratively-developed solutions based on theoretical insights and practical insights. Pragmatism is in recognition of the complexities of steering a course through these two legal systems and deploying principles such as FAIR (Findable, Accessible, Interoperable, Reusable) and FAIR-OLR (Ownership, Localization, Regulatory) toward addressing these complexities.

As a researcher, I am both an analyst and an interaction: I analyze the existing literature, which captures the inner workings of the PIPL and the GDPR and apply the FAIR principles to build a uniform framework. Recognizing the shortcomings of theoretical validation, I seek to assure that the proposed framework stays malleable enough to withstand real-world situations while carrying value to the academic and business environments.

1.9. Research design

This research uses a Design Study approach to develop a data management framework to bridge the regulatory differences between China's PIPL and the European Union's GDPR.

The choice of Design Study as the approach was made due to its strong suitability for addressing complex, practice-oriented research questions. Given the unique challenges arising from the differences between PIPL and GDPR, a design-driven methodology allows for creating a solution that is theoretically sound and practically viable. The iterative Design Study allows for the continuous improvement of the framework and an epitome of relevance throughout the process as new learnings are obtained through testing and expert feedback.

1.10. Research objectives

This research primarily aims to discover data management strategies to manage the conflicts posed by the differences between China's PIPL and the European Union's GDPR.

To achieve this, the study will focus on the following specific objectives:

- The first part of the project aims at conducting an in-depth analysis of PIPL and GDPR's core regulations. This encompasses an assessment of the key obligations on personal data protection, cross-border data transfer, and obligations to comply.
    - In comparing these two legal frameworks, the other major objective is to see if their comparison in terms of data processing impacts on individual rights and corporate compliance are similar or different.
- The second goal will be to identify possible applications of the FAIR principles in an overall framework of data management to tackle the challenges of the PIPL and the GDPR.
- The goal is to develop a framework that allows not just for compliance but also for the more efficient and effective use of the data.

- The third objective is to test and implement a proof of concept based on the FAIR principles.
    - The goal is to evaluate how this framework performs, especially in transferring cross-border data under the legal constraints of PIPL and GDPR.

2. Theoretical framework

Rigorously understanding the theoretical frameworks that will underpin the analysis will provide an important basis for this study. The research looks at data protection frameworks represented mainly by the EU's GDPR and China's PIPL, framing context through the lens of the FAIR Principles. Balancing regulatory compliance with responsible data stewardship, this framework paves the way for focused rumination on how cross-border data governance can be optimized across a spectrum of legal and operational contexts.

2.1. The hourglass model

The hourglass model: which serves as a conceptual framework for standardizing interoperability in the Internet of FAIR Data and Services (IFDS). Influenced by the success of the present Internet IP/TCP protocol, the model points for the necessity of a single common definition for the central protocol-FAIR Digital Objects (FDOs)-to seamlessly integrate data across platforms and applications. The hourglass model indeed reflects a balance between flexibility and standardization(Cardoso Silva Ferreira & Van Reisen, 2023). The very start of the hourglass is where the raw data go through FAIRification, obtaining a structured actionable unit, and the bottom is advanced analytics, integration, and sharing; the narrowest part of the hourglass embodies the necessary standardization by which such diverse systems talk to each other.

The Hourglass Model provides interoperability at the level of Fair Data Objects so that any data can be reused, shared, or discovered, irrespective of the domain or technological context. For instance, in the case of VODAN-Africa, the nanopublication acted as an approximation to FDOs, fostering the open data sovereignty between hospitals across Africa(Cardoso Silva Ferreira & Van Reisen, 2023). These efforts demonstrate the model's potential to transform data governance and analytics by creating a unified, FAIR-compliant ecosystem. The hourglass model is described in Figure 1 below.

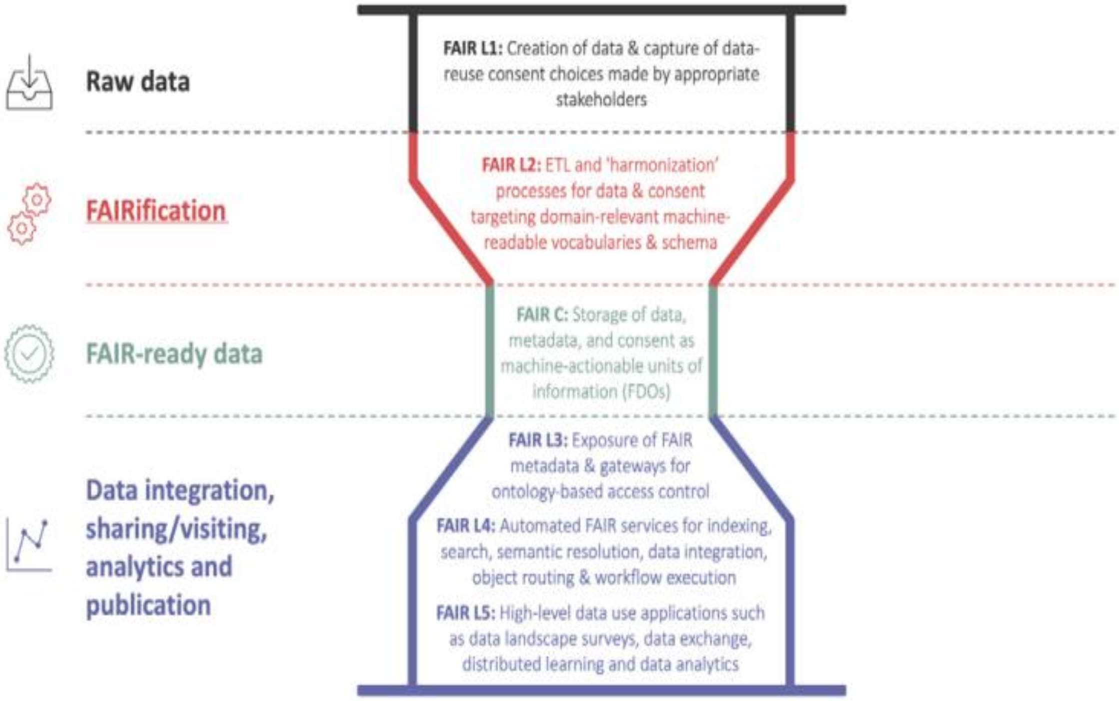

Raw data
FAIR L1: Creation of data & capture of data-reuse consent choices made by appropriate stakeholders
FAIRification
FAIR L2: ETL and 'harmonization' processes for data & consent targeting domain-relevant machine-readable vocabularies & schema
FAIR-ready data
FAIR C: Storage of data, metadata, and consent as machine-actionable units of information (FDOs)
Data integration, sharing/visiting, analytics and publication
FAIR L3: Exposure of FAIR metadata & gateways for ontology-based access control
FAIR L4: Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution
FAIR L5: High-level data use applications such as data landscape surveys, data exchange, distributed learning and data analytics

Figure 1 The hourglass model Illustrates the transformation of raw data into FAIR Digital Objects (Cardoso Silva Ferreira & Van Reisen, 2023).

2.2. Data sovereignty

Data sovereignty relates to the rights and authority that owners or organizations wield over decisions that involve the storage, access, and utilization of their data. In the current day of flowing information across borders, this concept is vital to ensure compliance by means of the diverse regulatory frameworks for data protection, e.g., the General Data Protection Regulation of the European Union and the Personal Information Protection Law of China. While protecting data privacy, data sovereignty empowers different stakeholders to dictate how their data is shared and utilized.

FAIR principles, especially Accessibility and Reusability, provide a structured way of dealing with issues in data sovereignty. Through the establishment of clearly defined access conditions, implementation of robust authentication mechanisms and an emphasis on metadata transparency, the FAIR principles support data sovereignty while, at the same time, enhancing global collaborations (Wilkinson et al., 2016). Controlled-access models, for instance, allow data providers to maintain ownership while authorizing users to access the data for legitimate purposes; this lecturing is, thus, between the conflict of privacy and data sharing.

In fact, data sovereignty requires careful consideration of balancing privacy and compliance. Integration of these principles will ensure that FAIR data remains findable, accessible, and interchangeable with full protection of the data subjects' rights and respect for local laws (Boeckhout et al., 2018).

2.3. FAIR Data Point (FDP)

FAIR Data Points underpin the implementation of the FAIR principles through the decentralized management and access of data. These data repositories will be designed to conform to the FAIR principles by having machine-readability in metadata and structured data storage. (da Silva Santos et al., 2023). FDFS allow local control over data governance while simultaneously enabling global data collaboration, with compliance with various legal and ethical frameworks.

FDP consists of metadata describing datasets and an interface for discovery, access, and analysis. In practice, FDPs empower data ownership at the source, following the principle of data sovereignty. This mechanism ensures that data remains with its producers while allowing its use for legitimate research purposes. An appropriate example would be from the VODAN- Africa project, where FDPs have been successfully implemented in several countries across Africa to support data collection and analytics related to COVID-19(van Reisen et al., 2021). These implementations showed FDPs' potential to tackle regional problems in data sharing, such as cross-border interoperability and compliance with jurisdiction-specific regulations.

By bringing FDPs into the fold of data management strategies, organizations can obtain real-time data analytics and use this for collaborative research(Wijnbergen et al., 2024). The architecture of FDPs supports both clinical and research data for seamless aggregation and analysis under a highly secured and privatised environment. This feature makes FDPs a bridging tool between local data governance and global research initiatives.

2.4. Federated analyses

Generative analyses present a manner of decentralizing data processing without compromising the data sovereignty or connecting it with local laws. Also called "data visiting," federated analyses has enabled organizations to extract insights from sensitive or proprietary data without compromising on privacy or ownership (Casaletto et al., 2023). To work in practice, this approach sends analytical algorithms toward the data rather than centralizing the data itself. In deploying machine-actionable metadata via FAIRE Data Points (FDP), federated analyses provide alternatives for secure and efficient collaboration between stakeholders by minimizing risks concerning data transfer and storage. The main benefits of federated analyses are, but are not limited to, those of strong security, compliance with diverse legal environments, and differing datasets integration.

2.5. Federated learning and AI

Federated learning is changing the landscape of artificial intelligence by enabling collaborative machine learning among decentralized datasets while ensuring data sovereignty and privacy. Unlike conventional centralized learning methods, federated learning permits algorithms to learn locally on distributed data sources(Van Reisen et al., 2021). Assessing this aspect will allow for concerns regarding data transfer and data privacy to be minimized, or rather answered, and would allow the use of the proper channels of communication with respect to the aforementioned principle among others, such as under the GDPR or PIPL."

In the context of the FAIR principles, federated learning supports the goals of accessibility and reusability through the ability of AI models to learn from different datasets across various jurisdictions(Nguyen et al., 2021). Federated learning facilitates machine-actionable data processing via FAIR-compliant metadata and infrastructure, streamlining their integration and analysis among different environments.

As artificial intelligence evolves, federated learning, along with FAIR principles, presents the future organization of ethical, transparent, and efficient data innovation. This partnership addresses critical challenges in cross-border data governance while maximizing the potential of AI in solving complex global problems.

2.6. FAIR-OLR

The principles of FAIR-OLR build on the foundation of FAIR by adding three vital aspects: ownership, localization, and regulatory compliance. This also targets the challenges of governing sensitive and personal data framed in specific jurisdiction contexts to uphold data sovereignty and global interoperability. The principles are developed to have data under the control of its producer or subject, for local storage at the point of generation, under a legal context. These three principles add to trust, transparency, and ethical stewardship in federated data architectures. This approach enhances the compliance requirements outlined in legislation, such as the GDPR, while creating a sustainable avenue for data reuse across borders and domains(Van Reisen et al., 2023).

The FAIR-OLR creates a model expanding on the FAIR principles by introducing essential considerations for data management:

Ownership: In determining the data management principles for the FAIR-OLR process, ownership is the pillar to be considered for the principle, keeping in place data control with the organizations or people that created such data or those who are represented. An example of the enactment of this principle is found in the VODAN-Africa project, which supports data "in residence" governed by data use agreements, subject to a changing set of laws and regulations in every jurisdiction (Van Reisen et al., 2023). The principle of ownership also extends to aggregated data, which, once anonymized and stripped of personal identifiers, can be utilized for broader research and statistical purposes (Van Reisen et al., 2023). Instead, it can guarantee personal privacy protection with a good business intelligence benefit on a wide scale.

Localization ensures that data physically and administratively remain located in the jurisdiction from where it is generated or collected. It ensures compliance with the local regulatory framework and allows the government to control data management. Practically, localization refers to data storage in local repositories such as AllegroGraph or other semantic data storage solutions while achieving both horizontal and vertical interoperability(Van Reisen et al., 2023). Data localization is particularly relevant in jurisdictions like China, where PIPL mandates that certain data must be stored within the country.

Regulatory Compliance: Compliance is the third component of the FAIR-OLR framework, which seeks to ensure that data handling processes adhere to the legal and ethical practices in the jurisdictions where they are produced. Compliance entails keen acquaintance with differing regulatory frameworks, including GDPR in the EU and other such policies in the non-EU areas.

For this analysis, there are strict rules laid forward for both GDPR and PIPL that must be met when processing and transferring data. Federated data architectures allow overregulation by ensuring that data can remain within the borders of its jurisdiction while still being available for authorized uses through federated learning and data visiting models. Such a setup collaborates with the principle of assigning the decision-making process to the data subjects while guaranteeing accountability in data processing through different regulatory environments (Van Reisen et al., 2023).

3. Research methodology

This section outlines the research methodology used in this study, which focuses on developing a data management framework that bridges the regulatory differences between China's PIPL and the European Union's GDPR. The chosen methods are designed to ensure that the study's objectives are met efficiently and effectively.

3.1. Research Design

The array of research work uses the Design Study approach to finding realistic solutions to complex problems. Iterations of design and testing of data management frameworks based on FAIR principles and FAIR-OLR principles are normally adopted into this framework. Such process involves defining the problem, developing the framework, prototyping and testing.

3.2. Data Collection

Data Sources: The primary data sources for this research are academic journals, legal commentaries, official regulatory documents, and policy papers, which will be accessed from Google Scholar and other through academic databases.

Data Collection Methods: The information used in this brief will be sourced from literature review. The literature review will provide guidance of a systematic kind starting from a broad searching for relevant literature on both PIPL and GDPR. The search shall be centered on data protection issues related to privacy laws, cross-border data transfer, data localization, and regulatory compliance. After that, the above literature, qualified according to relevance, can be applied for filtering and further classification in relation to certain clauses, articles, and principles within PIPL and GDPR which define data processing, storage, and transfer.

3.3. Methodology for Research Finding 1

In order to conduct a comparative legal analysis and a document analysis of these regulatory frameworks, focusing on balancing data privacy with national security interests, the research methodology will investigate how the PIPL and the GDPR regulate cross-border data transfer.

The first part consisted of a thorough legal framework analysis of PIPL and GDPR. This involved a thorough examination of the basic provisions and highlights of both sets of regulations, focussing especially on the issue of cross-border data transfer.

Following that was the comparative legal analysis, identifying salient differences between the two frameworks. This step was central to understanding the major regulatory differences between PIPL and GDPR, specifically with respect to their perspectives on data privacy and national security. While PIPL emphasizes data sovereignty and exercises harsher controls over data, GDPR offers easier traversability for cross-border data transfers.

At last, it delved into analyzing the crux of the divergence between PIPL and GDPR in ways of export-oriented data flows and national security. This provided important insights that could pave the way toward developing FAIR-OLR-based architecture for both regulators.

On the basis of such methodology, the research identified regulatory requirements that would govern cross-border data transfers and possible implications for setting up a compliant, interoperable, and secure data exchange system between China and the EU.

3.4. Methodology for Research Finding 2

To identify the ways in which Chinese Data Policies can promote the implementation of the principles of FAIR, and how these principles can be adjusted to meet the challenges posed by PIPL regarding strict data localization rules and GDPR's emerging cross-border data transfer requirements, this research methodology has primarily centered on the development of a Cross-border Data Exchange Architecture, which was conceptually and technically evaluated.

The methodology begins with the development of a Cross-border Data Exchange Architecture that integrates the principles of FAIR with the core components of Ownership (O), Localization (L), and Regulatory Compliance (R). This architecture was built to conform with legal requirements under both regulations and was developed using the analysis and initial conclusions drawn from Research Finding 1, which compared the approaches of PIPL and GDPR about data localization and cross-border data transfers. Thus, it provides for data localization to comply with PIPL while catering for cross-border data sharing in a secure manner through an array of mechanisms such as SCC under GDPR.

The architecture is to be designed in such a manner that it enables the autonomy of data collection with respect to the informed consent of the multi-cloud enabled smart contracts, which will allow both the regions to take an upper hand over their own data while also allowing certain data-sharing provisions.

3.5. Methodology for Research Finding 3

To evaluate the implementation and performance of FAIR-OLR-based architecture, the research methodology will focus on the implementation of the architecture itself, the testing of the system, and the definition of key metrics that will assess its effectiveness in ensuring regulatory compliance and its adherence to FAIR-OLR principles.

The first step is the implementation of the FAIR-OLR-based architecture. This includes the establishment of the FDP where sensitive healthcare data is stored in a FAIR-compliant format. FAIRifying raw data was done with the use of tools like CEDAR templates and BioPortal ontologies to make sure that they became machine-readable and interoperable. In addition to that, smart contracts and informed consent mechanisms were integrated into the architecture to manage data ownership and control data access. These mechanisms are aimed at automating the granting and revocation of data access to comply with PIPL and GDPR requirements.

System testing followed the implementation of the architecture. This testing aimed at checking the ability of the architecture to comply with PIPL's local data process laws and GDPR's laws on handling cross-border data between China and the EU. Systems Integration Testing was aimed at checking whether the components of architecture function harmoniously, even with FDP, AllegroGraph for query, and smart contracts for data access. Regulatory Compliance Testing examined whether the system was compliant with PIPL and GDPR's requirements, ensuring that all personal data were stored securely, localized where required, and shared within a compliant process.

To further assess and evaluate the efficacy of the architecture, some metrics were proposed for examining its performance in real-life scenarios. These metrics were set to check other parameters like the architecture's ability to maintain data ownership, compliance with localization requirements as well as compliance with general regulatory rules. Such main parameters are data findability, which would define how easily they could obtain data in a SPARQL query; data accessibility, which defines how easily authorized users can access it when needed, and compliance rate, which elaborates on how often the system applies PIPL and GDPR during cross-border data transfers. Additionally, metrics for interoperability and data security were also considered to ensure smooth shaping of security and interoperable data exchange across regimes without derogation of any privacy or security protocols.

Lastly, the scalability of the architectures was checked to make sure that it could manage the increasing volumes of data and requests from users in the course of time. System load was evaluated under high-load conditions to stress-test its remaining capacity for large datasets and multiple users to ensure it remains effective and secure as it scales. Performance metrics such as response time, data throughput, and system load capacity would be measured to evaluate how well the architecture worked in high-demand settings.

4. Research finding 1: The Specific Regulatory Requirements for Cross-Border Data Transfers under PIPL and GDPR

The first research finding examined the particular regulatory mechanisms that these frameworks employ in considering how PIPL and GDPR regulated cross-border data transfer-especially in balancing data privacy and national security. Both PIPL and GDPR have different modes of approach to achieving harmony between the personal data protection regime and the national security interests of their jurisdictions. An extensive analysis of the regulatory frameworks within which both laws operate would be provided in this section, including a discussion on key points such as data transfer restrictions, adequacy determinations, and the permissible circumstances under which data may be transferred internationally.

4.1. The mechanisms that Chinese data legal framework deals with the regulations of cross-border data transfers, particularly in balancing data privacy and national security

During these times of speedy digital transformation, cross-border data transfer is becoming the central pillar of the global economy. For international trade, technological development, and socio-economic growth, businesses and governments depend on instant access to data exchange (Tehrani et al., 2018). Even though this interlinkage bears considerable risks, including data breaches, cybersecurity espionage, and inability to patrol across border data movement.

Based on the above concerns, China as one of the biggest digital economies has put in place an elaborate legal regulatory framework on cross-border data transfer that weighs the need for supporting economic growth through data-driven innovation versus the need to protect national security by placing restrictions on the free flow of sensitive information. The core legislative cornerstones of this framework comprise the Cybersecurity Law (2017), the Data Security Law (2021), and the Personal Information Protection Law (2021). Together, they provide a basis for a highly structured and rigorous regulatory regime (Calzada, 2022).

China's regulatory approach, based on data classification and tiered protection, is at the center of all considerations regarding the level of compliance necessary for various types of data (Riccio, 2024). These tasks thus refer to sensitive and critical data remaining with strict controls but may be transferred across borders in lawful and safe means, given conditions are met.

The section at hand delves into China's mechanisms for regulating cross-border data flows, an examination that details the legal framework, step-by-step process of compliance, and challenges that actors face amid this complex regulatory environment.

4.1.1. Legal Framework and Key Principles

China's cross-border data transfer mechanism is based upon three principal laws that provide the regulatory framework for balancing data flows while paying attention to national security, economic development, and individual privacy concerns(Calzada, 2022).

4.1.1.1. Cybersecurity Law (2017)

The Cybersecurity Law (CSL), enacted in 2017, is the cornerstone of China’s data governance framework. It aims to enhance cybersecurity and ensure the protection of critical information (Creemers, 2023). More generally, it provides for all precautionary arrangements for those Protective Critical Information Infrastructure Operators (CIIOs), with operators/equipment in sectors such as finance, energy, telecommunications, and transportation. The mandate of the law is that all CIIOs must store any critical data within the borders of the country, ensuring governance over the sensitive information so that it remains under the Chinese authorities(Parasol, 2018). This data localization requirement aims at mitigating the risk of unauthorized foreign access and related potential exploitation. CIIOs would face rigorous security reviews by the Cyberspace

Administration of China (CAC) before data was transferred abroad(Creemers, 2023). Such reviews will look at whether or not the transfer poses risks to national security, public interest, or personal privacy. Transfers that don't satisfy the security standards are disallowed, as this shows that the government cares for something far more than something in the nature of national interest (Creemers, 2023).

Apart from the CIIOs, the CSL sets several obligations for network security for all organizations operating within China. Among the measures prescribed are data encryption, user authentication, and emergency plans with a view to preventing ability in new incidents against cyber-spacial reach (Sacks & Li, 2018). Periodic reviews and inspections by regulatory bodies are necessary for compliance with these standards. In addition, the CSL encourages a framework for privacy protection through the mandates for consent for collection, use, and disclosure of any individual's personal data(Parasol, 2018). Most of the provisions were further expanded under the PIPL; thus, it seems the initial draft of the Personal Data Protection was made under the framework of the CSL.

There reigned fear about growing cyber threats ranging from espionage, cyberattacks, and foreign exploitation of critical data that consequently brought about the formulation of the CSL. Yet this law runs parallel to the Chinese superior in seeking cyber sovereignty, where provisions center on national control over a nation's digital infrastructure and the data within its territory(Creemers, 2023). Challenges are nevertheless reported to have taken place in its implementation. The broad definition of CIIO generates ambiguities, thus, many enterprises are uncertain regarding whether they fall under this category. Moreover, the requirement to store data in the country raises the cost of compliance for multinational corporations operating in China and creates a potential conflict with international data transfer regulations(Sacks & Li, 2018). Despite these challenges, the CSL has remained at the center of China's legal framework for data governance. It set in motion subsequent regulations, including the Data Security Law and the Personal Information Protection Law, that together are creating an increasingly nuanced regulatory regime for governing cross-border data transfers (Li, 2021).

4.1.1.2. Data Security Law (2021)

The Data Security Law (DSL) that came into being in 2021 forms the basis of the regulatory environment already carved out by the Cybersecurity Law but suggests an increasing complexity of data governance in a digital economy (J. Chen & Sun, 2021). The DSL introduces a detailed framework for data protection based on a classification scheme that aims at striking a balance between fostering data-driven innovations and protection of national security and public interest. The very heart of the DSL is a requirement that organizations classify and manage data, depending on its level of significance, and the highest scrutiny is for that data under a designation as "important data"(J. Chen & Sun, 2021).

Pursuant to DSL, important data is defined as the information damage caused by leaking or mishandling could jeopardize national security, economic stability, or public interests. Organizations managing important data must enhance their security measures - these include risk assessments, data audits, and contingency plans (Miao & Lei, 2016). Furthermore, the Law stipulates that companies are required to undertake a self-assessment to determine what risks, if any, would be posed by each of their cross-border data transfers. The self-assessment is done with regard to the necessity of the transfer and the foreign receivers' capability figures for security as well as whether such data transfer would affect national security. The self-assessment reporting is then submitted to the Cyberspace Administration of China (CAC) for approval before the transfer can take place(Lee, n.d.). The DSL is known for its strict measures relative to certain supervised processes necessary for companies that handle sensitive data. Nonetheless, the focus is on preventing unauthorized access to or misuse of this information.

Such as other legal instruments, DSL also covers those companies and persons outside of China. If foreign entities work with or collaborate with local Chinese companies on data concerning China, they too will fall under DSL, provided their actions are determined by China to be related to national security or public interest. This extraterritorial effect highlights the government's attempt to retain strict control over any data exported from China, no matter where it is processed.

Some of the notable features of the DSL are its emphasis on setting up a strong accountability system. They require organizations to employ or appoint data security officers and establish internal data governance structures to guarantee compliance. The consequence of non-compliance with DSL laws includes huge fines and can even reach a standstill of activities when the severity warrants such a measure(J. Chen & Sun, 2021). For example, violations regarding important data can attract penalties of up to ten million renoumbers, underscoring the government's concrete approach to data security.

The DSL builds on the groundwork laid by the CSL, underpinned by the challenges faced by operators of critical information infrastructure. In shifting toward a classification-based approach, the DSL focuses the spotlight on the nature of the data, quite apart from the issue involving the identity of the business managing it. However, this development has also ushered in different kinds of vagueness, especially regarding clarity in defining "important data." Without precise guidance on determining valuable data, different industries and regions often end up with incongruous interpretations which posed compliance problems for affected businesses under the law. Nevertheless, the DSL had represented a pivotal turn in data governance in China. While it reinforces the regulatory framework introduced by the CSL, it clumsily paves the way for the incorporation of data safety into economic growth and national defense strategy.

4.1.1.3. Personal Information Protection Law (2021)

The PIPL, which came into effect on November 1, 2021, represents a major milestone in China’s data protection framework (Torrisi, 2023). Being the first comprehensive personal information protection legislation in China, this document shows China's growing interest in privacy, security, and all things digital sovereignty in a highly connected world. The PIPL shall further build upon the foundations laid by the Cybersecurity Law (CSL) and the Data Security Law (DSL), seeking to protect individual rights, regulate corporate behavior, and align personal information management with national security needs(Tan & Zhang, 2021).

PIPL also reflects the strategic perspectives of China in the light of global data protection standards. It reduces various administrative and regulatory barriers for the free flow of data internationally, having been inspired by a handful of international frameworks such as free templates to GDPR of the European Union(Calzada, 2022). By emphasizing transparency, accountability, and individual control over personal data, it responds to domestic privacy concerns while meeting the global need for interoperable data governance. Its extraterritorial provisions further demonstrate China’s intent to maintain control over data originating within its borders, regardless of where it is processed, reinforcing its broader goal of cyber sovereignty (Liu & Chen, 2024).

This groundbreaking legislation applies to a wide range of activities, including data collection, storage, processing, and transfer, affecting entities both within China and internationally. With detailed provisions on lawful data processing, cross-border data transfer mechanisms, and individual rights, the PIPL establishes a strong and enforceable framework (Torrisi, 2023). It not only protects data subjects but also offers businesses clearer compliance pathways. As a cornerstone of China’s approach to regulating its digital economy, the PIPL positions the country as a key player in global data governance discussions (Creemers, 2022).

Meanwhile, the PIPL directly solve several challenges identified in the CSL and DSL. One significant improvement is the precise categorization of data. While the CSL introduced the concept of critical information infrastructure operators (CIIO) and the DSL classified data based on its importance, these laws left ambiguities in defining critical and important data, which created compliance challenges for businesses (Torrisi, 2023). The PIPL provides clarity by focusing on two specific categories and three primary mechanisms.

Two specific categories:

1. Personal Information: Personal Information refers to any data that identifies or could identify an individual, such as names, contact details, or ID numbers. Anonymized data, however, falls outside the scope of regulation (Creemers, 2022).

2. Sensitive Personal Information: Includes data such as biometrics, religious beliefs, health and financial information, personal whereabouts, and information about minors under 14. For this category, stricter requirements are applied.

Three primary mechanisms:

1. Security Assessments: Organizations handling large volumes of personal data or sensitive information must undergo security assessments organized by the Cyberspace Administration of China (CAC). These assessments evaluate the necessity of the transfer, the receiving party's security capabilities, and potential risks to national security or individual rights.

2. Standard Contractual Clauses: For transfers not exceeding specified thresholds, organizations can rely on CAC-approved contractual clauses that outline the rights and responsibilities of both parties, ensuring compliance with Chinese data protection standards.

3. Certification: Companies can choose for personal information protection certification through authorized institutions, simplifying cross-border transfers by demonstrating compliance with regulatory requirements.

What's more, one of PIPL's notable contributions is its effort to reduce the gap between China's regulatory framework and international standards, particularly the EU's GDPR. These points of coincidence include both legislation insist that explicit consent should be sought before processing sensitive data; access rights, rights of correction, and deletion of personal data provided to individuals promote their privacy; organizations have to put accountability measures in place, including appointing data protection officers and conducting impact assessments.

4.1.2. Mechanisms for Cross-Border Data Transfers

Cross-border data transfer is now a cornerstone of international digital economy, enabling cross-border trade, technological modification, and international connectivity. China subscribes its framework regarding cross-border data transfers to data types, entities processing the data, and data quantity. Such transfers are further fine-tuned by certain mechanisms-a security assessment, certification, and standard contractual clauses, depending on the type and scope of each transfer(Li, 2021).

This section further discusses China's cross-border data transfer framework by analyzing the mechanisms described here. It details the procedures that need to be followed by CIIOs as well as non-CIIOs, mainly based upon the type of data and volume which dictates the alternate compliance pathways. The next discussion also highlights the differences between mandatory security assessments, certifications, and contractual agreements, emphasizing their respective roles in getting lawful and secure data transfers from China.

4.1.2.1. Mechanisms for CIIOs

The cross-border data transfer framework in China subjects the Critical Information Infrastructure Operators, or CIIOs, to stringent rules depending on the type of data they hold. The classification of data into important data and personal data provides the basis for specific compliance requirements for CIIOs (Corrales Compagnucci et al., 2021a).

Important Data means any information which, if leaked or altered, or accessed without authorization, could prove detrimental to China's national security, economic stability, or public interest. Examples of important data include operational data related to critical sectors such as energy, telecommunications, transportation, finance, and public health(Li, 2021). The specific scope of sensitive data will be determined on a case-by-case basis depending upon the severity of the case, which often necessitates the CIIOs internally assessing which datasets meet these criteria. Sensitive data management requires the maximum protection, while cross-border transfer would undergo a very stringent process of thorough security assessments.

Personal Data, on the other hand, including information related to identified or identifiable individuals, such as names, contact information, identification numbers, and online behaviours (Torrisi, 2023). Such sensitive data make up a major part of the business operations of many CIIOs. For example, CIIOs in telecommunications could be processing massive datasets containing the communication records of their users, while CIIOs in the financial industry are likely handling personal financial data(Li, 2021).Yet, even if not affecting national security directly, incorrectly handling personal data can incur significant privacy breaches, reputational losses, and decreases in public trust.

Together, these two categorizations ensure that the regulatory approach for CIIOs is both comprehensive and targeted. Critical data receives stricter oversight owing to its governance implications, while personal data must receive adequate privacy protection to cater for personal rights. The classification system allows regulators to prioritize their supervisory resources on the basis of the sensitivity and expected risk of each type of data(Li, 2021).

What's more, to ensure the secure management and transfer of these data types, CIIOs must take a mandatory security assessment before engaging in cross-border transfers (*Cross-Border Data Regulations in the European Union and South Korea*, n.d.). The process for the security assessment begins with the CIIO initiating an internal review of the data to be transferred, with this self-assessment providing an assessment for some of the more critical areas: whether transfer is essential; data volume and sensitivity; and the national, public, or individual legislative impact. Thereafter, the CIIO must formulate and forward a summary risk assessment report to the CAC, explaining the classification of the data and the intent behind the transfer in question(Savona, n.d.), the receiving party's security capabilities, and the safeguards implemented to ensure secure data handling.

The CAC subsequently conducts a comprehensive examination of the submission. This examination is directed toward determining whether or not the transfer complies with the requirements laid down in the treaties, and what the possible risks are that may occur with the foreign entity receiving the data (Guo & Li, 2025). Key considerations include:

- **Necessity**: Is the cross-border transfer necessary for the CIIO's operations or public services?
- **Security Measures**: Does the receiving party possess sufficient data protection measures to prevent unauthorized access, breaches, or misuse?
- **Risk Analysis**: Could the transfer expose the data to risks that may threaten China's national security or public interests?

Upon approval of the transfer by the CAC, authorization is granted for a period of validity not exceeding two years. During this time, the CIIO must have stricter oversight of the data transfer by implementing continued monitoring and compliance assessments in order to ensure that the transfer remains secure(Li, 2021). In the case of security breaches or compliance failures, the CIIO shall immediately inform the CAC. Such authority will, however, ensure that needed steps are taken to suspend approval for transfer(Guo & Li, 2025).

Where the CAC identifies significant risks or deficiencies during the review, denial of the transfer request must ensue. The CIIO then must address the specified issues via either bolstering data protection measures, modifying the scope or specific purpose of the transfer, or using alternate solutions to meet the operational requirement while respecting regulatory requirements(Liu & Chen, 2024).

The compulsory security assessment embodies how Central China has placed priority upon national security and public interests in its cross-border data governance framework (M. Chen, 2024). By imposing stringent evaluation processes, the CAC sees to it that the CIIOs are held responsible for data management to minimize risks while supporting critical infrastructure operations(Li, 2021). This embodies China in reinforcing its broader agenda of data sovereignty and cybersecurity, creating a very robust and enforceable mechanism for high-risk data transfer.

4.1.2.2. Mechanisms for None-CIIOs

For CIIOs falling outside this scope, China's cross-border data transfer framework is less rigid. These entities, termed non-CIIOs, deal with more varied data types, including personal information and important data, but do not provide support or enable activities that have serious implications for national security(M. Chen, 2024). Consequently, the regulatory mechanisms controlling non-CIOs are, in a way, lesser strict comparatively, yet compliance and data protection are still their focus.

The management of data by non-CIOs begins with data categorization, a critical step for determining the specific compliance pathway. Similar to CIIOs, non-CIIOs must classify their data into important data and personal data (Torrisi, 2023). Important data under non-CIIOs may include information that, while significant for certain industries, does not have direct ties to national security but could still disrupt public interests or industry operations if mishandled (Guo & Li, 2025). Personal data for non-CIIOs typically involves information related to individuals, such as contact details, purchasing habits, or demographic information, often collected at a large scale for commercial purposes. Although personal data handled by non-CIIOs may not carry the same security implications as data managed by CIIOs, its improper use can still lead to privacy violations and reputational risks for organizations.

Once data is categorized, non-CIIOs follow distinct cross-border transfer steps that vary based on the volume and sensitivity of the data. For transfers involving personal data, the compliance requirements are divided into two pathways depending on whether the data volume exceeds regulatory thresholds.

For data exceeding volume thresholds, organizations must first conduct a self-assessment to evaluate the necessity of the transfer, the receiving party's data protection capabilities, and potential risks (Li, 2021). Thereafter, the self-assessment results would be submitted to foreign authorization institutions for certification by its founding institution, for example, the China Cybersecurity Review Technology and Certification Center. Such a process would entail an intense review of data protection standards within the institution, including those relating to the secure storage of data and the protocols established at the point of transmission and during its transmission to a given recipient (Guo & Li, 2025). This certification, once obtained, is valid for a tenor of 3 years and extends formal and flexible processes in respect of higher risk data transfers.

With data volumes falling beneath their respective threshold limits, non-CIIOs may make use of standard contractual clauses sanctioned by the Cyberspace Administration of China. These clauses outline basic obligations for both the transferring entity and the receiving party with respect to data security, breach notification, and the rights of data subjects(Y. Chen & Song, 2018). This mechanism, then, is simplified in terms of transferring types and is thus most suited to small and medium enterprises (SMEs) and startups undertaking cross-border transactions.

Unlike CIIOs, non-CIIOs are not required to undergo mandatory security assessments for every transfer across borders. As for the rest, under a newly developed regulatory regime focusing more on contractual obligations and voluntary certification, non-CIIOs enjoy more operational flexibility but sufficient safeguards remained. In other words, through a tiered compliance framework for non-CIIOs, China manages to balance the need for robust data protection with practical requirements for businesses in less critical industries. The less stringent modalities for non-CIIOs compared to CIIOs reveal a consistent push for protection of cross-border transfers of data merely from interrupting business activities(M. Chen, 2024). This unique approach should show the broad versatility of China's regime of data governance in considerations to accommodate different levels of risk exposure among the entities, while unifocal with a standard commitment to data security. Figure 2 provides a clear view of cross-border data circulation under China's data law framework.

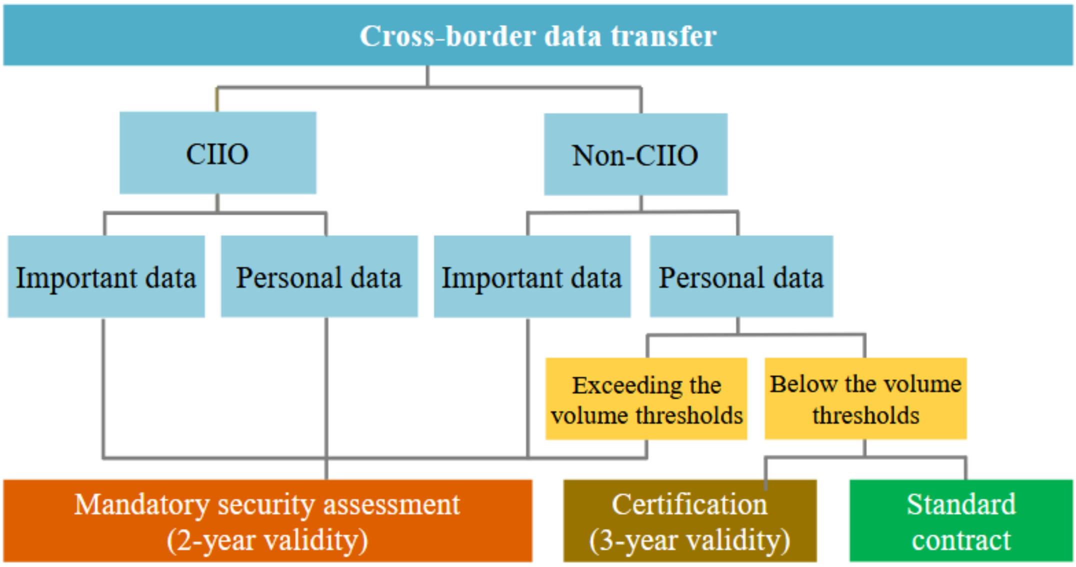

Cross-border data transfer
CIIO
Non-CIIO
Important data
Personal data
Important data
Personal data
Exceeding the volume thresholds
Below the volume thresholds
Mandatory security assessment (2-year validity)
Certification (3-year validity)
Standard contract

Figure 2 Cross-Border Data Transfer Mechanisms in China

4.2. The mechanisms that GDPR deals with the regulations of cross-border data transfers, particularly in balancing data privacy and national security

The European Union is a world leader in the field of data protection, and its legal framework often sets the tone for global standards. At the center of it is the General Data Protection Regulation (GDPR), effective from May 2018 (Bakare et al., 2024). GDPR is considered one of the most extensive data protection laws worldwide, focusing primarily on transparency, accountability, and the protection of individual rights. As the law has extraterritoriality functioning, any organization outside the EU handling the personal data of EU citizens has to comply with the rule, highlighting the law's far-reaching influence (Demirer et al., 2024).

Transfers of data across borders are considered an essential component of the EU digital economy, which spans the areas of international trade, innovation, and cooperation. At the same time, the transfer of data also creates certain problems for data security and privacy, especially if the data is transferred to jurisdictions with less stringent protection measures(Bakare et al., 2024). The GDPR provides a number of mechanisms for the protection of personal data, regardless of physical borders and data processing location(Sullivan, 2019). These mechanisms, found in Articles 44-50 of the GDPR, include adequacy decisions, SCCs, and Binding Corporate Rules (BCRs) (Rzayeva, 2024). Each of them is supposed to give a legitimate basis for data transfers while retaining the necessary protection of the data within the level of expectation found in the EU.

This section covers a deeply analytical discussion of EU mechanisms for cross-border data transfer that discusses the legal bases, mechanisms for implementation, and limitations presented. Therefore, the analyses of the key processes delineated under the GDPR and their implications for businesses and individuals are intended to present a balanced view about how the EU balances an opportunity for international engagement against the demand for individual data privacy(Corrales Compagnucci et al., 2021b).

4.2.1. Legal Framework

The principal framework guiding the EU on this cross-border data transfer reflects its engagement with individual privacy and the safe passage of information across borders. The GDPR-which was effective from May 2018-is indeed the main element under the framework governing the legislation, repealing the 1995 data protection directive and making modifications therein to customize and reflect the complexity of the modern digital landscape.

The GDPR establishes principles for the lawful processing and transferring of personal data, focusing on announced expectations, organizational accountability, and individual rights. Articles 44 to 50 specifically address the mechanisms by which data transfers will occur, setting out the mechanisms and conditions under which personal data may be transferred to third countries or international organizations(Bakare et al., 2024).

The regulation makes sure that the provisions apply not only to the entities within the EU but also to those outside the EU which process personal data of residents of the EU. This extraterritorial application guarantees that protections equivalent to those of the GDPR go beyond the EU being made globally impactful for the system of data protection.

The main intention of the GDPR, therefore, rests on the protection of lost individuality, protecting its own identity by the aforementioned assertion that personal data remain safe under a care of autonomy by the individuals that have it, even when it is processing outside that jurisdiction. Still, it permits global data flow contrary to the traditional boundaries created by data protection policies, thus granting interoperability for international collaboration and economic activity through adaptable and standardized paths for legitimate transfers. Uniform rules among EU member states, meanwhile, coordinate standards in data protection practices with global ones. Besides, the regulation puts the onus onto organizations that are required to ensure sufficient measures for data protection so that any transferee in a third state complies with rules of the GDPR.

The implementation of the GDPR is underpinned by a system of regulatory authorities that oversee both compliance and enforcement. Each Member State of the EU shall have its own Data Protection Authority (DPA), enforcing the GDPR at the national level and providing guidance regarding cross-border data transfers. The European Data Protection Board (EDPB) provides this guarantee of uniformity throughout the EU by issuing opinions on adequacy decisions and standard contractual clauses, as well as handling disputes involving multiple DPAs. The European Commission, upon assessing, will accord an adequacy decision based on its opinion on a behind-closed-doors deal providing protection for data transfers to specific third countries(Demirer et al., 2024).

With regard to the adoption of the GDPR, it converges with the other mechanisms provided by international treaties and initiatives for global data protection, thereby extending its reach beyond the borders of the EU. Other adequacy decisions such as those made for Japan, South Korea, and the UK, support the free exchange of information in that their data protection legislative polities are in fact equivalent to that of the GDPR. However, with the invalidation of the EU-US Privacy Shield framework following the decision in Schrems II, any future approaches at facilitating transatlantic data flows will need to look for other alternatives to ensure continued compliance. All over the world, GDPR is an instrument used in the building of similar laws like Brazil's LGPD and China's PIPL, showing how it serves as a model for international data governance.

4.2.2. Mechanisms for Cross-Border Data Transfers

The GDPR provides a robust legal framework to ensure the lawful and secure transfer of personal data to third countries or international organizations (Bakare et al., 2024). These mechanisms are contained in Articles 44 to 50 of the GDPR and are meant to maintain a level of protection for personal data that is almost equal to that which lies east of the EU. The principal data transfer mechanisms are adequacy decisions, SCCs, BCRs, and exceptional cases of specific derogations.

Article 45 makes an adequacy decision the most evident mechanism for transferring personal data outside the EU under the GDPR. A third country, a territory, or a specific sector within that country is assessed by the Commission in terms of providing adequate protection for the processing of personal data. The European Commission take into account different factors in determining their completeness, including international commitments, the country's legal framework, and finally, but by no means least, its enforcement mechanisms. Countries such as Japan, South Korea, and the UK are among those designated as so-called adequate, whereby their personal information can, therefore, be transferred without any additional restraints(*Cross-Border Data Regulations in the European Union and South Korea*, n.d.). These decisions are subject to periodic review to ensure continued compliance, with the possibility of revocation if a country’s data protection standards lost trust (Demirer et al., 2024). For example, the invalidation of the EU-US Privacy Shield after the ruling in Schrems II shows the Commission's flexible monitoring and the necessity for equivalent data protection levels. The ruling in Schrems II by the Court of Justice of the European Union has essentially changed the way these policies are implemented(Corrales Compagnucci et al., 2021a). To this end, it invalidated the EU-US Privacy Shield but underscored the use of supplementary measures in the case of transferring information such as encryption and pseudonymization to ensure equivalent protection in the countries that are deemed non-adequate. Consequently, organizations are therefore required to conduct extensive risk assessments of their transfers to countries like the US, supposedly to comply with GDPR standards.

When not in possession of an adequacy decision, organisations may lean on 'appropriate safeguards', as elaborated in Article 46. Probably the most used tool within this context is SCCs. These standard contractual clauses, previously authorised by the European Commission, set out legally binding contractual obligations between data exporters and importers with respect to such matters as GDPR compliance. Other areas touched by these standard contractual clauses include provisions on data minimisation, breach notifications, and enforceable rights for data subjects (Peloquin et al., 2020). Following the Schrems II ruling, the European Commission updated SCCs in 2021, strengthening obligations for data importers and addressing concerns about government access to data in third countries (Corrales Compagnucci et al., 2021a). For multinational corporations, BCRs under Article 47 offer an alternative mechanism, allowing data transfers within corporate groups under a unified privacy governance framework (Demirer et al., 2024). BCRs require supervisory authority approval and typically include robust compliance measures, such as data protection principles, employee training, and accountability mechanisms. To further reduce risks, organizations transferring data to non-adequate jurisdictions are now required to adopt supplementary measures, such as encryption, pseudonymization, or tokenization, to address concerns about unauthorized government access and data breaches.

In situations where neither adequacy decisions nor safeguards are suitable, the GDPR permits data transfers based on reductions under Article 49 (Li, 2021). These include cases such as the specific agreement of the data subject, the necessity of the transfer for contractual performance, or a important reasons of public interest. However, these reductions are exceptional measures and are interpreted narrowly to prevent abuse.

While GDPR primarily focuses on personal data protection, its approach to non-personal data is slightly more flexible, reflecting the varied risks associated with different data types. To be more specific, for cross-border data transfer, the GDPR classifies data into two primary categories—personal data and non-personal data—each subject to varying levels of oversight based on its sensitivity and associated risks. For personal data, transfers are further divided into “normal” and “sensitive” categories (Sullivan, 2019). Regular personal data can be transferred to States that have been accorded an adequacy decision by the European Commission, which attests to the fact that the recipient country provides a level of protection for personal data that is deemed adequate to that existing within the European Union. Such transfers will go through unimpeded and will need no supplementary safeguards. However, it will be when the adequacy decision has not been rendered whereby the transferring entities shall have to obtain appropriate supporting mechanisms, such as SCCs, BCRs, or individual certifications to support compliance with GDPR. Sensitive personal data, such as health or ethnic origin, requires additional safeguards, including encryption, pseudonymization, and supplementary risk assessments.

In the case of non-personal data, the GDPR is more flexible. In general, a non-personal data transfer may proceed without restriction, so long as the transfer is carried out in accordance with overarching EU regulations. Examples of data that do require additional scrutiny include sensitive non-personal data, for example from the telecommunications or the energy sector. The assessment looks at risks posed by any transfer initiatives and applies regulations for a specific sector, along with taking competent authority opinion when necessary (Bakare et al., 2024). Such transfer may only occur upon satisfaction of these conditions, a reflection of the objectives of the GDPR regarding striking a balance between security and operational flexibility.

In sum, the framework introduced in the GDPR on transborder data exchange distinguishes unrestricted from conditional. Unrestricted transfers are possible when the recipient is an entity that meets GDPR standards through adequacy decisions or recognized safeguards (Demirer et al., 2024). In contrast, conditional transfers require strict conformity with sectoral requirements, requirements for data protection, and approaches to risk mitigation. This tiered approach reiterates GDPR's commitment to a high standard of data protection, while permitting the international flow of data vital to global commerce and innovation.

The Figure 3 shows a clear cross-border data transfer mechanisms in EU.

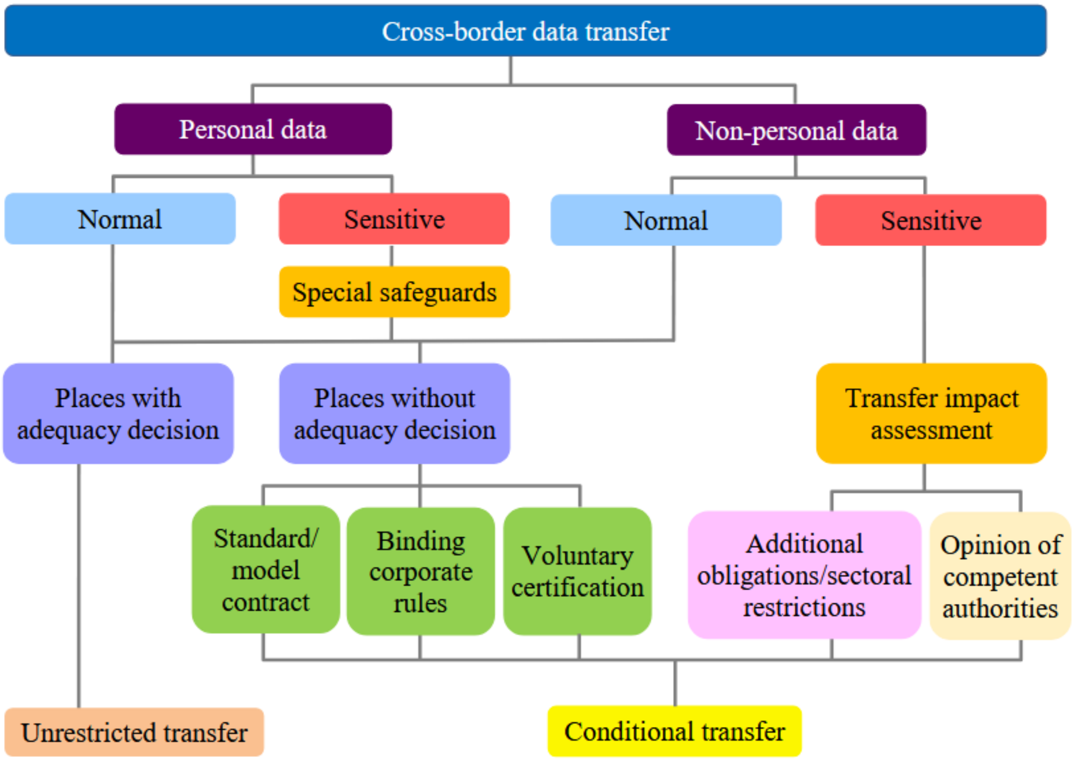

Cross-border data transfer
Personal data
Non-personal data
Normal
Sensitive
Normal
Sensitive
Special safeguards
Places with adequacy decision
Places without adequacy decision
Transfer impact assessment
Standard/ model contract
Binding corporate rules
Voluntary certification
Additional obligations/sectoral restrictions
Opinion of competent authorities
Unrestricted transfer
Conditional transfer

Figure 3 Cross-Border Data Transfer Mechanisms in EU

4.2.3. Challenges and Practical Applications

The implementation of the GDPR's cross-border data transfer framework has run into many challenges with paralysis from the Court of Justice of the European Union ruling in the case of Schrems II invalidating the EU-US Privacy Shield, making it much more difficult to comply(Sullivan, 2019). The court stressed that complementary measures, including encryption and pseudonymization, should apply for better safeguarding against foreign government surveillance. For this reason, transferring organizations must now conduct Transfer Impact Assessments (TIAs) to consider the risks posed by the legal environment of the recipient country(Savona, n.d.). Many assessments are felt with particular acuteness by organizations dealing with high volumes of data or jurisdictions in which they operate, as they require repeated revisions dictated by changing regulatory landscapes.

Practical applications of GDPR further illustrate the tensions between demands for data localization and the need for global interoperability. In particular, the sectors of finance and healthcare in the EU have imposed even more restrictions on the transfer of sensitive data. Consequently, it is mandated that they adhere to sector-specific compliance guidelines(Sullivan, 2019). This fragmented regulatory environment complicates cross-border data flows, especially for multinational corporations that must bring their operations in line with both EU and non-EU regulations. On the other hand, the principles of GDPR have motivated more global standardization efforts, prompting the likes of Brazil's LGPD and China's PIPL to implement similar frameworks. These developments present the GDPR's key role in reordering global data governance even if varying in alignment(Demirer et al., 2024).

Nonetheless, the framework, which undergirds the general data protection regulation, has remained integral in fostering accountability and transparency for cross-border maintenance and transport of data. The GDPR upholds a secure and legal flow of data by ensuring that compliance pathways exist and that organizations are held accountable for safeguarding personal data(Savona, n.d.). Moving forward, enhancing the practicality and scalability of compliance mechanisms will be essential to support both SMEs and large enterprises in navigating the evolving global data economy.

4.3. The main differences between the PIPL and GDPR regulations regarding the handling and protection of sensitive data

The European Union (EU) and China have different cross-border data transfer frameworks that reflect their varying legal traditions and strategic priorities. While both aim to balance personal data protection with economic growth, the specific legal bases, regulatory mechanisms, and enforcement practices diverge significantly(*The GDPR vs China's PIPL*, n.d.). These differences highlight opportunities for integrating FAIR principle （Findability, Accessibility, Interoperability, and Reusability） as a bridge to address shared challenges and foster global data collaboration.

4.3.1. Legal Foundations and Core Principles

While the foundational principles guiding GDPR and China's PIPL may vary considerably, these discontinuities imply divergences in projected goals. GDPR holds the notion of privacy as a fundamental right and therefore aims at a horizontal approach within the EU, with emphasis on individual rights and greater transparency. Rather, China's apparatus encompassing PIPL, DSL, and CSL prioritizes national security and data sovereignty with the aim of protecting critical infrastructure and controlling outbound data flow(Virtosu & Li, 2024).

EU (GDPR): The provision for extraterritorial application makes sure that any non-EU entity involved in processing data about EU citizens must comply with the GDPR. Article 45 permits data transfers to third countries that offer adequate safeguards for data protection, aiding global interoperability. This speaks readily to the promotion by the EU of privacy as a worldwide norm.

China (PIPL): Chinese laws put national interests ahead. Among other things, PIPL Article 38 states that major data needs security assessments to export data, while DSL Article 21 introduces "important data," which suggests that data important to public interests will remain under the management of the domestic country. These are elements of China's particular interest in cyber sovereignty rather than global harmonization.

Comparison and FAIR Alignment: While GDPR's coordinated policy encourages global interoperability, the importance attached by China to localization and security reviews gives rise to barriers for a seamless flow of data. Integrating FAIR principles could soothe this tension. For instance, the article on data portability in GDPR serves to make possible the transfer of data between systems in line with FAIR's accessibility. In the same way, data-sharing provisions in China could be directed by PIPL Article 41, which could establish FAIR-aligned metadata standards, to facilitate interoperable control over approved cases of data sharing- for example in scientific research.

4.3.2. Data Classification and Sensitivity Levels

Both frameworks rely on data classification, but their categorizations reflect different regulatory goals. The GDPR focuses on personal data and sensitive personal data, ensuring additional protections for the latter (e.g., health, biometrics). China adopts a broader categorization, introducing "important data" as a unique classification with stricter controls (Virtosu & Li, 2024).

EU (GDPR): Under GDPR Article 9, sensitive personal data requires explicit consent and additional safeguards, reflecting the EU's focus on protecting individual privacy. For instance:

"Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership is prohibited unless specific conditions are met."

China (PIPL and DSL): China's DSL and PIPL categorize data into general, important, and sensitive personal information. PIPL Article 29 requires clear agreement for processing sensitive personal information, while DSL Article 21 mandates strict localization for important data, such as public health or infrastructure information.

4.3.3. Mechanisms for Cross-Border Data Transfers

Cross-border data transition methods show huge differences in the flexibility of their regulations. The GDPR offers several appropriate channels such as adequacy decisions, SCCs, and BCRs. China relies instead on "mandatory" security checks and government-approved contracts.

EU (GDPR): The GDPR allows for great flexibility and interoperability by providing numerous mechanisms such as adequacy decisions, SCCs, and BCRs to allow lawful data flows to cross borders. While the adequacy decisions (Article 45) of GDPR allow effortless transfers to countries that provide equivalent data protection, SCCs (Article 46) and BCRs (Article 47) allow for flexible transfers to jurisdictions that have been determined not adequate. For instance, SCCs provide standardized obligations that data exporters and importers must adhere to, allowing for consistency and traceability in data-sharing practices.

China (PIPL): In its design, the system adopted by China leans toward centralized oversight and national security, with critical or very important data requiring mandatory security assessments prior to their being transferred. What torture standard contracts and certifications might be available for less risky situations, but all methods are heavily regulated and scrutinized by a conscious body in

China, the Cyberspace Administration of China (CAC). China makes the security assessments a category for important data and CIIOs. The article requires the organization to assess the need and security of data transfers as part of the review process. Standard contracts and certifications may address lower-risk scenarios, similar to GDPR SCCs but far more tightly controlled.

4.3.4. Data Localization Requirements

Data localization is one of the key differences. Under GDPR, there are no localization requirements, just that data flow is adequately protected. On the other hand, China requires localization for critical and important data to enhance security for the nation.

EU (GDPR): The GDPR is a flexible approach, whereby there are no explicit localization requirements but ensured secure and lawful cross-country transfer of personal data using mechanisms such as adequacy decisions, SCCs, and BCRs. For instance, under GDPR Article 45, data may flow freely to those countries which are determined to have an adequate level of protection. In the same way, Article 46 allows organizations to provide such appropriate safeguards to effectuate cross-border transfers into non-adequate jurisdictions. Such flexibility supports accessibility and interoperability, enabling free flow across borders that is shared among global industries like research, technology, and finance.

China (DSL and PIPL): In marked contrast, China has stringent localization mandates to safeguard national security and public interest. DSL Articles 25 demand CIAOs to store all important data collected through various sources in China to take care of personal and financial security regarding telecommunications data. PIPL Article 40 underlines that security assessments are required for anybody from the data processors handling a large quantity of data prior to permitting any cross-border transfers. Various sector-specific regulations reinforce localization demands; for instance, health data is frequently mandated to remain within the Peoples Republic of China, limiting its availability for international research. Beyond strengthening China's grip on sensitive data, these mandates are hindering interoperability and global collaboration.

4.4. Conclusion

This section then assesses the ways in which PIPL and GDPR define cross-border data transfers, especially as they relate to data privacy and national security. In contrast to traditional approaches to sensitive data handling, localization of data consent, and multinationals, the article's thorough comparative analysis pinpoints significant divergences in ways of approaching those things. Such divergences highlight Europe's and China's likely dissimilarity in philosophical and regulatory tendencies yet indicate certain areas of confluence regarding the organization of using FAIR principles.

To answer Research Question: How do PIPL and GDPR deal with the regulations of cross-border data transfers, particularly in balancing data privacy and national security?

The PIPL and the GDPR really do differ in their general dispositions regarding cross-border transfers, making major nods to their respective legal histories and strategic agendas. Privacy is framed as a universal right and an absolute centerpiece of the entire GDPR framework. Every means is accepted in making sure that personal data that goes outside the EEA maintains the same level of protection--these include adequacy, SCCs, or BCRs. This flexibility is about maintaining open and transparent policies, the rights of individuals, and interoperability among nations so that data can flow across national borders while still respecting privacy.

On the other hand, PIPL marries the protection of privacy issues with issues of national security and data sovereignty. In this sense, it compels mandatory security assessments for transfers of critical or large-scale personal data, per PIPL Article 38, with Article 25 of DSL mandating localization of “important data.” All this serves to exert stricter control over sensitive information, by emphasizing public welfare and stability for China. While GDPR stands for international cooperation, PIPL opts for the preservation of domestic administration and control in regulation.

To answer sub-Research Question: What are the main differences between the PIPL and GDPR regulations regarding the handling and protection of sensitive data?

The GDPR clearly defines sensitive personal data in Article 9, requiring explicit consent and additional safeguards, simplifying compliance for organizations and strengthening individual rights. China’s framework introduces "important data," a category broader than GDPR’s classifications, encompassing information critical to national security under DSL Article 21. However, ambiguities in defining important data create compliance challenges for businesses.

5. Research finding 2: FAIR implementation in China and FAIR-OLR based cross-border data exchange architecture

5.1. The proof that Chinese Data Policies enable FAIR Principles, and the impacts of FAIR Principles implemented in China

As we mentioned above, a strong legal framework for data governance, composed by CSL, DSL and PIPL, was built in China. Based on this framework, an analysis of data policies across various levels in China highlighted four key perspectives: Management of Data by Levels, Lifelong Data Stewardship, Data Publishing and Reusing, Long-term Data Preservation (Li et al., 2019, p. 287).

Management of data by levels, as emphasized in China's data policies such as the DSL, refers to the classification of data based on its sensitivity and importance. This approach ensures that data is categorized according to national security, public interest, and usability priorities. Metadata schemas utilized by platforms such as China Scientific Data label their datasets to some degree of classification, enabling Findability and Accessibility to guarantee that researchers find it easy to locate and to distinguish that data relevant to their projects under appropriate data labels.

Lifelong data stewardship ensures that data is managed responsibly at all points throughout its lifecycle-from creation to preservation for reuse in other contexts. Such stewardship calls on institutions or custodians of data to produce ample high-quality metadata, ensure long-term accessibility, and guarantee that data remains interoperable and reusable across varying platforms. Examples include the systems employed by GigaDB that enforce lifecycle stewardship by tagging datasets with metadata readable by machine, thus allowing researchers to locate datasets and reuse them while guaranteeing their compliance with international standards such as RDF and OWL(Edmunds et al., 2017). Such practices greatly improve Findability, Accessibility, Interoperability, and Reusability, guaranteeing that these data will stay useful for immediate and future research purposes.

Publishing and reusing research data is more than just making publicly funded research data easy and effective to reuse ethically. With policies in place, datasets must be published with clear licensing and usage conditions, guaranteeing they can be reused in an ethical and legal manner. For instance, repositories like China's Scientific Data, the first multidisciplinary data publication journal in the country, and National Microbiology Data Center are concerned with reuse through metadata standardization and the provision of persistent identifiers (such as DOIs) to allow datasets to be cited and traced.

Long-term data preservation focuses on the secure and sustainable storage of valuable datasets to ensure they remain available for future use. This requires robust filing solutions and standardized data formats that guarantee long-term (re)usability.

Based on these four views of legal frameworks, we can see China's existing legal frameworks support FAIR. In summary, categorizing the four perspectives on data policies based on the characteristics of FAIR data, as outlined in Table 1, highlights how these policies can enhance the capabilities required to support the implementation of FAIR data principles.

Data policies focus	Findable	Accessible	Interoperable	Reusable
Management of data by levels	✓	✓		
Lifelong data stewardship	✓	✓	✓	✓
Data publishing and reusing				✓
Long-term data preservation				✓

Table 1 Chinese data policies enabling FAIR data (Li et al., 2019, p. 289)

Recently, open data become more and more essential for advancing modern scholar communities and has appeared as an advantage trend in today's interconnected global landscape. Effectiveness and efficiency play a crucial role in advancing FAIR data principles, developing data ecology, and enhancing practices within the scientific community. FAIR data principles can serve as a guideline for the domestic data development field and provide inspiration for the development of data science in China. During the 2016 G20 Summit in Hangzhou, China, world leaders gathered to discuss strategies for fostering global innovation and sustainable development. One of the key highlights of the summit was the emphasis on the FAIR principles as an essential framework for advancing open science and data sharing within the global scientific community. The G20 Innovation Action Plan, which calls for investments in science, technology, and innovation as well as other things that serve to support knowledge diffusion and open access to the results of publicly funded research, was also introduced during the summit. Such an undertaking is aligned with the paradigms through which China's commitment to the advancement of data sharing and data science via global cooperation is on the rise. By hosting this important summit, China has manifested the leadership role in implementing the framework of the FAIR principles and already in setting an international dialogue about transparency, sustainability, and innovations in science and technology.

FAIR data is a good beginning for China, but, by itself, FAIR is not enough. This is also the basis for this thesis examining the application of various FAIR-based frameworks.

5.2. The methods that FAIR principles are adjusted to meet the challenges of PIPL's strict data localization rules and GDPR's cross-border data transfer requirements

This section presents the FAIR-OLR framework, incorporating the principles of Ownership, Localization, and Regulatory Compliance into the already established FAIR guidelines. FAIR-OLR offers practical solutions for meeting principles for PIPL and GDPR. These regulations, with their strictest localization of data requirements and prohibition of cross-border data transfer, are stiff challenges that call for figuring out innovative means around compliance so as to balance data sharing and reuse.

Therefore, this section proposes an architecture designed on the basis of FAIR-OLR in order to tackle the stated challenges. Among other things, this set of architecture brings to the merger and constitutes a federated model for managing data, thus allowing the localization criteria to be satisfied with the proper possible interoperability and accessibility of the data.

5.3. Strategies and technologies that can be implemented within a FAIR-based framework to ensure compliance with both PIPL and GDPR

In this section, I will construct a FAIR-OLR-based Architecture and outline specific technologies and strategies that can be practically deployed to ensure compliance with both PIPL and GDPR. The following subsections will detail the technologies and strategies supporting the implementation of this Architecture.

Before the architecture, there are some key technologies which will be used in the architecture:

CEDAR and BioPortal in FAIRification

CEDAR is a web-based platform developed at Stanford University, designed to facilitate the creation and management of metadata templates for scientific data. It provides open-source tools and REST APIs that enable data providers to create machine-actionable metadata for various types of data, including healthcare and clinical data.

At the heart of FAIRification process is the enrichment of metadata using ontologies, which are stored in the BioPortal repository. BioPortal is a comprehensive web-based portal that serves a variety of biomedical ontologies via several services. The combination of these ontologies contributes to the semantic enrichment of metadata, whereby the data in general are adequately described and standardized. BioPortal affords the use of existent ontologies; however, in instances whereby an ontology does not exist, bespoke ontologies can be made for the needs particular to the data.

Although BioPortal offers the great advantage of flexibility in selecting ontologies, some are faced with limitations of interoperability because its ontologies still retain certain idiosyncrasies in their within-systems basic structure and accompanying formal standards aimed at integration across heterogeneous systems. To tackle the challenge, a two-fold approach is used. BioPortal is put into service whenever there is room for flexibility; domain-specific vocabularies that retain much of the standardization process originating from the OBO Foundry are employed under this dual approach to enhance interoperability. This confluence of ontologies enhances the enrichment of healthcare data through the benefits of flexibility and standardization.

AllegroGraph For Triple Storage

This process produces a data model through which data representation of a conceptual schema can take place, yielding a machine-interoperable language for encoding and language linking known as the Resource Description Framework (RDF). Such a semantic graph has its Universal Resource Identifiers (URIs) as nodes facilitating metadata queries and access across healthcare facilities.

Data will then be modelled in RDF format and then put into storage for easy querying. To this end, as our triple store solution, we have chosen AllegroGraph. AllegroGraph is a multi-model graph database that can accommodate every kind of graph request that we want to apply to all kinds of types of data storage.

5.3.1. Ensuring Ownership (O) in the Model

This part refers, first of all, to the recommended architecture in alignment with the principles of FAIR-OLR via Ownership. Ownership means granting data subjects or institutions complete control over their data to guarantee that it is used only for the purpose agreed upon. Smart Contracts and Informed Permission mechanisms are implemented here in order to establish and reinforce ownership rights effectively. The Figure 4 describe the complete process how to ensure data ownership in the architecture.

Step 1: Initiating the Data Query via SPARQL API

The process begins with a query request being sent to the SPARQL API to access specific data stored in a system. This query acts as a formal request for data access, triggering ownership validation. SPARQL is a standard query language used for RDF-based data to enable machine-readable and interoperable queries in compliance with FAIR principles. This API acts as the entrance to access the data and initiates the ownership check mechanism.

Step 2: Granting Informed Permission by the Data Provider

The system notifies the Data Provider upon peer review request, and the Data Provider will consider the reason, scope, and conditions under which the request is made and will consider granting access to that information. The notion of Informed Permission here is implemented, where the Data Provider specified approval for data collection and running the query. Informed Permission ensures transparency by providing clear communication about how data will be used and aligning with PIPL's emphasis on informed consent and GDPR's data subject rights.

Step 3: Signing the Smart Contract

Once the Data Provider grants permission, the system generates a Smart Contract. This contract is a self-executing digital agreement that clarifies the permissions and conditions under which the data can be accessed. It provides several key functionalities:

- Ensuring only authorized queries can proceed.
- Creating an invariable, tamper-proof record of all transactions for audit purposes.
- Supporting dynamic updates to permissions when needed, without disrupting the system.

Step 4: Authorizing the Query in AllegroGraph

The smart contract communicates with the AllegroGraph database, a FAIR-compliant RDF storage solution, to authorize the requested query. This step ensures that only queries with valid permissions are executed. The database strictly brings into operation rules defined by the smart contract, firmly enhancing compliance with the ownership principles.

Step 5: Execution of the Authorized Query

Ultimately, the SPARQL API dispatches a query-Is undertaken according to permissions gained from the smart contract. The query retrieves those data which indeed are allowed under their mutual agreement, respectively bringing a total piece of action towards ownership right. The results thereafter are delivered securely to the requester for closure.

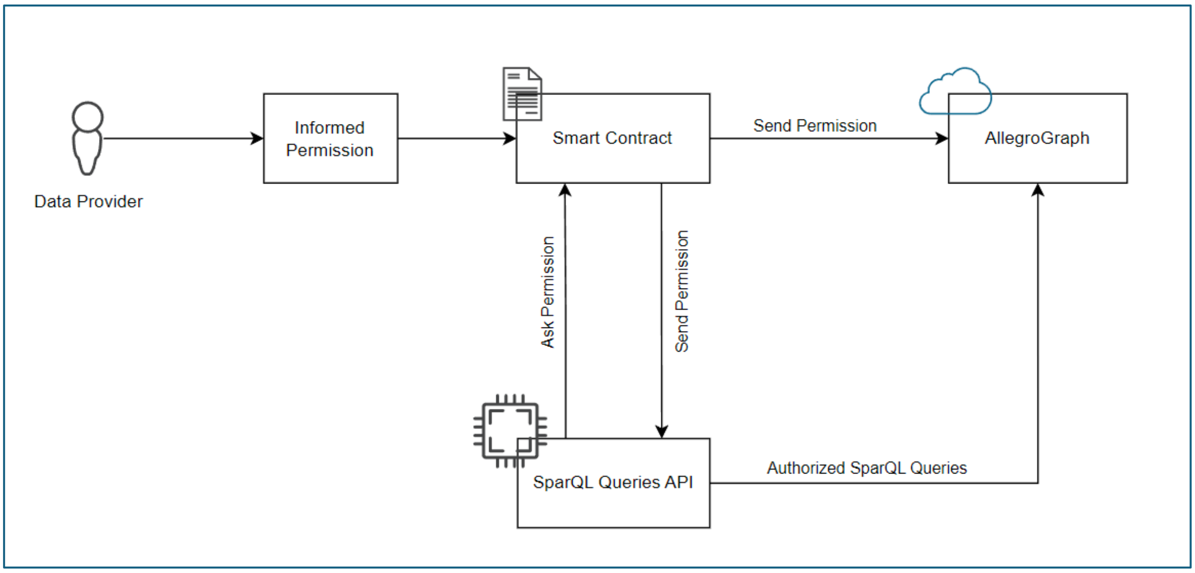


Figure 4 This figure explains the complete process of ensuring data ownership

5.3.2. FAIRification Progress

The next huge step in the proposed architecture is FAIRification which comes after ownership (o) has been dealt with in the previous section. It has laid a solid foundation of secure single ownership and controlled access secured by smart contracts to turn raw data into a FAIR-compliant state. FAIRification is a systematic and semantic modelling process to structure the raw data in ways leading to its being machine-readable and published for secure and compliant access. This step ensures that the data indeed becomes FAIR. The Figure 5 shows a very detailed FAIRification process, which will serve as a guideline for this section.

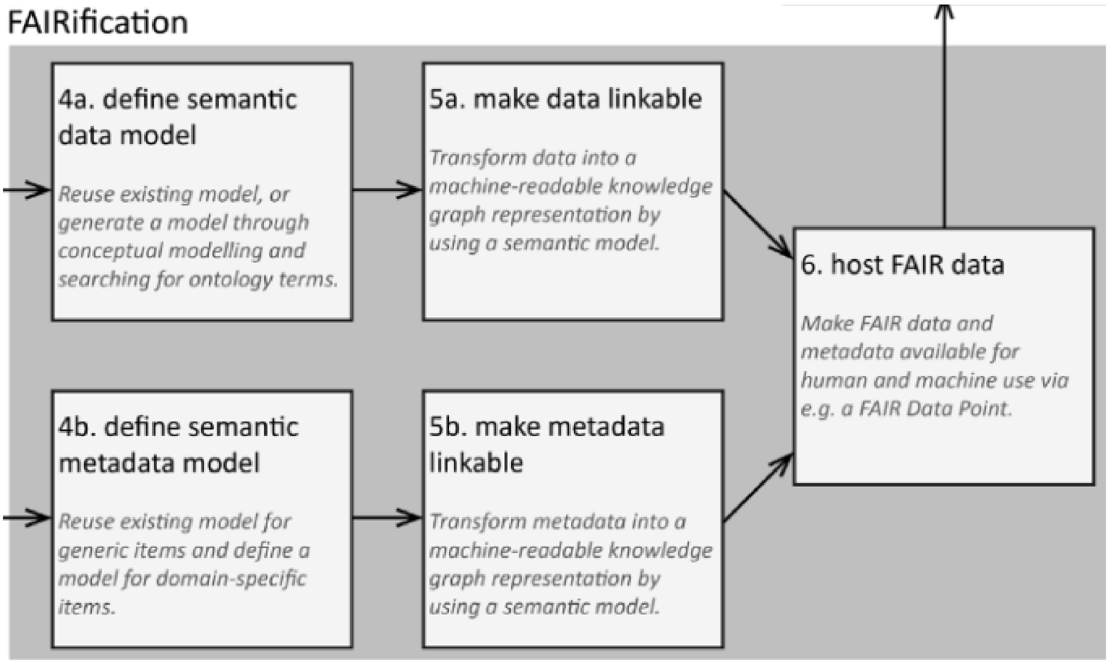

FAIRification
4a. define semantic data model
Reuse existing model, or generate a model through conceptual modelling and searching for ontology terms.
5a. make data linkable
Transform data into a machine-readable knowledge graph representation by using a semantic model.
6. host FAIR data
Make FAIR data and metadata available for human and machine use via e.g. a FAIR Data Point.
4b. define semantic metadata model
Reuse existing model for generic items and define a model for domain-specific items.
5b. make metadata linkable
Transform metadata into a machine-readable knowledge graph representation by using a semantic model.

Figure 5 A generic step-by-step workflow for the process of making data FAIR (Jacobsen, Kaliyaperumal, et al., 2020)

Data Input

The data provider kicks off the process by feeding the raw data to the system, and this step transpires to be a crucial part of the process because it marks the origin of the data's journey towards FAIR compliance. The data is submitted into the system using structured templates, such as a CEDAR template, to ensure conformity to a given structure. After going through standardized input, that stage is necessary for the further steps of the FAIRification process, where the data is ready for semantic modelling and other forms of processing. This architecture will adopt CEDAR template as a tool to turn raw data into RDF data.

Defining the Semantic Data Model

Once the data have been inputted, the next step will be an initial definition of the semantic data model. The semantic data model is designed to ensure that there is a two-fold capacity for data not only to be structured but also to carry connotations defined by rule-based terminologies and classifications. At this time, the system calls on tools like BioPortal to search for ontology terms or conceptual models that could most appropriately be used to describe the data in question. If an existing one is not there, new domain-specific terms are created to represent the data appropriately. This guarantees the interoperability of the semantic data model, meaning that it can be combined readily with other datasets and understood in various systems.

Furthermore, a semantic model establishes links between data and external datasets, making it possible for machines to search and utilize the data across different contexts. Through a semantic model, the system ensures that data is interoperable and thus reusable in the broader data ecosystem.

Making Data Linkable

Linkable data is the next stage of the FAIRification process. Data transforms into a semantic machine-readable knowledge graph using discussions like RDF. The data is given a unique URI functioning as a permanent identifier which allows it to be referred to and linked everywhere.

The RDF format ensures that the data is not only structured but also linked to other relevant data sources, thus, making it more useful for cross-border data exchange. The transformed RDF data gets a URI and is put into a triple storage database like AllegroGraph. The database is cloud-based, allowing secure, scalable, and FAIR-compliant preservation and access to the data, and thus ensuring the ability of the data to be found and accessed.

Hosting FAIR Data

Step five entails the hosting of the FAIR data. This comprises making both the data and the related metadata available, which further allows use by people and machines. The FAIR Data Point will be important in this phase, since it is the portal that will allow authorized users or systems to access, query, and use the data. The data is stored in a way that will allow both the interoperability and reusability of the information in a larger context-based data ecosystem.

By hosting the data in FDP, it becomes available for various use cases-ranging from research to policymaking-while adhering to the regulatory approaches like PIPL and GDPR. Indeed, it will remain open to query through standardized APIs, thus ensuring its ongoing potential for access and use in a safe and possibly compliant way.

The Figure 6 explains how this section solve the FAIRification and shows the technologies in this process.

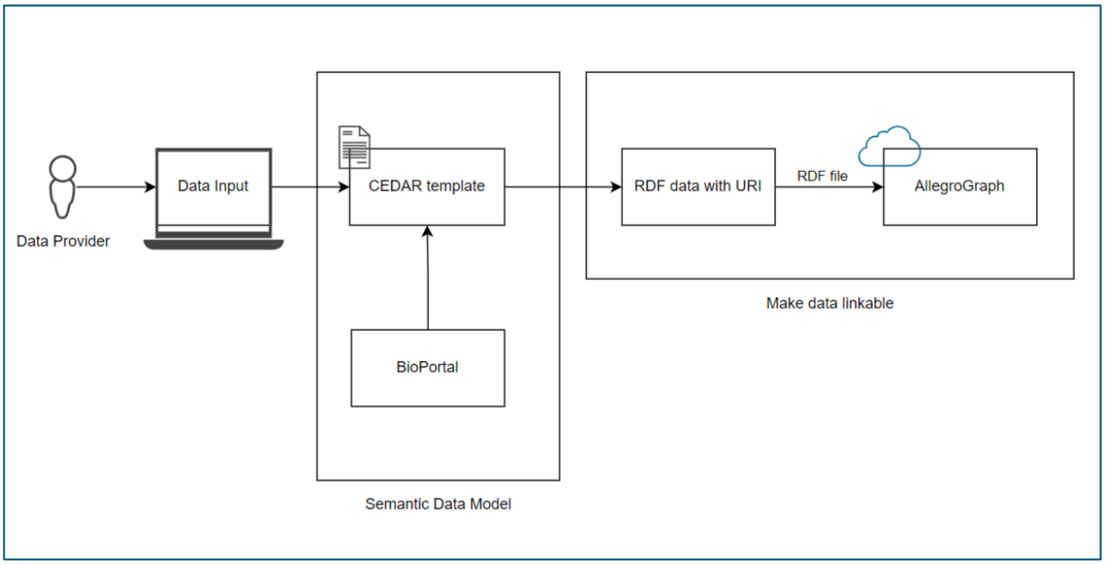


Figure 6 This figure illustrates the process of making raw data FAIR

5.3.3. Ensuring Localisation (L) and Regulatory compliance (R) in the Model

This section explains how the much-discussed FAIR-OLR architecture ensures localisation and compliance with particular reference to PIPL in China and GDPR of the EU. These matter immensely because such considerations keep the data safe within the ambit of data localization laws.

PIPL stands for Personal Information Protection Law in China which mandates data localization as it must operate at least in some geographical boundaries. Data coming from China must be locally kept away from upload and storage into AllegroGraph.

As described in the proposed architecture, compliance with the data protection rules within the concerned jurisdiction, PIPL in China, and GDPR in the EU is built into all processes. More specifically, several mechanisms under the architecture ensure that data handling, storage, and sharing are compliant with the national regulatory framework in place.

A key component of this architecture is the Security Assessment, which is specifically tailored to comply with China's data protection requirements under PIPL. This assessment process ensures that any data transfers, especially cross-border data flows, are conducted securely and in compliance with PIPL regulations.

5.3.4. The complete Cross-border data exchange architecture between China and EU

These components represent the basis of the Cross-border Data Exchange Architecture, established between China and the EU, for secure and compliant data exchange between data providers working across different borders. The complete Cross-border Data Exchange Architecture can be found in Figure 7.

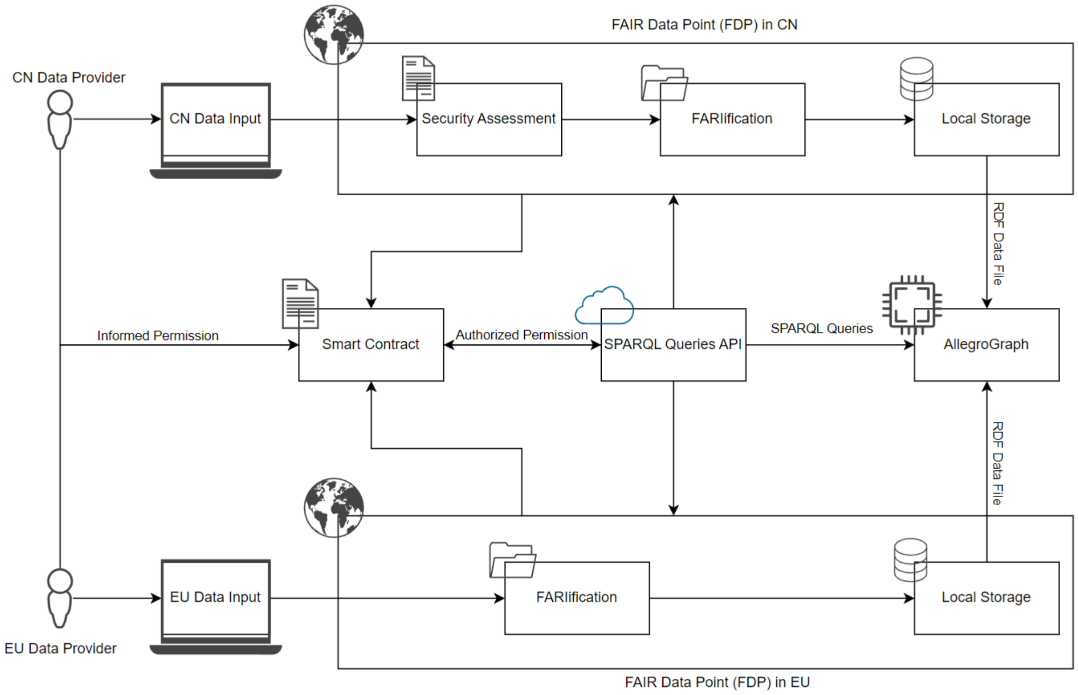


Figure 7: Cross-border data exchange architecture between China and EU

6. Research finding 3: Implementing the proof of concept for Cross-border data exchange architecture between CN and EU

The first step towards gauging the viability of the proposed architecture will involve making the reference architecture operational along with a gathering of critical data input for assessing the performance and legal compliance. The key objective is to ensure that this architecture can provide means for the cross-border data exchange to take place safely, efficiently, and in compliance with stringent laws.

To ensure this, this requires establishing all data models and data exchange mechanisms to be fully FAIRified, becoming machine-actionable and compliant with Findable, Accessible, Interoperable, and Reusable principles.

The development of a suitable CEDAR template for data entry begins with: CEDAR template is an indispensable tool that ensures the consistency of any data set through a standard format that allows linking them to relevant ontologies. The Figure 8 shows the example of creating a CEDAR template.

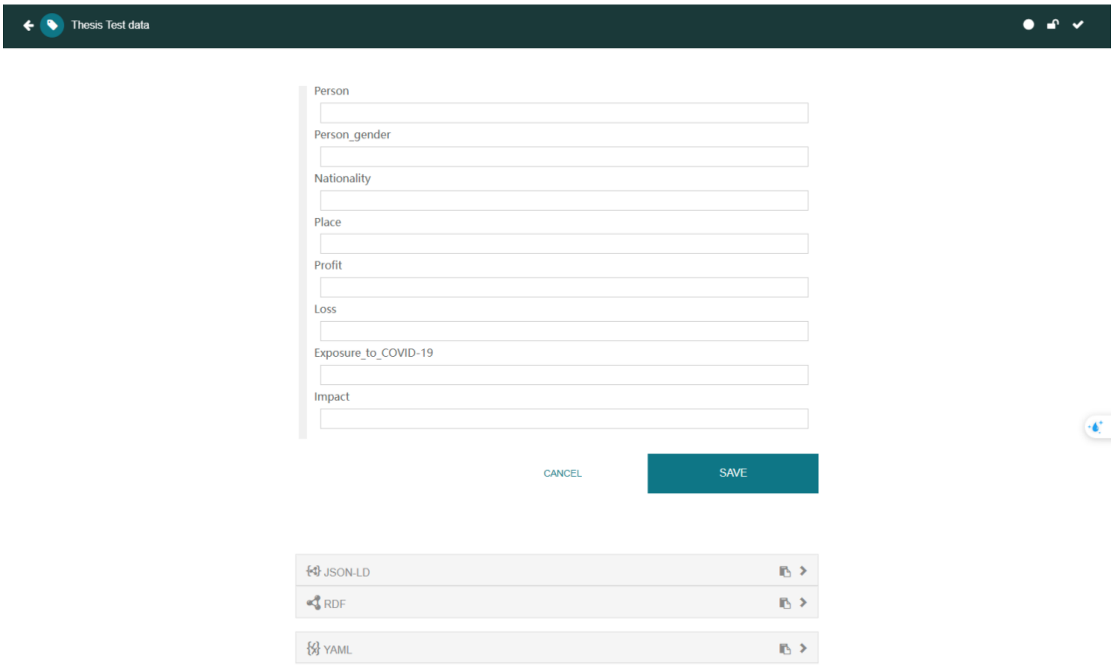
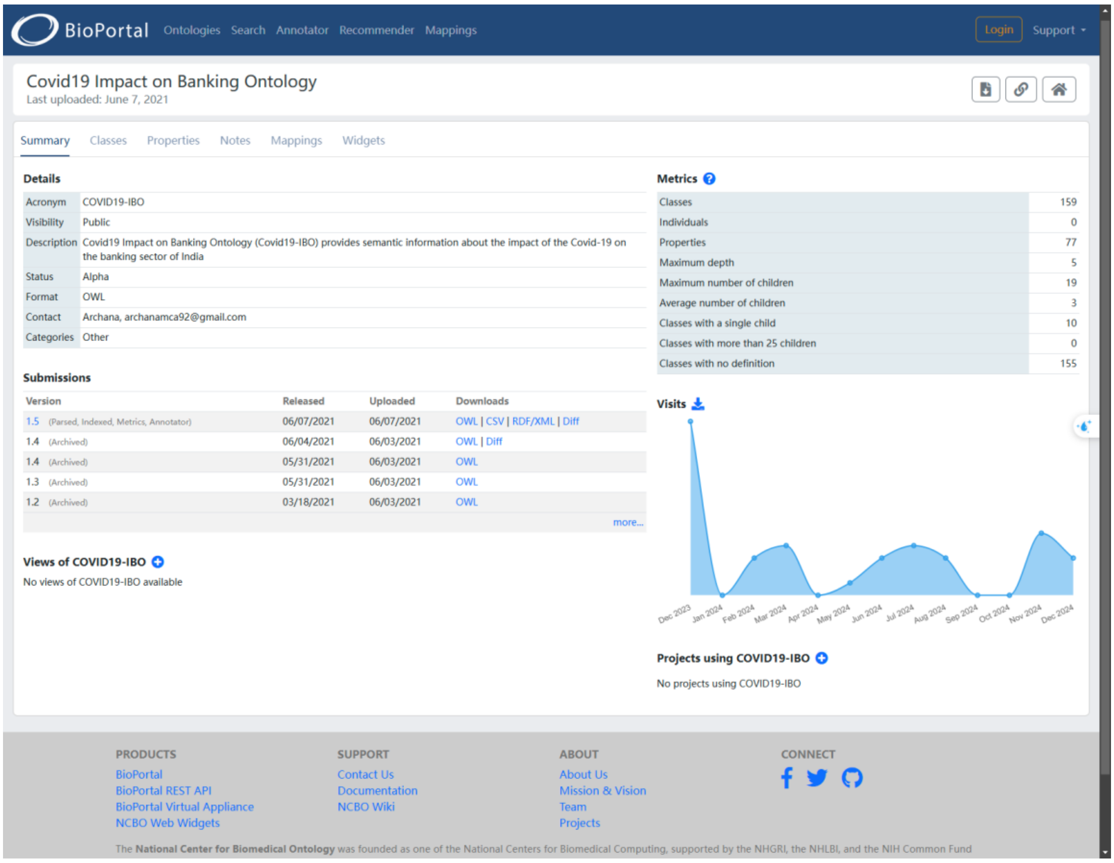


Figure 8 This figure illustrates the process of creating a suitable CEDAR template

After creating the template, we will identify the relevant BioPortal ontologies to enrich the data. BioPortal is a repository that houses a wide variety of biomedical ontologies, which will be used to enhance the metadata of the input data, making it more specific and semantically rich. In Figure 9, it shows an ontology in BioPortal that will be used in the implementation.



Covid19 Impact on Banking Ontology
Last uploaded: June 7, 2021

Details

Acronym	COVID19-IBO
Visibility	Public
Description	Covid19 Impact on Banking Ontology (Covid19-IBO) provides semantic information about the impact of the Covid-19 on the banking sector of India
Status	Alpha
Format	OWL
Contact	Archana, archanamca92@gmail.com
Categories	Other

Submissions

Version	Released	Uploaded	Downloads
1.5 (Parsed, Indexed, Metrics, Annotator)	06/07/2021	06/07/2021	OWL | CSV | RDF/XML | Diff
1.4 (Archived)	06/04/2021	06/03/2021	OWL | Diff
1.4 (Archived)	05/31/2021	06/03/2021	OWL
1.3 (Archived)	05/31/2021	06/03/2021	OWL
1.2 (Archived)	03/18/2021	06/03/2021	OWL

more...

Views of COVID19-IBO

No views of COVID19-IBO available

Metrics

Classes	159
Individuals	0
Properties	77
Maximum depth	5
Maximum number of children	19
Average number of children	3
Classes with a single child	10
Classes with more than 25 children	0
Classes with no definition	155

Visits

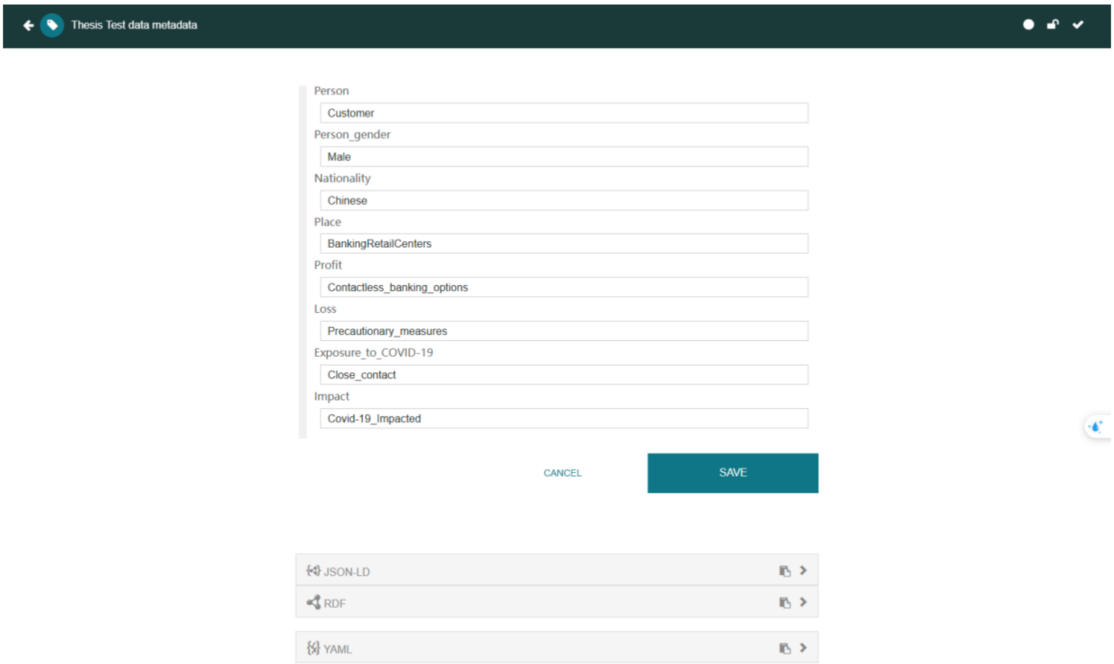
Projects using COVID19-IBO

No projects using COVID19-IBO

The **National Center for Biomedical Ontology** was founded as one of the National Centers for Biomedical Computing, supported by the NHGRI, the NHLBI, and the NIH Common Fund

Figure 9 This figure is the proof of how to find a suitable ontology in BioPortal and connect with CEDAR template

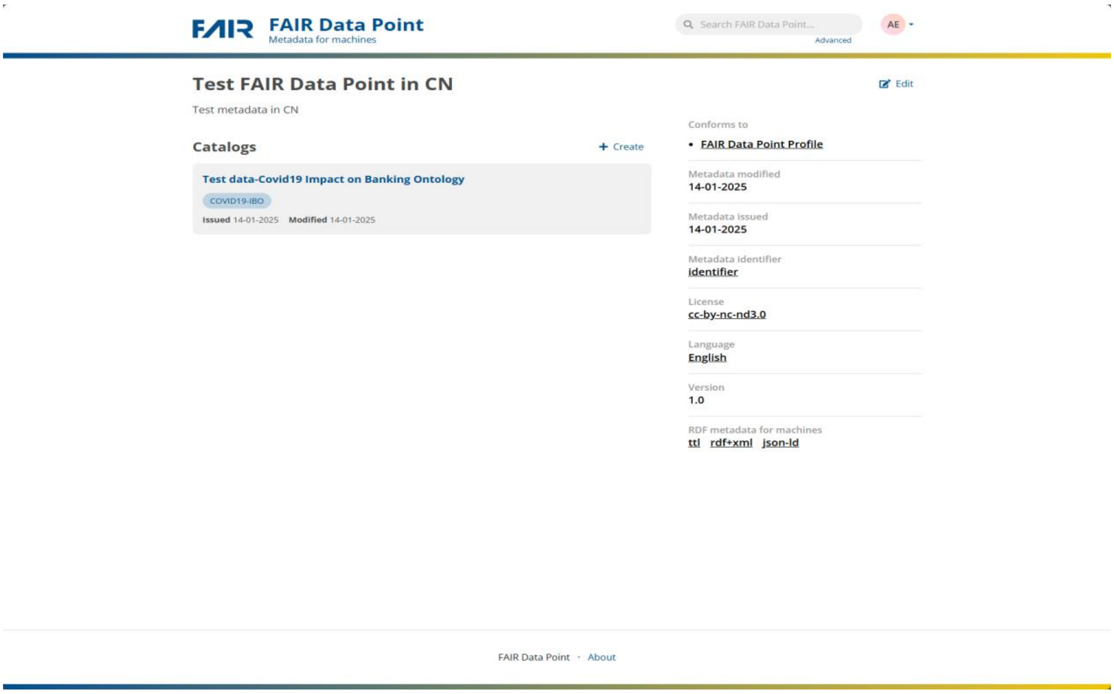
Once the data is linked to the appropriate ontologies, it will be input into the CEDAR template, ensuring it is structured and ready for the next phase. And the Figure 10 serves as an example of data input in CEDAR template.



Person
Customer
Person_gender
Male
Nationality
Chinese
Place
BankingRetailCenters
Profit
Contactless_banking_options
Loss
Precautionary_measures
Exposure_to_COVID-19
Close_contact
Impact
Covid-19_Impacted

Figure 10 This figure illustrates the process of loading data into the CEDAR template

The data will then be saved locally, and we will establish the FDP in China. This FDP will serve as the local interface for securely storing and managing the data. The Figure 11 shows the implement of FDP in China.



Test FAIR Data Point in CN

Test metadata in CN

Catalogs

Test data-Covid19 Impact on Banking Ontology
COVID19-IBO
Issued 14-01-2025 **Modified** 14-01-2025

Conforms to
- FAIR Data Point Profile

Metadata modified
14-01-2025

Metadata issued
14-01-2025

Metadata identifier
identifier

License
cc-by-nc-nd3.0

Language
English

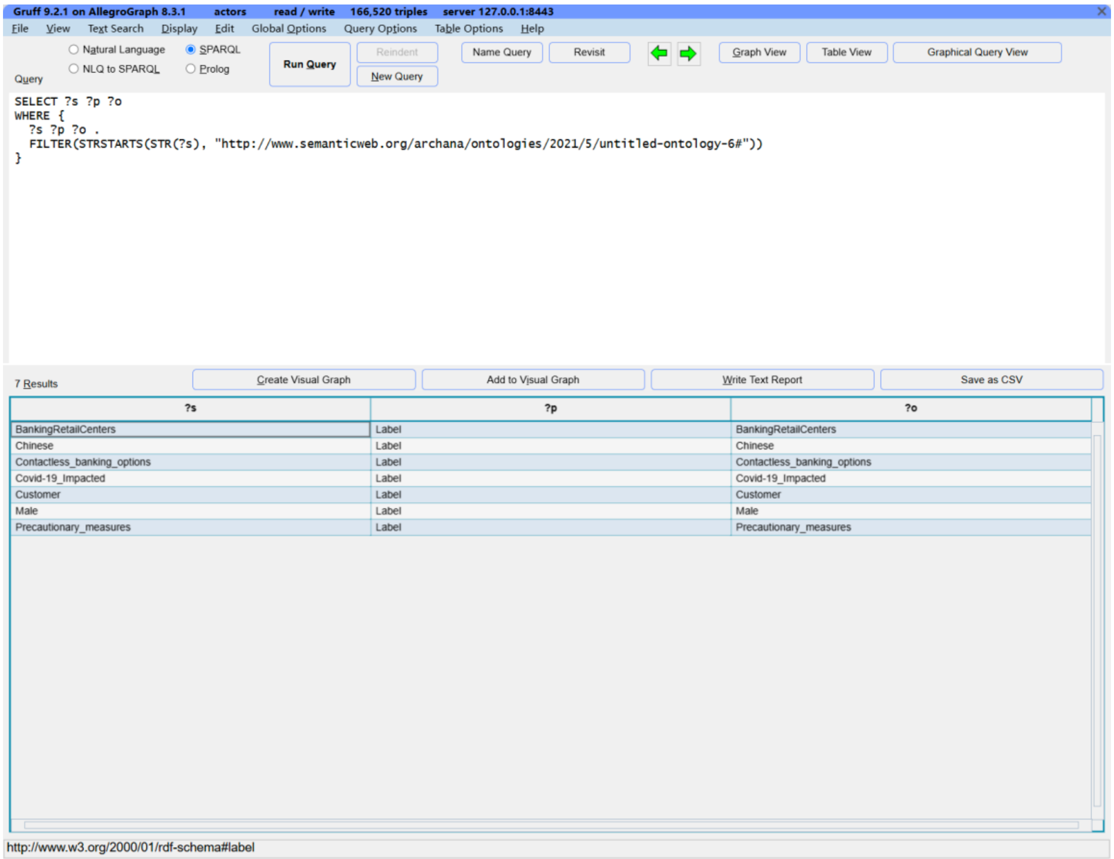
Version
1.0

RDF metadata for machines
ttl rdf+xml json-ld

Figure 11 The implementation of a FAIR data point in China

Next, we will link the FDP with AllegroGraph, ensuring that the data from both the EU and China are securely accessible and can be queried across jurisdictions in compliance with PIPL and GDPR. Once the FDP is connected to AllegroGraph, we will use the FDP interface to get the data and query the data through SPARQL in AllegroGraph.

The Figure 12 shows the result of SPARQL query in AllegroGraph. And the Figure 13 shows the visualization of AllegroGraph.



○ Natural Language ● SPARQL
○ NLQ to SPARQL ○ Prolog

Query

```
SELECT ?s ?p ?o
WHERE {
  ?s ?p ?o .
  FILTER(STRSTARTS(STR(?s), "http://www.semanticweb.org/archana/ontologies/2021/5/untitled-ontology-6#"))
}
```

7 Results

?s	?p	?o
BankingRetailCenters	Label	BankingRetailCenters
Chinese	Label	Chinese
Contactless_banking_options	Label	Contactless_banking_options
Covid-19_Impacted	Label	Covid-19_Impacted
Customer	Label	Customer
Male	Label	Male
Precautionary_measures	Label	Precautionary_measures

http://www.w3.org/2000/01/rdf-schema#label

Figure 12 The SPARQL query request and result in AllegroGraph

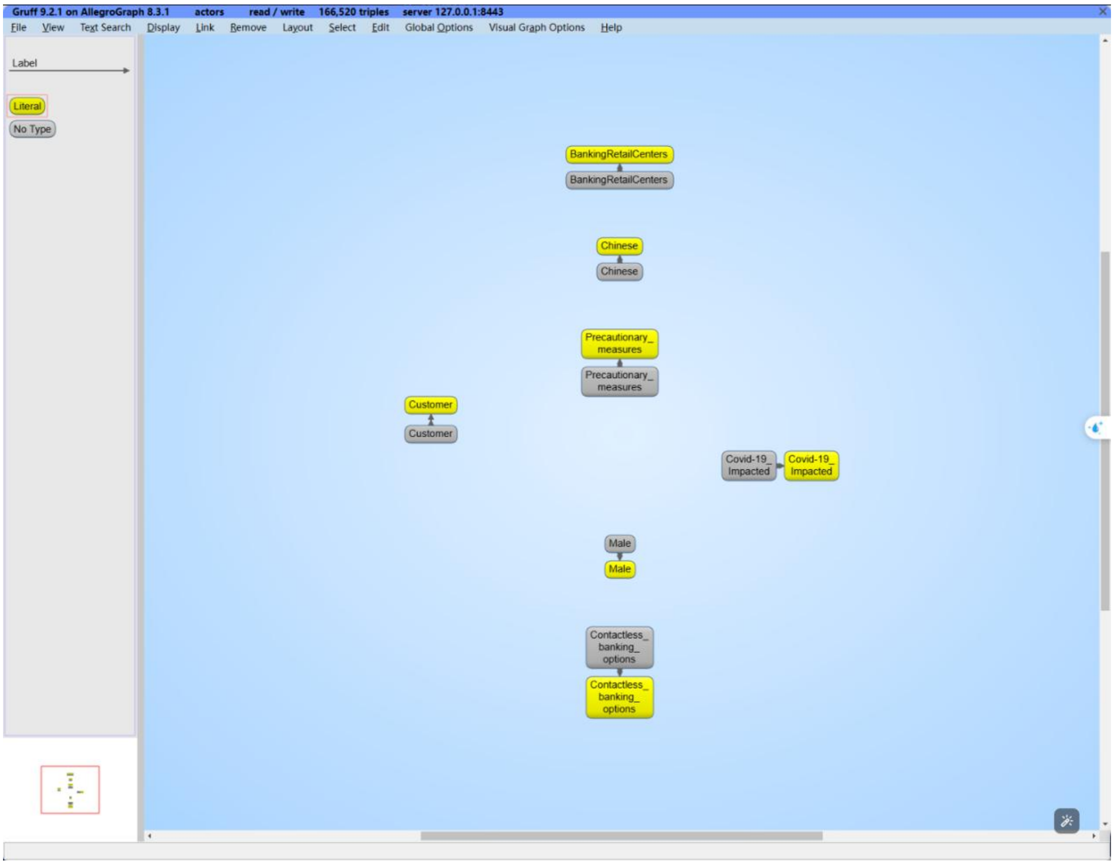

Gruff 9.2.1 on AllegroGraph 8.3.1 actors read / write 166,520 triples server 127.0.0.1:8443
File View Text Search Display Link Remove Layout Select Edit Global Options Visual Graph Options Help
Label
Literal
No Type
BankingRetailCenters
BankingRetailCenters
Chinese
Chinese
Precautionary_measures
Precautionary_measures
Customer
Customer
Covid-19_Impacted
Covid-19_Impacted
Male
Male
Contactless_banking_options
Contactless_banking_options

Figure 13 The visualization of query result.

7. The key metrics to assess the architecture's performance

To evaluate the performance of the FAIR-OLR-based architecture, it’s essential to identify key metrics that assess how well the system adheres to the Ownership (O), Localization (L), and Regulatory Compliance (R) principles, while also ensuring the data remains FAIR (Findable, Accessible, Interoperable, Reusable). Below are some key metrics that can be used to assess the architecture's performance in each of these areas:

7.1.1. Ownership (O) Metrics

Data Consent Rate: The percentage of data providers who have granted informed consent for their data. A higher rate indicates effective implementation of informed permission mechanisms and proper ownership management.

Smart Contract Compliance: The number of data access requests successfully handled through smart contracts, ensuring that only authorized users can access the data according to the predefined rules set by the data owners.

Audit Trail Completeness: The extent to which every data transaction (e.g., access, modification, sharing) is logged and traceable. This helps ensure that ownership control is transparent and properly managed.

7.1.2. Localization (L) Metrics

Data Residency Compliance: The percentage of data that is stored in compliance with localization requirements (e.g., PIPL for China and GDPR for the EU), depending on whether the data is stored correctly according to the jurisdictions that collected the data.

Cross-Border Data Transfer Incidents: Unauthorised transfers of data across borders; their frequency shows failures in enforcement of localization rules.

Data Storage Availability by Region: The availability of local storage solutions in both regions (China and the EU), ensuring that every jurisdiction should have its own data storage resources.

7.1.3. Regulatory Compliance (R) Metrics

Regulatory Compliance Rate: The percentage of data handling and data-sharing processes that are in accordance with relevant regulations (e.g., PIPL and GDPR). This may be tracked through regular audits and legal checks.

Security Assessment Success Rate: The number of successfully conducted security assessments for cross-border data transfer where the requirements for data protection and security in accordance with PIPL and GDPR are adhered to.

Data Access Audits: Frequency and scope of data access audits to ensure that data sharing practices conform to requirements of consent and authorization from PIPL and GDPR.

7.1.4. System Performance Metrics

Query Execution Time: The average time taken to process and return results for SPARQL queries from the FAIR Data Points (FDP), which gives a good idea of the overall capability of the system to handle larger data volumes.

Scalability: The capability of the system to cope with an increasing amount of data or user queries with little or no performances degradation, hence emphasizing on the robustness of the architecture as it scales.

System Uptime: The percentage of time when FAIR Data Points (FDPs) and the database running it and/or one or more database or software that support it (such as AllegroGraph) are available and usable. This depicts the degree of dependability of the system.

7.1.5. User Satisfaction and Engagement Metrics

The user adoption rates provide information on the number of data providers, health facilities, and other users actively using the FAIR-OLR-based architecture. The higher percentage of adoption will signify the effectiveness of the system to satisfy its users.

The feedback received from data providers is an indication of satisfaction regarding the ease of use of the system, the transparency of ownership mechanisms, and adherence to the principles of data protection laws.

The frequency with which users perform queries through the SPARQL interface shows the effectiveness of data sharing and how effective the system is in line with the FAIR principles.

These metrics allow us to consider how well the architecture performs in terms of Ownership, Localization, and Regulatory Compliance. They will help measure the extent to which the system performs in real-life applications to show that the Cross-border Data Exchange Architecture is technically robust and aligned with legal requirements.

8. Discussion

8.1. The issue and what we have researched

The growing need for exchange of data across borders, particularly with respect to health data containing sensitive information about individual patients, has brought forth the very many intricacies involved when complying with divergent national regulations. This is even more manifest in the case of PIPL and GDPR, in so far as both of them provide for stringent measures related to data localization and cross-border data transfer. In particular use cases involving healthcare data, wherein privacy and security are prime considerations, these very rules build the barriers for frictionless international data exchange.

The question that this research addresses is of how to facilitate exchange that is secure, compliant, and efficient among healthcare systems in China and the EU, all the while ensuring ownership, localization, and regulatory compliance with respect to sensitive healthcare data. The research seeks to explore how FAIR principles can be used to help break down these barriers, and assure PIPL and GDPR compliance as well as usefulness of health data for cross-jurisdictional purposes.

To address this issue, the research, apart from providing a robust machination certified by FAIR-OLR, deals with data ownership (O), data localization (L), and regulatory compliance (R) while doing so in accordance with FAIR data principles in such a manner that breach of data rights and exchange between regions in varied legislative settings will not challenge interoperability or legal data exchange anymore and will promote such exchange in a secure, legal, and efficient manner.

8.2. The key findings

This research has shown that it is feasible and practical to develop a Cross-Border Data Exchange Architecture for the secure, compliant, and efficient exchange of data between China and the EU. The architecture incorporates a FAIR-OLR framework to ensure that the exchange of data is legally compliant and cross-jurisdictionally accessible.

One of the prominent findings is a deep exploration of China and the EU's data policy frameworks, leading to distinctive regulatory standpoints toward cross-border information transfer. Regarding China, it is legislatively mandated by PIPL that strict data localization and security assessment be conducted for any sensitive data in transferring across borders to ensure data remains within the confines of China unless certain conditions allow it. In comparison, GDPR adopts a more flexible framework in allowing cross-border data flow on the basis of SCCs or adequacy decisions while continuing to emphasize protection of data privacy.

The contrasting regulatory framework comparison highlighted major differences: on the one hand, the PIPL with intensive focus on data localization and data sovereignty explanations; on the other carton, the GDPR stressing the data subjects' right with a more flexible approach possible for cross-border data exchange. The differences became very fundamental in determining the construction of FAIR-OLR-based architecture which needed to balance both legal requirements while maintaining data ownership, localization, compliance, etc.

An important finding here is the successful implementation of the FAIR principles into architecture. The architecture ensures that through tools such as CEDAR templates and linking data to relevant BioPortal ontologies, the healthcare data are converted into a FAIR-compliant format. This enables the data to be discoverable, accessible, and interoperable. The architecture thus supports cross-border data exchange as well as compliance with regulatory requirements. The process of FAIRification is critical to ensuring safe data sharing and reuse in a manner that upholds the requirements of the PIPL and GDPR.

The architectural implementation of Ownership (O) through informed consent and use of smart contracts allows data providers the control of their data, ensuring PIPL and GDPR's requirements for explicit consent and data protection are fulfilled. Automation of data access rules is enforced via smart contracts for compliance and transparency.

The architecture, too, ensures Localization (L) through the local enrollment of RDF data before uploading to the AllegroGraph. This speaks to the local storage requirement of the PIPL and the secure cross-border transfer provision of the GDPR. Security assessments are conducted any time sharing of the data crosses borders.

Finally, Regulatory Compliance (R) encompasses compliance with PIPL and GDPR in relation to data handling processes. The security assessments, SCCs, and smart contracts provide a means for secure data transfer mechanisms to ensure compliance with privacy protection requisites by both regions and enable legal cross-border data exchange.

In conclusion, the findings confirm that architecture based on FAIR-OLR can be taken as a possible solution to solve cross-border data exchange challenges between China and the EU. By combining the FAIR principles into Ownership, Localization, and regulatory compliance, this architecture ensures data remains secure, compliant, and ready for global sharing while respecting data sovereignty in both jurisdictions.

8.3. How these findings can be interpreted

The growing demand for cross-border data transfer creates an urgent need for solutions to help bridge existing discrepancies in data policies across countries and regions. These researches demonstrate the essential role of drawing upon the FAIR principles and FAIR-OLR to address these challenges, with a specific emphasis on cross-border data transfer challenges between China and the EU. With the study comparing their respective policies, it enters such a framework that can effectively ensure both legal approval and the secure moving of sensitive data while observing data sovereignty, PIPL, and GDPR requirements.

8.3.1. Understanding the Regulatory Landscape

Within the Chinese context, however, PIPL is the paramount regulation governing data protection and the transfer of personal information across borders. It calls for a very strict data localization requirement, thereby preventing sensitive data from being moved out of China unless stipulated conditions-obtaining a security assessment-are met. In short, their main focus on data sovereignty means the data controllers and organizations in China must comply with not just the local data storage and processing laws; they are also further constrained by the requirements of Chinese regulatory framework when it comes to cross-border data transfer. The establishment of Security Assessment as a mandatory function demonstrates China's caution with international data flows, particularly in regard to sensitive personal information.

Conversely, the GDPR provides a comprehensive uninterfered regulatory framework for data protection at EU-wide levels. While the PIPL would create strict localization of data, the GDPR emphasizes data protection rights and permits international data transfer International data transfer to be done legally only if the receiving states afford such an adequate mechanism; that would include various other obstacles for making sure that data is fully protected, such as Standard Contract Clauses or Adequacy Decisions. The realization of that is that it into a more flexible organization of the cross-border data transfer compared to PIPL. But it, again, emphasizes some heavyweight legal guarantees for individuals' rights.

8.3.2. Comparative Analysis of Data Policies

In comparing the data policies of China and the EU, one can see crucial differences in their approach toward cross-border data transfer. China, through its PIPL, enforces data localization to ensure that sensitive personal data remains in the country, paying attention to data sovereignty and national security. Such a policy requires organizations to conduct security assessments prior to the cross-border movement of data, thus establishing a controlled mechanism on cross-border data transfer.

On the contrary, the GDPR allows for cross-border data transfer under certain conditions, guaranteeing that the rights of data subjects remain valid. This system provides for use of mechanisms such as SCCs and adequacy decisions resulting in flexibility towards data transfer while upholding privacy protections. The differences in the values attached to sovereignty over data orders on one side and fundamental human rights of data protection on the other is illustrated in such an approach.

8.3.3. The Role of FAIR-OLR in Solving Cross-Border Data Challenges

The FAIR-OLR framework offers a practical solution that integrates the FAIR principles with Ownership, Localization, and Regulatory Compliance components that would keep the data secure, accessible, and compliant with respective national and international data privacy laws. So, why FAIR-OLR?

The choice of FAIR-OLR is predicated upon the need for specific regulatory requirements based on PIPL in China and GDPR in the EU. Ownership gives data provision providers the right to control access to their data, which offers clear consent mechanisms in line with the PIPL emphasis on informed consent and GDPR's focus on data subject rights. Localization ensures that it takes care of PIPL's data localization measures, thereby ensuring that data does not leave Chinese jurisdiction, except in accordance with certain regulations, which allow cross-border data sharing under certain regulated conditions. Finally, Regulatory Compliance ensures that all processes are in conformity with strict requirements for cross-border data transfers, like GDPR SCCs or security assessments of PIPL, thus making sure that our research remains in compliance with all applicable laws.

By incorporating these elements into the FAIR-OLR framework, we allow for a flexible-scaled composition to any of those complexities with global data exchange. It gives findable and accessible data in a secure way, thus achieving privacy and then promoting interoperability.

8.3.4. Steps in building the Cross-Border Data Exchange Architecture

Implementation of the architecture is structured along several key milestones designed to ensure data ownership, localize data, and achieve regulatory compliance. In the initial phase, Data Input and FAIRification occurs: raw healthcare data is processed and transformed into a FAIR-compliant standard. This is done via the use of CEDAR templates that codify the data with relevant BioPortal ontologies which link the data sets with information about them and thereby make them semantically compatible across systems.

Data are FAIRified, then stored within local FDP which ensures that the localization requirement of data is abided by. Data is stored securely within the jurisdiction where it had been collected within China in order to comply with PIPL's stringent data localization regulations and the data protection provisions of the GDPR. At this stage, Security Assessments are performed for any potential cross-border data exchanges to ensure compliance with the requirements of any security agreement outlined by both PIPL and GDPR.

Subsequently, Smart Contracts and Informed Permission mechanisms insist upon Ownership throughout the process. Data providers consent explicitly to the use of the data; Smart Contracts will automatically enforce these permissions thus allowing only authorized parties access to the data. This automated method helps streamline data sharing while maintaining accountability regarding the data which matches with PIPL emphases on informed consent and GDPR's explicit consent requirements.

After the local storage is in place and regulatory frameworks have chosen to comply, data can be accessed and queried by means of the SPARQL the use of AllegroGraph, a graph database specifically optimized for storing and querying RDF data. AllegroGraph allows the data to be queried in an efficient manner, thereby assuring her findability and interoperability across systems. Data can be queried securely through this mechanism, following both the PIPL and GDPR laws regarding access to data.

8.4. Limitation of this research

Although the FAIR-OLR-based architecture proposed in this research provides a reasonable solution to the problem of cross-border data exchange between China and the EU, this, however, is limited in its implementation and application in reality against various unexpected factors.

8.4.1. Implementation Challenges

Inherent to this research is a disconnect between the conceptual architecture and its real-world implementations. Although designed to comply with the requirements of the PIPL and GDPR, real-world go-ahead in live healthcare systems has not been explored. Challenges may arise in implementing the architecture in real-world scenarios-engineering its integration with legacy systems that may not be fully capable of supporting the RDF format or the SPARQL query system accommodated in the architecture. Significant technical and logistical challenges could also arise in creating FDPs and ensuring compliance with both data localization standards of PIPL and cross-border data transfer requirements of GDPR.

8.4.2. Regulatory and Legal differences

Even though PIPL and GDPR are considered, the regulatory climate in the country and EU is in knelt uncertain movement. PIPL and GDPR Updates and amendments since their promulgation may witness other amendments in the future, with the importance of revising the compliance of the FAIR-OLR architecture. Promotion data sovereignty and data sovereignty are still very much in a gray area, a lot still needs to be defined. The nuances behind PIPL could well bring on interpretations of GDPR's adequacy decisions or Standard Contractual Clauses through an uneven process that may frustrate architect's adaptability.

More so, there are a number of differences between the rights of the data subjects under the GDPR and the mode of enforcement of the PIPL, yet the combined effect leaves both of them as cumbersome statutes that may require continuous changes to the architecture in an attempt to comply with the governing laws. Therefore, it is one other challenge to design upon that would keep modifying and adopting to the legal changes accordingly.

8.4.3. Technical Limitations

Although the proposed architecture integrates state-of-the-art technologies such as smart contracts, SPARQL queries, and AllegroGraph, there are inherent technical limitations. Implementation of the blockchain and smart contracts to manage data ownership and access control is still in an emergent stage where the scalability or integration in large-scale production applications has yet to be adequately tested. This will add complexity in a situation where implementations of distributed ledger technology for automated permissions and compliance would be new territory for systems not yet equipped to handle blockchain solutions in an efficient manner.

AllegroGraph and FDP may also face challenges in large-scale deployments and interoperability issues with legacy systems that do not apply or conform to RDF standards. This ensures that the greatest conflicts among respondents to visit different healthcare systems do not altogether make sure of its disparate data formats or standards.

8.4.4. Cost and Resource Implications

The FAIR-OLR architecture requires big investment in infrastructure and maintenance. The operational cost of implementing FDP in either region may entail great costs in data storage, data processing, and security assessments for compliance under PIPL and GDPR. On the whole, the financial and technical demands may turn small or medium-sized organizations away from adopting the architecture into wider use.

Further, there are several unignorable educational and training requirements for healthcare and data-providing professionals, enabling them to use the FAIR-OLR architecture with full cognizance. Coupled with the need for continuous compliance under regulatory checks and data reviews, operational costs could increase and render sustainability of the architecture in a less resource-rich environment tricky.

8.5. Future work

Although the FAIR-OLR-based architecture does seem to suggest security and compliance possibilities for cross-border data exchange between China and the EU, there is still much research and development to be done. Future work will focus on addressing the limitations identified in the current research and improvement of scalability, flexibility, and adaptability of the architecture when applied to real-world scenarios.

8.5.1. Real-World Implementation and Test Projects

The next critical step would be implementing the FAIR-OLR architecture in real-world hospitals. While this work presents a cognitive architecture, the live testing of it will be crucial to checking its feasibility and effectiveness. Test projects for actual data exchange between providers of China and the EU should be undertaken to identify challenges that might arise during the integration process. This phase would further concern technical and regulatory issues, such as integration with legacy systems, scalability for the FDP, and ensuring smooth functioning of cross-border data sharing across jurisdictions. With successful pilot-testing, refinements can be made, aiding the case for a broader adoption of the architecture.

8.5.2. Continuous Adaptation to Regulatory Changes

Since PIPL and GDPR are constantly evolving, an immediate configuration is to set up adaptive mechanisms to align the architecture with the regulations. This might include using automated control tools to address legal prescriptions by monitoring updates and inserting them in the architecture. For instance, the conditions for cross-border data transfer provisions can alter under the Control of GDPR, and in relation thereto, new security requirements or localization expectations may also enter into force under PIPL. This might also mean flexibility in establishing the architecture to accommodate such changes without in any way disrupting processes of data exchange-which is important for continuing in compliance.

8.5.3. Improving Interoperability and Standardization

On one hand, the architecture integrates RDF and SPARQL to facilitate data exchange; on the other hand, the work is traditionally recognized for further advancing the system's interoperability in the coming years, especially when dealing with various datasets, all possibly utilizing a different standard. Developing and adopting more standardized metadata models and ontologies to provide effective solutions in integrating data would bring a lot of improvement. Increased cooperation with organizations developing international standards could also serve to improve standardization in data exchange protocols and ontologies and facilitate wide adoption of the architecture across multiple sectors in healthcare and elsewhere.

8.5.4. Leveraging Advanced Technologies

The use of advanced technologies such as AI, intelligent learning, and federated analytics is another place that would need future exploration. AI can enhance data analysis, prediction, and decision-making, while federated learning could allow for a collaborative effort in machine learning around the globe without compromising data privacy. With decentralized data processing, federated learning can meet the localization requirement of PIPL and GDPR and allow sensitive data to remain under the jurisdiction while still benefiting global insights. Their application would take the interoperability, scalability, and efficiency of the whole system a notch higher, thus increasing its power in real-world applications.

8.5.5. Addressing Cost and Resource Implications

The highlights suggest that the FAIR-OLR architecture requires great resources to implement and maintain. Future work should research ways to reduce the operational costs and resource needs of the system. This could involve establishing budget-friendly cloud-based solutions for FDP initiatives or devising partnerships with public health or governmental organizations to share operational costs. Moreover, the usability and accessibility of the architecture for smaller healthcare providers would be instrumental in cementing the providing framework for scaling the system for larger adoption, especially in areas with lesser means of livelihood.

9. Conclusion

This thesis presented a Cross-border Data Exchange Architecture that integrates the FAIR principles with the Ownership, Localization, and Regulatory Compliance (OLR) framework to address the regulatory and technical challenges of cross-border data exchange between China and the EU. The research aimed to explore how to facilitate secure and compliant data sharing in the context of healthcare data, particularly in light of the strict regulatory requirements of PIPL in China and GDPR in the EU.

The FAIR-OLR-based architecture developed in this research offers a comprehensive solution to these challenges by combining key components: Ownership (O), which ensures that data providers retain control over their data; Localization (L), which ensures compliance with PIPL's data localization requirements; and Regulatory Compliance (R), which guarantees compliance with both PIPL and GDPR for cross-border data transfers. Through the use of smart contracts, informed consent mechanisms, and security assessments, the architecture was supposed to respect the legal frameworks of both jurisdictions while promoting interoperability and findability of data.

Research clearly indicates that the architecture based on FAIR-OLR is a tried-and-true method of managing sensitive healthcare data across borders. By combining principles from FAIR and OLR, the architecture includes a sound and legally compliant structure that keeps data accessible, usable, and protected while allowing for appropriate cross-border sharing and collaboration. The application of FAIRification processes, along with the deployment of RDF and SPARQL queries through AllegroGraph, has shown data can easily and compliantly be queried and transferred across jurisdictions without compromising on data protection laws.

However, limitations highlighted include the need for real-world testing, the impossibility of supporting continuous change due to evolving nature of regulations, and ultimate difficulties in integrating blockchain and smart contracts in healthcare systems. Such limitations espouse the necessity for further testing projects and dynamic adaptation to the regulatory labyrinth in order to allow the construction to remain functional and compliant as the law evolves.

For the future work, testing the architecture within the live healthcare settings of test studies between China and the EU would impart a practical approach towards possible implementation problems. The architecture of the model may also be enlarged, interoperable, and data-privacy-protective through AI, federated learning, and other advanced technology integrations. The architecture must be integrated in an economical and sustainable manner toward wider adaption, the latter particularly for smaller healthcare providers which usually are constrained in their resources.

The thesis hundreds of them address a strong foothold for cross border data exchange in healthcare, providing an architecture that has scalability and compliance to the two respective areas under consideration: China and the EU. The combined manifestation of FAIR principles with a law compliance aspect through the proposed architecture indicates an initial step toward establishing secured, streamlined, and lawful data sharing within the global healthcare ecosystem.

10. References

Almada, M., & Radu, A. (2024). The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy. *German Law Journal*, *25*(4), 646–663. https://doi.org/10.1017/glj.2023.108

Bakare, S. S., Adeniyi, A. O., Akpuokwe, C. U., & Eneh, N. E. (2024). DATA PRIVACY LAWS AND COMPLIANCE: A COMPARATIVE REVIEW OF THE EU GDPR AND USA REGULATIONS. *Computer Science & IT Research Journal*, *5*(3), Article 3. https://doi.org/10.51594/csitrj.v5i3.859

Bansal, G., & Warkentin, M. (2021). Do You Still Trust?: The Role of Age, Gender, and Privacy Concern on Trust after Insider Data Breaches. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *52*(4), 9–44. https://doi.org/10.1145/3508484.3508487

Bielialov, T., Kalina, I., Goi, V., Kravchenko, O., & Shyshpanova, N. (2023). Global Experience of Digitalization of Economic Processes in the Context of Transformation. *International Journal of Professional Business Review: Int. J. Prof.Bus. Rev.*, *8*(6), 10.

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, *26*(7), 931–936. https://doi.org/10.1038/s41431-018-0160-0

Bradford, A. (2020). The Brussels Effect. In A. Bradford, *The Brussels Effect* (1st ed., pp. 25–66). Oxford University PressNew York. https://doi.org/10.1093/oso/9780190088583.003.0003

Calzada, I. (2022). Citizens' Data Privacy in China: The State of the Art of the Personal Information Protection Law (PIPL). *Smart Cities*, *5*(3), Article 3. https://doi.org/10.3390/smartcities5030057

Cardoso Silva Ferreira, A., & Van Reisen, M. (2023). Hourglass-based interoperability through nanopublications in VODAN-A. *FAIR Connect*, *1*(1), 5–11. https://doi.org/10.3233/FC-221512

Casaletto, J., Bernier, A., McDougall, R., & Cline, M. S. (2023). Federated Analysis for Privacy-Preserving Data Sharing: A Technical and Legal Primer. *Annual Review of Genomics and Human Genetics*, *24*(Volume 24, 2023), 347–368. https://doi.org/10.1146/annurev-genom-110122-084756

Changing Data Policies in China: Implications for Enabling FAIR Data | SpringerLink. (n.d.). Retrieved 23 December 2024, from https://link.springer.com/chapter/10.1007/978-3-030-28061-1_28

Chen, J., & Sun, J. (2021). Understanding the Chinese Data Security Law. *International Cybersecurity Law Review*, *2*(2), 209–221. https://doi.org/10.1365/s43439-021-00038-3

Chen, M. (2024). Developing China's Approaches to Regulate Cross-border Data Transfer:Relaxation and Integration. *Computer Law & Security Review*, *54*, 105997. https://doi.org/10.1016/j.clsr.2024.105997

Chen, Y., & Song, L. (2018). China: Concurring regulation of cross-border genomic data sharing for statist control and individual protection. *Human Genetics*, *137*(8), 605–615. https://doi.org/10.1007/s00439-018-1903-2

Christen, E., Meyer, B., Oberhofer, H., Hinz, J., Kamin, K., & Wanner, J. (n.d.). *The Brussels Effect 2.0: How the EU sets global standards with its trade policy*.

Corrales Compagnucci, M., Aboy, M., & Minssen, T. (2021a). *Cross-Border Transfers of Personal Data after Schrems II: Supplementary Measures and New Standard Contractual Clauses (SCCs)* (SSRN Scholarly Paper No. 3951085). Social Science Research Network. https://doi.org/10.2139/ssrn.3951085

Corrales Compagnucci, M., Aboy, M., & Minssen, T. (2021b). *Cross-Border Transfers of Personal Data after Schrems II: Supplementary Measures and New Standard Contractual Clauses (SCCs)* (SSRN Scholarly Paper No. 3951085). Social Science Research Network. https://doi.org/10.2139/ssrn.3951085

Creemers, R. (2022). China's emerging data protection framework. *Journal of Cybersecurity*, *8*(1), tyac011. https://doi.org/10.1093/cybsec/tyac011

Creemers R. (2023). *Cybersecurity Law and Regulation in China: Securing the Smart State*. https://doi.org/10.1163/25427466-06020001

Cross-border data regulations in the European Union and South Korea. (n.d.).

da Silva Santos, L. O. B., Burger, K., Kaliyaperumal, R., & Wilkinson, M. D. (2023). FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication. *Data Intelligence*, *5*(1), 163–183. https://doi.org/10.1162/dint_a_00160

Demirer, M., Jiménez Hernández, D. J., Li, D., & Peng, S. (2024). *Data, Privacy Laws and Firm Production: Evidence from the GDPR* (Working Paper No. 32146). National Bureau of Economic Research. https://doi.org/10.3386/w32146

Edmunds, S. C., Li, P., Hunter, C. I., Xiao, S. Z., Davidson, R. L., Nogoy, N., & Goodman, L. (2017). Experiences in integrated data and research object publishing using GigaDB. *International Journal on Digital Libraries*, *18*(2), 99–111. https://doi.org/10.1007/s00799-016-0174-6

Greenleaf, G. (2021). *China's Completed Personal Information Protection Law: Rights Plus Cyber-security* (SSRN Scholarly Paper No. 3989775). https://doi.org/10.2139/ssrn.3989775

Guo, S., & Li, X. (2025). Cross-border data flow in China: Shifting from restriction to relaxation? *Computer Law & Security Review*, *56*, 106079. https://doi.org/10.1016/j.clsr.2024.106079

Ikram, N. A. H. S. (2024). DATA BREACHES EXIT STRATEGY: A COMPARATIVE ANALYSIS OF DATA PRIVACY LAWS. *Malaysian Journal of Syariah and Law*, *12*(1), Article 1. https://doi.org/10.33102/mjsl.vol12no1.458

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., … Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, *2*(1–2), 10–29. https://doi.org/10.1162/dint_r_00024

Jacobsen, A., Kaliyaperumal, R., Da Silva Santos, L. O. B., Mons, B., Schultes, E., Roos, M., & Thompson, M. (2020). A Generic Workflow for the Data FAIRification Process. *Data Intelligence*, *2*(1–2), 56–65. https://doi.org/10.1162/dint_a_00028

Kravchenko, O., Leshchenko, M., Marushchak, D., Vdovychenko, Y., & Boguslavska, S. (2019). The digitalization as a global trend and growth factor of the modern economy. *SHS Web of Conferences*, *65*, 07004. https://doi.org/10.1051/shsconf/20196507004

Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. L., Chue Hong, N., Goble, C., & Capella-Gutierrez, S. (2020). Towards FAIR principles for&nbsp;research&nbsp;software. *Data Science*, *3*(1), 37–59. https://doi.org/10.3233/DS-190026

Lee, J. (n.d.). *Cyberspace Governance in China: Evolution, Features and Future Trends*.

Li, Y. (2021). *CROSS-BORDER DATA TRANSFER REGULATION: A COMPARATIVE STUDY OF CHINA AND EUROPE*. https://u-pad.unimc.it/handle/11393/283978

Liu, L., & Chen, Y. (2024). A Triple-Layered Comparative Approach to Understanding New Privacy Policy Practices of Digital Platforms and Users in China After Implementation of the PIPL. *Social Media + Society*, *10*(4), 20563051241301265. https://doi.org/10.1177/20563051241301265

Miao, W., & Lei, W. (2016). Policy review: The Cyberspace Administration of China. *Global Media and Communication*, *12*(3), 337–340. https://doi.org/10.1177/1742766516680879

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), 49–56. https://doi.org/10.3233/ISU-170824

Mons, B., Schultes, E., Liu, F., & Jacobsen, A. (2020). The FAIR Principles: First Generation Implementation Choices and Challenges. *Data Intelligence*, *2*(1–2), 1–9. https://doi.org/10.1162/dint_e_00023

Neto, N. N., Madnick, S., Paula, A. M. G. D., & Borges, N. M. (2021). Developing a Global Data Breach Database and the Challenges Encountered. *Journal of Data and Information Quality*, *13*(1), 1–33. https://doi.org/10.1145/3439873

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Vincent Poor, H. (2021). Federated Learning for Internet of Things: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, *23*(3), 1622–1658. IEEE Communications Surveys & Tutorials. https://doi.org/10.1109/COMST.2021.3075439

Parasol, M. (2018). The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. *Computer Law & Security Review*, *34*(1), 67–98. https://doi.org/10.1016/j.clsr.2017.05.022

Peloquin, D., DiMaio, M., Bierer, B., & Barnes, M. (2020). Disruptive and avoidable: GDPR challenges to secondary research uses of data. *European Journal of Human Genetics*, *28*(6), 697–705. https://doi.org/10.1038/s41431-020-0596-x

Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1). https://doi.org/10.1093/cybsec/tyy001

Purnama Jati, P. H., van Reisen, M., Flikkenschild, E., Oladipo, F., Meerman, B., Plug, R., & Nodehi, S. (2022). Data Access, Control, and Privacy Protection in the VODAN-Africa Architecture. *Data Intelligence*, *4*(4), 938–954. https://doi.org/10.1162/dint_a_00180

Riccio, G. M. (2024). Data protection and appropriate measures: Too many uncertainties in the judicial applications? *UNIO – EU Law Journal*, *10*(1), Article 1. https://doi.org/10.21814/unio.10.1.5782

Rzayeva, J. (2024). *Standard contractual clauses as a gdpr safeguard: Implementation and challenges /* [Vilniaus universitetas.]. https://epublications.vu.lt/object/elaba:191367268/

Sacks, S., & Li, M. K. (2018). *How Chinese Cybersecurity Standards Impact Doing Business in China*. Center for Strategic and International Studies (CSIS). https://www.jstor.org/stable/resrep22317

Savona, M. (n.d.). *Data Governance: Main Challenges*.

Stocker, M., Stokmans, M., & Van Reisen, M. (2022). Agenda Setting on FAIR Guidelines in the European Union and the Role of Expert Committees. *Data Intelligence*, *4*(4), 724–746. https://doi.org/10.1162/dint_a_00168

Sullivan, C. (2019). EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era. *Computer Law & Security Review*, *35*(4), 380–397. https://doi.org/10.1016/j.clsr.2019.05.004

Tan, Z., & Zhang, C. (2021). China's PIPL and DSL: Is China following the EU's approach to data protection? *Journal of Data Protection & Privacy*, *5*(1), 7–25.

Tehrani, P. M., Sabaruddin, J. S. B. H., & Ramanathan, D. A. P. (2018). Cross border data transfer: Complexity of adequate protection and its exceptions. *Computer Law & Security Review*, *34*(3), 582–594. https://doi.org/10.1016/j.clsr.2017.12.001

The GDPR vs China's PIPL. (n.d.). Privacy Policies. Retrieved 15 August 2024, from https://www.privacypolicies.com/blog/gdpr-vs-pipl/

Torrisi, R. <1998>. (2023). *China's New Personal Information Protection Law (PIPL): Implications for Companies and Human Resources Management*. http://dspace.unive.it/handle/10579/24279

Van Reisen, M., Amare, S. Y., Nalugala, R., Taye, G. T., Gebreselassie, T. G., Medhanyie, A. A., Schultes, E., & Mpezamihigo, M. (2023). Federated FAIR principles: Ownership, localisation and regulatory compliance (OLR). *FAIR Connect*, *1*(1), 63–69. https://doi.org/10.3233/FC-230506

van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihgo, M., Folorunso, S., Schultes, E., Basajja, M., Aktau, A., Amare, S. Y., Taye, G. T., Purnama Jati, P. H., Chindoza, K., Wirtz, M., Ghardallou, M., van Stam, G., Ayele, W., Nalugala, R., Abdullahi, I., Osigwe, O., … Musen, M. A. (2021). Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Advanced Genetics (Hoboken, N.J.)*, *2*(2), e10050. https://doi.org/10.1002/ggn2.10050

Van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihgo, M., Folorunso, S., Schultes, E., Basajja, M., Aktau, A., Amare, S. Y., Taye, G. T., Purnama Jati, P. H., Chindoza, K., Wirtz, M., Ghardallou, M., Van Stam, G., Ayele, W., Nalugala, R., Abdullahi, I., Osigwe, O., … Musen, M. A. (2021). Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Advanced Genetics*, *2*(2), e10050. https://doi.org/10.1002/ggn2.10050

Virtosu, I., & Li, C. (2024). Navigating face recognition technology: A comparative study of regulatory and ethical challenges in China and the European Union. *International Conference on Machine Intelligence & Security for Smart Cities (TRUST) Proceedings*, *1*, 111–140.

Wijnbergen, D., Kaliyaperumal, R., Burger, K., Santos, L. O. B. da S., Mons, B., Roos, M., & Mina, E. (2024). *The FAIR Data Point Populator: Collaborative FAIRification and population of FAIR Data Points* (p. 2024.09.06.611505). bioRxiv. https://doi.org/10.1101/2024.09.06.611505

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Zaeem, R. N., & Barber, K. S. (2020). The Effect of the GDPR on Privacy Policies: Recent Progress and Future Promise. *ACM Trans. Manage. Inf. Syst.*, *12*(1), 2:1-2:20. https://doi.org/10.1145/3389685