Developing a federated data sharing network across multiple health care centers.

Contents

Introduction

Sharing of health data is an essential component in the provision of healthcare, in medical research, and disease surveillance. Health data sharing is subject to regulatory frameworks that vary across jurisdictions. There is a recent article published by (Nienaber McKay, Brand , & Botes, 2024) on the sharing of health data in Africa. Numerous factors complicate the regulation of health data sharing, including technological, motivational, economic, and political barriers, as well as ethical and legal challenges. The comparative study examines the regulation of health data sharing in Africa by comparing the legal and policy frameworks of five African countries. The study identifies gaps and inconsistencies in the current regulatory regimes and provides recommendations for improving the regulation of health data sharing in Africa. The five countries taken up into the publication are Ghana, Nigeria, Kenya, South Africa and Uganda.

https://pmc.ncbi.nlm.nih.gov/articles/PMC10800019/

Electronic (medical) records are a prerequisite for data sharing in a federated network. For more information on how to transition from paper medical records to electronic medical records we advise to consult the CTCAN Playbook: Paper to eMR Implementation.

1. Value of data sharing between multiple health care centers

In general, the value of data sharing between different healthcare centers is multifaceted:

- **Enhanced Patient Care**: Sharing patient data allows healthcare providers to access a more complete medical history, leading to better-informed decisions and improved patient outcomes.
- **Efficient Coordination**: Seamless data exchange ensures efficient coordination between centers, reducing redundant tests and ensuring continuity of care.
- **Public Health Insights**: Aggregated data helps to track health trends, identify outbreaks, and inform public health strategies.
- **Research and Evidence**: Shared data supports research efforts, enabling evidence-based practices and medical advancements.
- **Cost Savings**: Data sharing streamlines processes, reducing administrative costs.
- **Interoperability**: Achieving secure data exchange across systems benefits patients and providers.

Overall, data sharing fosters collaboration and contributes to better healthcare outcomes.

2. Designing a Federated Data Sharing Network for Healthcare Centers – critical steps

1. Define Objectives and Scope:

- Clearly outline the goals of the federated data sharing network.
- Identify the specific data types and use cases that will be addressed.
- Set boundaries on the scope to ensure manageable implementation.

2. Regulatory Compliance:

- Understand and comply with relevant healthcare data protection regulations (e.g. HIPAA, GDPR).
- Establish robust security and privacy protocols to safeguard patient information.
- Develop legal agreements and data sharing policies to govern the network.

3. Establish Data Governance Framework:

- Define data ownership, access controls, and responsibilities.
- Implement standardized data formats and coding systems for interoperability.
- Create a data quality management plan to ensure accuracy and consistency.

4. Infrastructure and Interoperability:

- Choose a scalable and flexible infrastructure to accommodate varying data volumes.
- Implement interoperability standards (e.g. FHIR) to facilitate seamless data exchange.
- Ensure compatibility with existing healthcare information systems.

5. Secure Communication Protocols:

- Use secure communication channels (e.g. TLS/SSL) to protect data during transit.
- Implement encryption mechanisms for data at rest to prevent unauthorized access.
- Regularly update and monitor security protocols to adapt to emerging threats.

6. Data Localization and Residency:

- Determine the geographical location of data storage to comply with regional regulations.
- Assess the need for data residency based on legal and ethical considerations.
- Establish protocols for data migration and backup to prevent data loss.

7. Participant Onboarding and Training:

- Develop a comprehensive onboarding process for healthcare centers joining the network.
- Provide training on data sharing protocols, security measures, and compliance requirements.
- Maintain ongoing communication channels to address queries and ensure adherence.

8. Monitoring and Auditing:

- Implement robust monitoring tools to track data access, usage, and anomalies.
- Conduct regular audits to ensure compliance with data sharing policies.
- Establish mechanisms for incident response and resolution in case of security breaches.

9. Establish Data Access Controls:

- Define roles and permissions for different stakeholders within healthcare centers.
- Implement a granular access control system to restrict data access based on user roles.
- Regularly review and update access controls to align with organizational changes.

10. Continuous Improvement and Evaluation:

- Establish metrics and key performance indicators (KPIs) to measure the success of the network.
- Solicit feedback from healthcare centers to identify areas for improvement.
- Iteratively update the network based on evolving technologies and healthcare needs.

By following these critical steps and criteria, you can develop a robust federated data sharing network that promotes collaboration while ensuring the security, privacy, and compliance of healthcare data across multiple centers.

3. Enhancing Federated Data Sharing Network with GDPR Compliance

- *Data Minimization and Purpose Limitation:*
  - Ensure that only necessary and relevant patient data is shared among healthcare centers.
  - Clearly define and communicate the specific purposes for which the data will be used

- *Lawful Basis for Processing:*
  - Identify and document the lawful basis for processing healthcare data under GDPR.
  - Obtain explicit consent from patients when required, ensuring transparency in data usage.

- *Data Subject Rights:*
  - Implement mechanisms to facilitate data subject rights, such as access, rectification, and erasure.
  - Establish procedures to respond to data subject requests within GDPR-mandated timelines.

- *Privacy by Design and Default:*
  - Integrate privacy measures into the design and development of the federated data sharing network.
  - Implement default privacy settings that prioritize data protection and minimize user intervention.

- *Data Protection Impact Assessment (DPIA):*
  - Conduct DPIA to assess and mitigate risks associated with data processing activities.
  - Document and address potential privacy and security concerns before deploying new features or functionalities.

- *Data Transfers:*
  - Ensure that cross-border data transfers comply with GDPR regulations.
  - Implement safeguards such as Standard Contractual Clauses (SCCs) or Binding Corporate Rules (BCRs) when transferring data outside the European Economic Area (EEA).

- *Security Measures:*
  - Enhance security measures to align with GDPR requirements, including encryption, pseudonymization, and access controls.
  - Regularly conduct security assessments and audits to identify and address vulnerabilities.

- *Data Breach Notification:*
    - Establish a robust incident response plan to promptly detect and respond to data breaches.
    - Adhere to GDPR's notification requirements, including notifying relevant authorities and affected individuals within 72 hours of discovering a breach.

- *Accountability and Record-Keeping:*
    - Implement mechanisms to demonstrate compliance with GDPR principles.
    - Maintain comprehensive records of data processing activities, risk assessments, and privacy measures.

- *Data Processor Agreements:*
    - Establish clear contracts and agreements with data processors, outlining their responsibilities and obligations.
    - Ensure that third-party entities processing healthcare data adhere to GDPR standards.

By incorporating these GDPR elements, the federated data sharing network not only promotes efficient collaboration but also prioritizes the protection of patient data in accordance with European data protection regulations. Compliance with GDPR principles enhances transparency, accountability, and trust in the healthcare data sharing ecosystem.

4. Reference to a Data Sharing Playbook

For IMI/IHI projects a Data Sharing Playbook is created, which provides strategies and resources to navigate common challenges associated with the provision of data in IMI/IHI projects, thus bringing efficiency into the process. (EFPIA, 2024)

The Playbook has been coordinated, managed and published by EFPIA and drafted with the help of several EFPIA partners.

In the Data Sharing Playbook, they have focused on challenges and consensual solutions. At the same time, they underline the need for early planning and multistakeholder involvement to bypass common obstacles. By proactively considering data sharing aspects from the development of a research idea to project implementation, most bottlenecks and challenges can be prevented or significantly mitigated. In this sense, they clearly advocate for more structured, well-thought processes that can be adopted by all actors, thus unlocking the value of the data more swiftly. Organizations are encouraged to proactively adapt internal processes, develop more standardized resources and create new roles to facilitate data provision.

It is to be expected that the introduction of the European Health Data Space in 2025 will underpin this approach.

Available here:

Open science | IHI Innovative Health Initiative (europa.eu)

5. Example of federated data networks on the African continent

LAISDAR – A federated data network to support COVID-19 research in Rwanda. LAISDAR is the first project in Africa implementing COVID-19 data harmonization (Nishimwe, Ruranga, & Musanabaganwa, 2022)

Background:

COVID-19 was first confirmed in Rwanda on March 14, 2020. By August 30, 2021, Rwanda conducted 2.4 million tests, confirming 87,131 positive cases and 1,083 deaths. The true number of cases is believed to be higher due to limited testing. Rwanda's COVID-19 data is scattered across multiple institutions, making comprehensive analysis challenging.

Project Objective:

The LAISDAR project aims to create a federated data network to unify and analyze COVID-19 data from multiple sources. The project is managed by the University of Rwanda, involving multiple Rwandan and Belgian institutions. Funded by Canada's International Development Research Centre (IDRC) under the Global South AI4COVID program.

Methods:

The network will use the OMOP Common Data Model (CDM) and OHDSI tools for data analytics and integration. It will include data from several hospitals, harmonized to OMOP CDM, and enriched with COVID-19 test and survey data.

Two open-source EHR systems, OpenMRS and OpenClinic GA, will be used with respective ETL processes. The Arachne platform will be used for integrating the sites with a central hub.

Deployment at hospitals will use Docker-based containerization, with pre-configured Mac Minis provided for consistent installations.

Training for end users will include documentation, online and live sessions, and local technical support to ensure continuity.

Results:

A proof-of-concept implementation was set up with a central LAISDAR instance and two data nodes (Mac Mini and AWS EC2). Initial tests of the integration layer and process flow for network studies were successful.
The project is advancing towards completing ETL implementations and integrating central COVID-19 test and survey results. The first phase includes 15 hospitals, with plans to expand.

Conclusion:

LAISDAR is the first project in Africa to implement COVID-19 data harmonization. It provides tools and data access for better disease tracking, outcome prediction, and response planning.
The infrastructure supports expansion to more hospitals and data sources and facilitates global collaboration through the OMOP CDM and OHDSI tools.

The project leverages advanced data science and AI techniques to support the Rwandan government in managing the COVID-19 pandemic, with a scalable and reproducible model that can be adopted by other regions.

The LAISDAR project provides challenges and solutions in data sharing:

Fragmentation and Standardization:

Challenge: COVID-19 data is fragmented across various institutions with different formats, making comprehensive analysis difficult.

Solution: LAISDAR addresses this by harmonizing data using the OMOP Common Data Model (CDM) to ensure consistency across different sources.

Privacy Concerns:

Challenge: Data owners are often reluctant to share data due to privacy concerns.

Solution: The project implements a federated data network, which allows data to be analyzed collectively without the need to centralize sensitive information. This approach mitigates privacy risks while enabling comprehensive data analysis.

Technical Integration:

Challenge: Integrating data from multiple systems with different structures and identifiers is complex.

Solution: The project uses two ETL (Extract, Transform, Load) processes for the different EHR systems (OpenMRS and OpenClinic GA). Unique identifiers are generated to match records from different sources reliably.

Infrastructure and Tools:

Challenge: Ensuring consistent and reproducible data integration across multiple sites.

Solution: The use of Docker-based containerization ensures uniform deployment of the LAISDAR software suite across all participating hospitals. The Arachne platform facilitates the integration of sites with a central hub for network studies.

Training and Sustainability:

Challenge: Ensuring that local personnel can maintain and utilize the data infrastructure beyond the project's lifetime.

Solution: The project provides comprehensive training and documentation for end users and technical staff. This includes user documentation, online training, live sessions, and the provision of ETL source code.

Key Takeaways:

Data Harmonization: Standardizing data formats across different institutions is crucial for effective data sharing and analysis.

Federated Approach: A federated data network can balance the need for comprehensive data analysis with privacy concerns by allowing decentralized data analysis.

Technical Solutions: Utilizing open-source tools and containerization ensures consistent implementation and facilitates the integration of diverse data sources.

Capacity Building: Providing extensive training and documentation ensures that local institutions can sustain and expand the data infrastructure in the future.

The LAISDAR project demonstrates that with the right infrastructure, tools, and training, it is possible to overcome the challenges of data sharing, particularly in the context of a public health crisis like COVID-19. This model can be replicated in other regions to enhance collaborative research and public health response efforts.

6. Tips for more in detail knowledge on Health Interoperability

Principles of Health Interoperability: FHIR, HL7 and SNOMED CT (Health Information Technology Standards). (Benson & Grieve, 2021)

7.Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
BCR	Blinding Corporate Rule
CDM	Common Data Model
DPIA	Data Protection Impact Assessment
EEA	European Economic Area
EFPIA	European Federation of Pharmaceutical Industries and Associations
eMR	electronic medical records
FHIR	Fast Healthcare Interoperability Resources
GDPR	Global Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
IHI	Innovative Health Initiative
IMI	Innovative Medicines Initiative
KPI	Key Performance Indicator
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
SCC	Standard Contractual Clauses
SSL	Secure Sockets Layer
TLS	Transport Layer Security

References

Benson, T., & Grieve, G. (2021). *Principles of Health Interoperability: FHIR, HL7 adn SNOMED CT.* Springer.

EFPIA. (2024). *Data Sharing Playbook.* Retrieved from Innovative Health initiative: https://www.ihi.europa.eu/sites/default/files/uploads/Documents/ProjectResources/IMI_IHI_DataSharingPlayBook_2024.pdf

Nienaber McKay, A. G., Brand , D., & Botes, M. (2024). The regulation of health data sharing in Africa: a comparative study. *journal of law and biosciences*. Retrieved from https://doi.org/10.1093/jlb/lsad035

Nishimwe, A., Ruranga, C., & Musanabaganwa, C. (2022). Leveraging artificial intelligence and data science techniques in harmonizing, sharing, accessing and analyzing SARS-COV-2/COVID-19 data in Rwanda (LAISDAR project): study design and rationale. *BMC medical informatics and decision making*. Retrieved from https://doi.org/10.1186/s12911-022-01965-9