Scientific Research Publishing

Journal of Information Security, 2026, 17(2), 25-42
https://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

Secure Data Lake Architecture Leveraging Jamii X-Change for Resilient Educational Assessment in Tanzania

Maduhu Mshangi Mlyatu, Margreth Venaely, Japhet Guyai

The National Examinations Council of Tanzania, Dar Es Salaam, Tanzania
Email: maduhumshangi@gmail.com, kileomargreth@gmail.com, guyai@necta.go.tz

How to cite this paper: Mlyatu, M.M., Venaely, M. and Guyai, J. (2026) Secure Data Lake Architecture Leveraging Jamii X-Change for Resilient Educational Assessment in Tanzania. *Journal of Information Security*, **17**, 25-42.
https://doi.org/10.4236/jis.2026.172003

Received: December 11, 2025
Accepted: February 3, 2026
Published: February 6, 2026

Abstract

This study proposes a system architecture design for a Student Data Lake integrated with Jamii X-Change, Tanzania's national interoperability platform, to improve educational assessment practices. With the growing volume and variety of student data across Tanzania's education system ranging from examination records and attendance to digital learning footprints traditional siloed databases are no longer sufficient for timely and accurate assessment. The key design challenge lies in enabling secure, scalable, and real-time integration of disparate educational datasets from various ministries, agencies, and institutions while ensuring privacy, data governance, and interoperability. To address this, the research adopts a secure by design approach and conceptualizes a data lake architecture that leverages X-Road protocols (underlying Jamii X-Change) for federated access, metadata driven governance, and layered analytics. The system comprises four layers: data ingestion, data lake storage, semantic integration, and analytics visualization. It also includes role-based access control and encryption to ensure data privacy and compliance with Tanzanian legal frameworks. A mixed-methods research methodology was employed; quantitative and Qualitative techniques were carried out to assess feasibility, usability, and security perceptions. Technical simulation of interoperability was performed using pilot data in Morogoro. However, concerns around data privacy, institutional silos, and limited infrastructure remain critical. This proposed system offers a scalable, secure, and interoperable foundation for real-time educational assessment, teacher feedback loops, and evidence-based policy making. Further pilot implementations and stakeholder capacity-building are recommended to validate its operational viability.

Keywords

Student Data Lake, Jamii X-Change, Educational Assessment,

Interoperability, Secure by Design, Leveraging Technology

1. Introduction

The digital transformation of public service delivery is reshaping sectors worldwide, with education emerging as one of the most profoundly impacted domains. Increasingly, digital technologies are being adopted not only to enhance operational efficiency but also to strengthen data security and promote inclusivity within education systems [1]-[3]. This transformation reflects a shift from the mere digitisation of processes to a fundamental rethinking of how educational data are collected, managed, and utilised to support evidence-based decision-making. Within this evolving context, the integration of digital platforms into educational assessment has become a strategic priority, as it supports the development of systems that are equitable, transparent, and resilient in meeting the demands of twenty-first-century education [4].

In Tanzania, however, the educational assessment system continues to rely heavily on manual, paper-based, and fragmented processes. Although initial efforts have been made to digitise specific functions, such as student registration, the broader assessment lifecycle from candidate verification to results processing and certification remains inefficient and vulnerable to data inaccuracies. These inefficiencies disproportionately affect rural and underserved regions, where access to timely and reliable educational services is limited [5]. Furthermore, the absence of interoperability between education systems and other key government sectors, including civil registration, health, and social welfare, constrains seamless data verification and information sharing. This fragmented environment increases administrative burdens, delays service delivery, and undermines the integrity of national educational statistics.

To address these systemic challenges, Tanzania has initiated the deployment of Jamii X-Change, a national digital interoperability platform designed to enable secure, real-time data exchange across public institutions. Jamii X-Change facilitates efficient communication between schools, examination councils, and civil registration authorities by standardising data formats and authentication protocols. Within the education sector, the platform plays a critical role in automating student identity verification, supporting accurate candidate registration, and enhancing the reliability of data used for planning and policy formulation [4] [6]. The operational structure and functionality of the platform, illustrated in **Figure 1**, demonstrate how Jamii X-Change functions as a central hub linking student information systems, national identification databases, and examination management systems into a unified and secure data flow. Through these integrations, Jamii X-Change not only streamlines assessment processes but also establishes the foundation for a more integrated, responsive, and data-driven educational infrastructure.

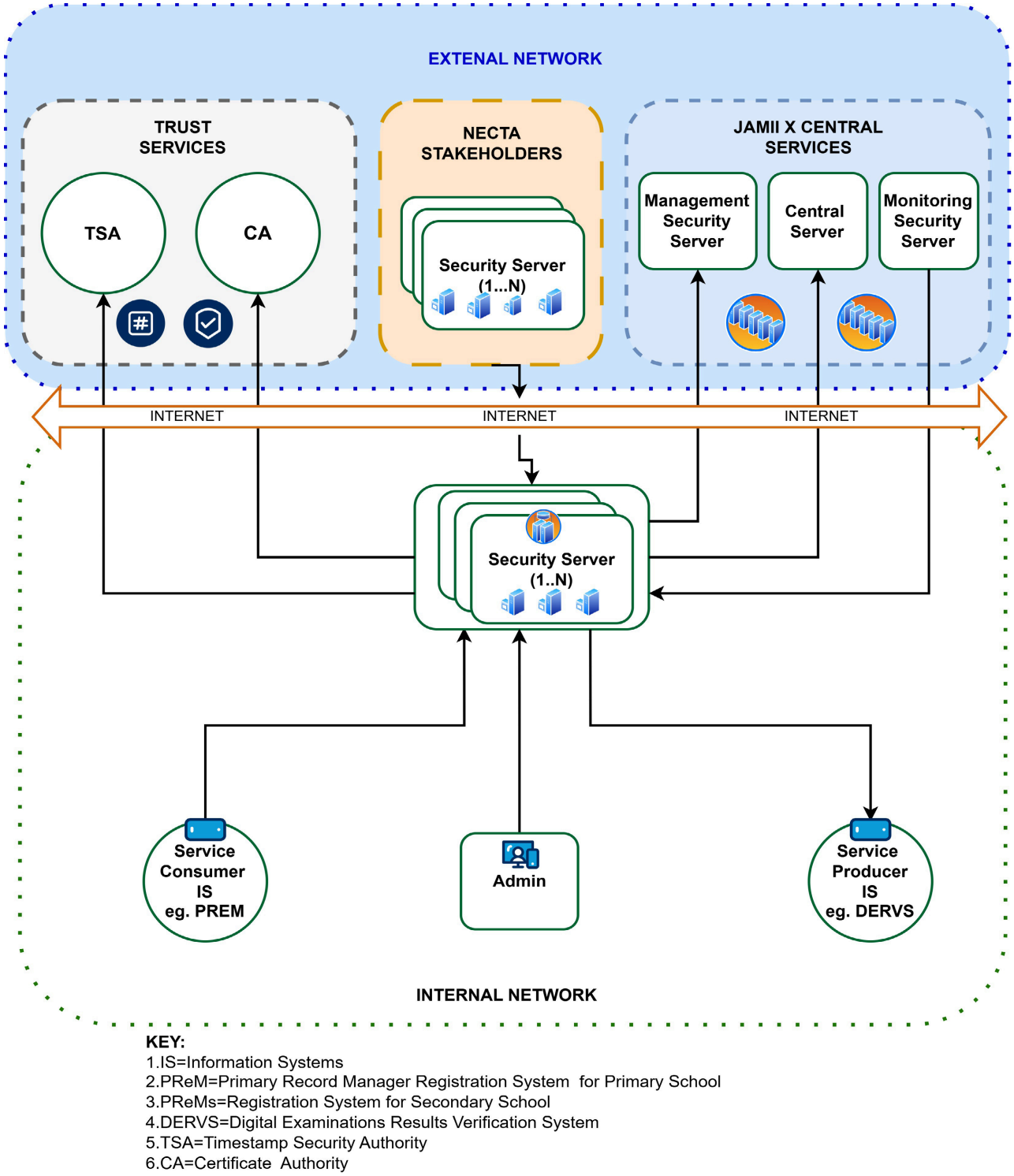

Figure 1. Jamii X-Change for education assessment.

Building on this foundation, this study proposes the design of a student data lake architecture model that leverages the Jamii X-Change platform. A student data lake functions as a centralised and scalable repository capable of storing large volumes of raw, structured, and unstructured data from diverse sources, including

school records, national examination results, attendance logs, and demographic information. Unlike traditional relational databases that enforce fixed schemas at the point of data ingestion, data lakes support a schema-on-read approach, thereby enabling flexible analytics and real-time reporting [7].

The proposed architecture has the potential to strengthen Tanzania's educational assessment ecosystem by enabling accurate and timely access to student records, reducing data duplication, streamlining examination processes, and supporting dynamic reporting and evidence-based policy interventions. Nevertheless, the implementation of such an architecture presents several challenges, including the need for reliable digital infrastructure, robust data governance frameworks, and adequate technical capacity among stakeholders to manage and interpret data effectively. Moreover, successful deployment requires alignment with national education policies, as well as sustained collaboration among government agencies, educational institutions, technology providers, and civil society organisations. These considerations underscore the importance of a holistic, multi-stakeholder approach to ensure the sustainable adoption and long-term impact of the proposed system.

Research Objective

The main objective of this study is to design and evaluate a student data lake architecture model leveraging the Jamii X-Change platform, with the aim of enhancing the efficiency, security, and inclusivity of educational assessment processes in Tanzania.

Research Questions

1) How can a student data lake architecture model, built on Jamii X-Change, improve the accuracy, efficiency, and security of educational assessment in Tanzania?

2) What are the key architectural components and integration strategies necessary for implementing a scalable and interoperable student data lake within Tanzania's education sector?

3) What infrastructural, technical, and governance challenges may arise in the deployment of a student data lake using Jamii X-Change, and how can these be mitigated?

4) How do stakeholders-including educators, administrators, policymakers, and technical staff-perceive the benefits and limitations of a student data lake architecture model for educational assessment?

This research study addresses the challenges posed by fragmented, limited-analytics, and non-interoperable educational assessment systems in Tanzania. The study systematically reviews relevant literature, identifies key research gaps, and outlines the research methodology, with findings and discussions organized around the study's research questions. It further proposes a scalable and secure Student Data Lake architecture designed to improve the accuracy, efficiency, and reliability of educational assessment processes. Moreover, the study examines the policy implications of implementing such a system in Tanzania and provides practical recommendations and concluding insights to support the development of a resil-

ient, interoperable, and evidence-based digital assessment ecosystem.

2. Literature Review

Student data lakes have increasingly become essential as educational institutions seek scalable and flexible architectures for integrating and analysing large volumes of data from multiple, often fragmented, sources. A data lake serves as a centralised repository that stores both raw and processed data in diverse formats, thereby enabling advanced analytics, real-time reporting, and machine learning applications [8]. In the education sector, such architectures facilitate the consolidation of student records, assessment results, attendance information, and demographic data, empowering institutions and policymakers with timely and accurate insights to improve learning outcomes and optimise resource allocation [9]. The inherent flexibility of data lakes makes them more suitable than traditional data warehouses for national education systems characterised by heterogeneous data types and evolving reporting requirements.

Several institutions have demonstrated the effectiveness of student data lake implementations. For instance, Ohio State University's Reporting and Analytics Environment integrates diverse institutional datasets into a governed analytics platform accessible to faculty members, administrators, and researchers [10]. Similarly, the New York City Department of Education employs a cloud-based data lake to integrate real-time school data, thereby enabling district-level monitoring and timely interventions [11]. These implementations illustrate how centralised and interoperable data environments enhance transparency and support evidence-based decision-making in education. A critical enabler of such data lake functionality is the underlying interoperability framework, which ensures secure, standardised, and real-time data exchange between heterogeneous systems.

Leveraging Jamii X-Change presents a strategic opportunity to consolidate fragmented education data sources, including student registration systems, examination boards, and teacher databases, into a unified repository. This integration supports longitudinal analysis, timely reporting, and improved credential verification across Tanzania's education ecosystem. Importantly, it addresses persistent challenges within the country's education system, such as manual data processing, limited system interoperability, and data accuracy issues that continue to constrain effective policy formulation and service delivery.

Emerging empirical research supports the development of student data lakes tailored to the Tanzanian context. Existing studies emphasise the need for scalable and context-sensitive data architectures capable of accommodating diverse educational environments characterised by variable digital infrastructure and pronounced rural-urban disparities [12]. Data lake architectures incorporating modular components, such as data ingestion pipelines, anonymisation modules, and analytics dashboards, enable phased implementation and functional flexibility, making them particularly suitable for resource-constrained environments [13]. Furthermore, the integration of Jamii X-Change Metrics enables real-time monitoring of data exchange health and service performance, which is essential for

maintaining stakeholder trust and ensuring long-term system reliability.

Despite these insights, a notable gap remains in the academic literature regarding the evaluation of student data lakes built on Jamii X-Change for improving educational assessment processes in Tanzania. Existing ICT research in the country has largely focused on infrastructure limitations, digital literacy, and isolated automation initiatives, without addressing comprehensive and interoperable data ecosystems that span the full assessment lifecycle. Moreover, the potential of Jamii X-Change to facilitate secure, real-time data interoperability has not been sufficiently leveraged or empirically examined within the context of educational data lakes. International best practices underscore the importance of robust data governance frameworks, security protocols, and sustained human capacity building in supporting successful digital transformation initiatives [3] [4]. Applying these principles to the design of a student data lake architecture underpinned by Jamii X-Change can help overcome existing challenges, reduce manual interventions, and provide policymakers and educators with reliable, actionable data. Accordingly, this study seeks to address this research gap by developing and evaluating a scalable and secure student data lake model that leverages Jamii X-Change's interoperability features and metrics capabilities to enhance educational assessment and data-driven decision-making in Tanzania.

3. Research Methodology

3.1. Research Design

This study adopted a mixed research methods design, integrating both quantitative and qualitative approaches to provide a comprehensive evaluation of the Jamii X-Change system within Tanzania's educational assessment context. The quantitative component involved the analysis of system performance indicators and registration metrics, enabling an objective assessment of the system's technical feasibility, operational efficiency, and reliability in supporting digital integration. The qualitative component comprised a documentary review of birth certificates issued by the Registration, Insolvency and Trusteeship Agency (RITA). These included typewritten birth certificates, older manually issued certificates, and newly issued digital birth certificates, which were examined to verify the consistency and accuracy of registration records against the RITA electronic registration system and NECTA students' registration systems, as well as other related documents presented during the pilot study.

In addition, qualitative data were collected through field observations conducted during the pilot study. This approach facilitated the collection of rich, contextual insights into user experiences, institutional preparedness, data quality challenges, and operational constraints associated with system implementation. The integration of quantitative and qualitative findings enabled data triangulation, thereby enhancing the validity, reliability, and interpretive depth of the study results. Mixed-methods research designs are widely recognized for their ability to combine numerical evidence with contextual understanding, resulting in more robust and comprehensive research outcomes [14].

3.2. Sampling Techniques

A purposive sampling technique was employed to select participants for the pilot study who were directly involved in, or affected by, the Jamii X-Change system integration. Two schools in the Morogoro Region, namely Mafiga Primary School and Morogoro Secondary School, were selected based on predefined criteria, including geographic accessibility, availability of ICT infrastructure, and administrative readiness to participate in the pilot study. These criteria ensured that the selected schools were representative of typical operational environments in which the system is intended to be deployed.

The sampling units comprised multiple participant groups to capture diverse perspectives across the educational assessment and civil registration ecosystem. These included all Standard Two students (n = 193) and Form Two students (n = 624) enrolled in the selected schools, as well as their parents or guardians, particularly those involved in birth certificate verification or digital certificate issuance processes.

In addition, the pilot study involved key institutional stakeholders selected based on their regulatory, managerial, and technical roles in the system's implementation. These included one registrar officer from the National Examinations Council of Tanzania (NECTA), one Regional Education Officer from the Morogoro Region, two District Education Officers from Morogoro Urban District, four registrar officers from the Registration, Insolvency and Trusteeship Agency (RITA), four RITA data entry officers, ten ICT officers from NECTA and RITA, two NECTA management officials, five RITA management officials, and eight management officials from the Ministry of Information and Communication Technologies. This multi-level purposive sampling approach ensured comprehensive stakeholder representation, which was essential for evaluating system performance and user experiences across policy, operational, and technical dimensions [15].

3.3. Data Collection Tools and Techniques

3.3.1. Quantitative Tools

Quantitative data collection involved extracting system log data and registration records. System logs from X-Road metrics and Jamii X-Change operational databases were analyzed to measure data exchange volumes, success rates, request-response times, and error incidences (e.g., error code ID 10702506125). These metrics provided objective indicators of system performance, reliability, and scalability during the pilot phase. Additionally, registration records from NECTA's PReM/PReMS systems were reviewed to quantify registration outcomes, including the proportion of students successfully registered, those issued new certificates, and cases excluded due to data inconsistencies or errors.

3.3.2. Qualitative Tools

Qualitative data were collected primarily through a documentary review conducted during the pilot study. This review focused on examining Birth certificates issued by RITA including typewritten birth certificates, older manually issued cer-

tificates, and newly issued digital birth certificates, with the aim of assessing data consistency, completeness, and accuracy when compared against records stored in the RITA electronic registration system and NECTA Students registrations systems. The documentary analysis enabled the identification of discrepancies, legacy data challenges, and procedural variations that could affect system interoperability and data integrity.

In addition, direct field observations were undertaken during onsite student registration, identity verification, and birth certificate issuance sessions. These observations systematically documented operational workflows, processing sequences, decision points, and stakeholder interactions among students, parents or guardians, school administrators, and registration officials. Particular attention was paid to identifying process bottlenecks, human-system interaction challenges, and deviations from standard operating procedures, which provided valuable insights into the practical realities of system implementation.

The combination of documentary review and field observation facilitated a context-sensitive understanding of institutional practices and user behaviour within real-world operational settings. This qualitative approach enabled the examination of institutional readiness, procedural efficiency, and implementation constraints that could not be fully captured through quantitative measures alone. Such methods are well established for generating in-depth contextual insights in applied information systems research and enhance the credibility and interpretive depth of the study findings [16].

3.4. Data Analysis Methods

3.4.1. Quantitative Analysis

Descriptive statistical techniques were employed to analyze student registration data, certificate issuance categories, and parental engagement metrics. System log data were also examined to evaluate technical parameters such as request sizes, latency periods, and frequency of request failures. These methods facilitated the assessment of the Jamii X-Change system's operational performance and technical reliability during the pilot implementation.

3.4.2. Qualitative Analysis

Qualitative data obtained through the documentary review were analysed using a thematic classification approach. Birth certificates issued by RITA including typewritten certificates, manually issued certificates, and newly issued digital certificates were systematically catalogued and compared with corresponding records in the RITA electronic registration system (e-RITHA) and the NECTA student registration systems (PReM and PReMS). This analysis focused on assessing data consistency, completeness, and accuracy, while identified discrepancies and legacy data challenges were categorised to evaluate their implications for system interoperability and data integrity.

Data collected from direct field observations were analysed through the systematic coding of observation notes, with particular emphasis on workflow sequences,

processing steps, stakeholder interactions, and decision points observed during onsite student registration and certificate issuance activities. Recurrent patterns related to process bottlenecks, human system interaction challenges, and deviations from standard operating procedures were identified and synthesised [17]. Findings from the field observations were subsequently triangulated with documentary evidence to enhance the validity, reliability, and interpretive depth of the qualitative analysis.

3.5. Ethical Considerations

Ethical protocols were rigorously observed throughout the study. Informed consent was obtained from all participants, including parents and school staff, ensuring voluntary participation and awareness of the study's objectives. Institutional approvals were secured from NECTA and RITA to access operational system data, reflecting adherence to governance and data stewardship policies. All collected data were anonymized during analysis and reporting phases to protect participant privacy and comply with Tanzania's data protection regulations. These ethical safeguards align with international standards for research involving human subjects and sensitive data [16].

3.6. Design of Student Data Lake Architecture Model

The study involved the development of a student data lake architecture model for leveraging the Jamii X-Change platform integrated with X-Road Metrics. This architecture aims to address the challenges of collecting, integrating, and analyzing student data across multiple educational institutions to support data-driven decision-making. The model for student data lake includes secure data exchange, real-time monitoring, usage analytics, and governance controls to ensure data quality and privacy. This design was validated through expert review and scenario-based testing (**Figure 2**).

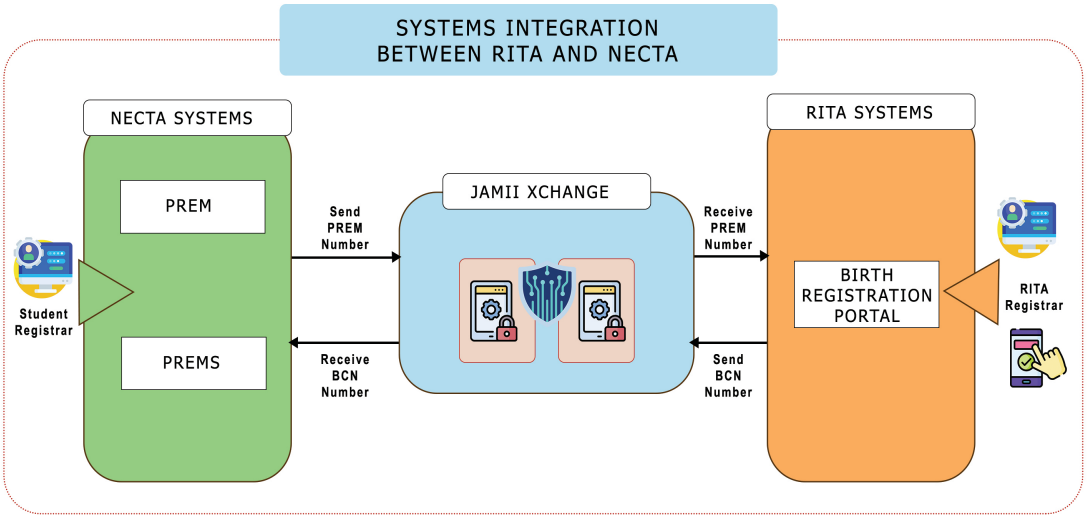

Figure 2. Integration overview between two entities.

4. Results Findings and Discussion

4.1. Implementation of Jamii X-Change Technology Affect the Accuracy and Efficiency of Educational Assessment

*RQ*1: *How does the implementation of Jamii X-Change technology affect the accuracy and efficiency of educational assessment processes in Tanzania*?

The results findings are as follows:

Case 1: *Pilot Integration of NECTA and RITA in Morogoro Schools*: The pilot implementation of Jamii X-Change for integrating NECTA and RITA systems in two Morogoro schools; one primary and one secondary demonstrated significant improvements in the accuracy and efficiency of student registration. At Morogoro Secondary School, among 154 Form Two students who submitted their birth certificates, 78 students (51%) were successfully verified and registered in NECTA's PReMS system through real-time data exchange with RITA. These findings highlight how interoperable systems can streamline registration processes, reduce manual data entry errors, and enhance data reliability.

Case 2: *log analysis of Jamii X-Change infrastructure showed efficient data exchange with typical request sizes between* 1 *KB and* 4 *KB*, *high success rates*, *and low latency. Occasional failed requests* (*e.g.*, *ID* 10702506125) *pointed to connectivity or sync issues*, *highlighting the importance of continuous system monitoring.*

Discussion:

These findings align with related studies emphasizing the critical role of interoperable platforms in enhancing educational data accuracy and operational efficiency [4] [13]. The positive pilot outcomes mirror global experiences where integrated digital systems reduce manual errors and speed up processes [3]. However, persistent infrastructure challenges, especially in connectivity, underscore the need for ongoing investment and monitoring to sustain reliable service delivery in resource-constrained settings.

4.2. Challenges Hindering the Effective Adoption of Jamii X-Change

*RQ*2: *What are the key infrastructural and technical challenges hindering the effective adoption of Jamii X-Change in Tanzania's education sector*?

The results findings are as follows:

Case 1-*Institutional Readiness*: The pilot exercise in Morogoro Region revealed encouraging signs of institutional capacity to implement Jamii X-Change. At Mafiga Primary School, 95 out of 193 students were registered using digital birth certificates, while at Morogoro Secondary School, 78 out of 154 students with digital certificates were successfully enrolled. Onsite issuance of new certificates further supported integration. However, challenges included outdated typewritten certificates (319 students), limited parental response (125 parents), and the need for enhanced digital literacy among school staff.

Case 2-*Logs analysis for System Performance*: Jamii X-Change infrastructure

demonstrated strong technical reliability, with high request success rates and sub-500 ms response times. The platform efficiently enabled data exchange between NECTA's PReM/PReMS and RITA's eRITA systems. Few failed requests were linked to connectivity or data sync issues, indicating the importance of sustained monitoring and support.

Discussion:

These findings support phased digital transformation through piloting and stakeholder engagement, aligning with global insights [3] [4]. Like Estonia's X-Road model [13] [18] [19], Tanzania's pilot shows that with targeted investments and capacity-building, foundational readiness can be converted into scalable digital infrastructure. The integration of Jamii X-Change demonstrated not only technical feasibility but also institutional willingness to adopt interoperable systems in education.

4.3. Perception of the Benefits and Limitations of Jamii X-Change Technology in Educational Assessment

*RQ*3: *How do stakeholders including educators, administrators, and policy-makers-perceive the benefits and limitations of Jamii X-Change technology in educational assessment*?

Results findings:

The pilot involving integration between NECTA and RITA through Jamii X-Change case study in Morogoro region revealed positive stakeholder perceptions. Educators and administrators noted enhanced data accuracy and reduced administrative workload, which increased their confidence in the student information systems. Policymakers recognized the system's potential to support data-driven planning and more efficient resource allocation. Despite these benefits, participants highlighted the need for continuous professional development, proactive technical support, and sustainable investment in infrastructure to ensure the system's effectiveness and scalability.

Discussion:

These findings align with similar studies emphasizing that interoperability platforms improve data quality and administrative efficiency in education [3] [4]. Consistent with global experiences, stakeholder buy-in depends heavily on capacity building and infrastructure readiness [13] [18] [19]. The pilot confirms that while technological solutions like Jamii X-Change enable real-time, accurate data exchange, the success of such digital transformations requires ongoing human and technical support, echoing lessons from Estonia's e-Governance model [18] and other digital education initiatives.

4.4. Sustainable Integration of Jamii X-Change Technology in Tanzania's Educational Assessment System

*RQ*4: *What strategies and policy interventions are necessary to enhance stakeholder capacity and ensure sustainable integration of Jamii X-Change technology in Tanzania's educational assessment system*?

Results Findings: The study identified critical strategies to support the sustainable integration of Jamii X-Change in Tanzania's educational assessment system. Key recommendations include phased rollout starting with pilot regions, ongoing capacity building focused on digital skills and cybersecurity, infrastructure expansion for improved connectivity and device access, and the development of governance frameworks for data privacy and usage. Metrics analysis of Jamii X-Change confirmed efficient data exchange with typical request sizes between 1 KB and 4 KB, high success rates, and low latency. Nonetheless, some failed requests highlighted areas needing better client-server synchronization and infrastructure reliability, emphasizing the need for continuous monitoring.

Discussion:

These findings align with recent research emphasizing the importance of incremental implementation and capacity development in digital education initiatives [3] [4]. Infrastructure challenges and governance remain common barriers in low- and middle-income countries' digital transformations [5]. The use of real-time operational metrics for system monitoring echoes best practices in interoperable digital ecosystems, as seen in Estonia's e-Governance and similar African digital platforms [6] [19]. This study confirms that combining strategic policy interventions with technical monitoring tools is vital to ensure effective and sustainable digital education reforms [20].

5. Proposed System Architecture Design for Student Data Lake Integrated with Jamii X-Change for Educational Assessment

Based on the results findings and stakeholder feedback, a student data lake architecture is proposed (**Figure 5**) to further address the challenges of data fragmentation and limited analytics capacity. The system design leverages Jamii X-Change for secure, interoperable data exchange and incorporates X-Road Metrics for operational monitoring. The main modules are:

1) Data Ingestion Module: Responsible to retrieve student data from Security Servers of NECTA, RITA, TCU, VETA, Loans Board, and other sources via Jamii X-change. The student data retrieved shall be inserted to the Mongo database. The execution of the collector module is performed automatically via a cron job task. The collector module needs HTTP-access to the Jamii X-change Central Server to get from global configuration list of members Security Servers. The collector module needs HTTP-access to an X-Road Security Server that acts as an Operational Monitoring Client to get the data collected. The collector module needs access to the Mongo Database.

2) Data corrector module: The Corrector module is responsible to clean the student data and derive monitoring metrics in a clean database collection. The execution of the corrector module is performed automatically via a service task. It is important to note that it can take up to 7 days for the Collector module to receive operational data from (all available) Security Server(s) and up to 3 days for

the Corrector module to clean the students raw data and derive monitoring metrics in a clean database collection.

3) Anonymizer: Is responsible of preparing the operational monitoring data for publication through the Opendata module. Anonymizer configuration allows Jamii x-change Metrics extension administrator to set fine-grained rules for excluding whole operational monitoring data records or to modify selected data fields before the data is published. The anonymizer module uses the operational monitoring data that Corrector module has prepared and stored to MongoDb as input. The anonymizer processes the student data using the configured rule-set and stores the output to the open data PostgreSQL database for publication (Figure 3).

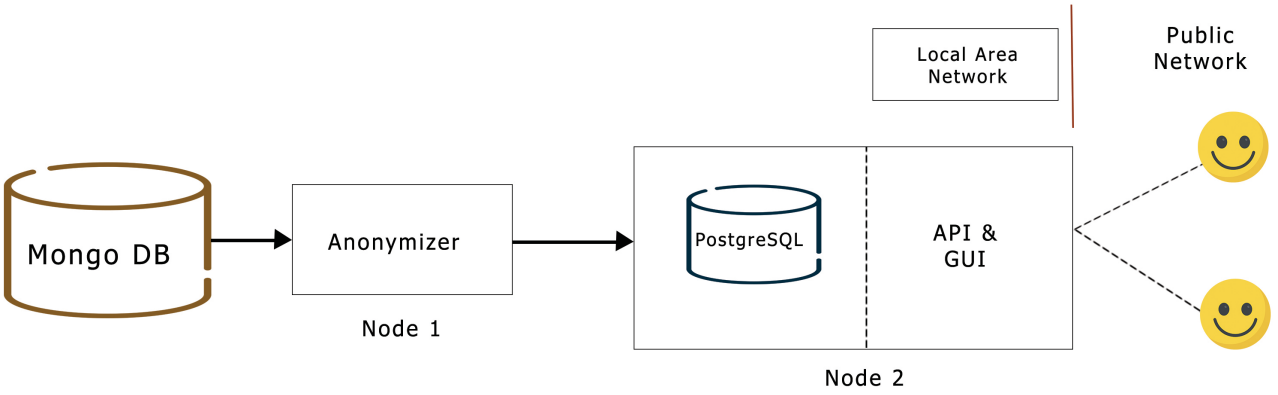

Figure 3. Anonymizer module.

MongoDb is used to store "non-anonymized" operational monitoring student data that should be accessible only by the Jamii X-Change Metrics administrators. Anonymized operational monitoring students' data that can be published for different stakeholders is stored in the PostgreSQL. The Opendata UI needs access only to the PostgreSQL (Figure 3). To follow the "principal of least privilege" it is recommended to install Opendata UI on a dedicated host that has no access at all to MongoDb. However, the Anonymizer module needs access also to the "not-public" data, so it should run on a host that has access to both MongoDb and PostgreSQL.

4) Networking/Visualizer Module: The purpose of the Networking module is to visualize the networking activity between the Jamii X-Change members (Figure 4). It consists of:

Data preparation R script that queries the Open data PostgreSQL database, does relevant calculations (query counts between Jamii X-Change members) and writes a table file to be used in the visualization web application.

Also networking module consists of RStudio Shiny-based web application to visualize the networking activity between X-change members in education sector.

Periodic crony job run to execute data preparation script. This can be started with the following steps: Read configuration parameters from a settings file, Read instance specific complementary files; establish a connection to Open data

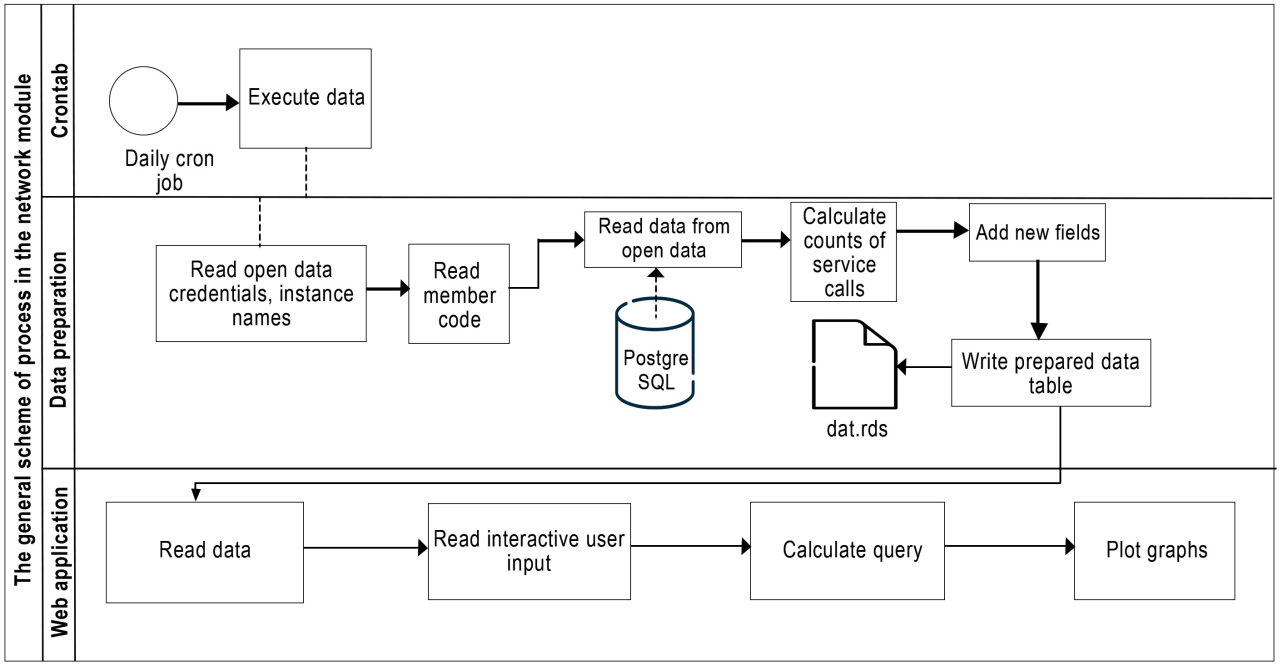

Figure 4. The general scheme of processes in the Networking module.

PostgreSQL and queries the most recent date in field; querie the most recent data from Open data dating back to certain number of days ("interval" in settings file, default 30 days) from the most recent data. The script retrieves only succeeded = True logs. The script retrieves the following fields: Calculates the count of service calls between each unique combination of the fields: Add the following concatenate fields client, producer, and producer. Service using newline escape sequence as a separator: Parses member names from riha.json and adds the names to the output data, writes the resulting table to RDS-file (R-specific binary file) in the shiny application's dynamic data directory.

5) Reports module: Is responsible for creating monthly reports about subsystems of X-Change members in education sector (datasets usage). The execution of the report's module can be either performed automatically (via cron job) or manually. Overall **system**, its users and rights, processes and directories are designed in a way, that all modules can reside in one server (different users but in same group "Jamii X-change-metrics") but also in separate servers. Also, the System designed in a way, that allows to monitor data from different X-change instances. However, the system designed in such a way that it can be used by all X-change members in education sectors as well as for Member own monitoring (includes possibilities to monitor also members data exchange partners). The proposed System architecture design for a Student Data Lake integrated with Jamii X-Change is presented in Figure 5.

This architecture (Figure 5) is modular, scalable, and designed to support both current and future data analytics needs in Tanzania's education sector. The proposed design for a Student Data Lake integrated with Jamii X-Change provides a practical solution for managing and gradually phasing out outdated RITA

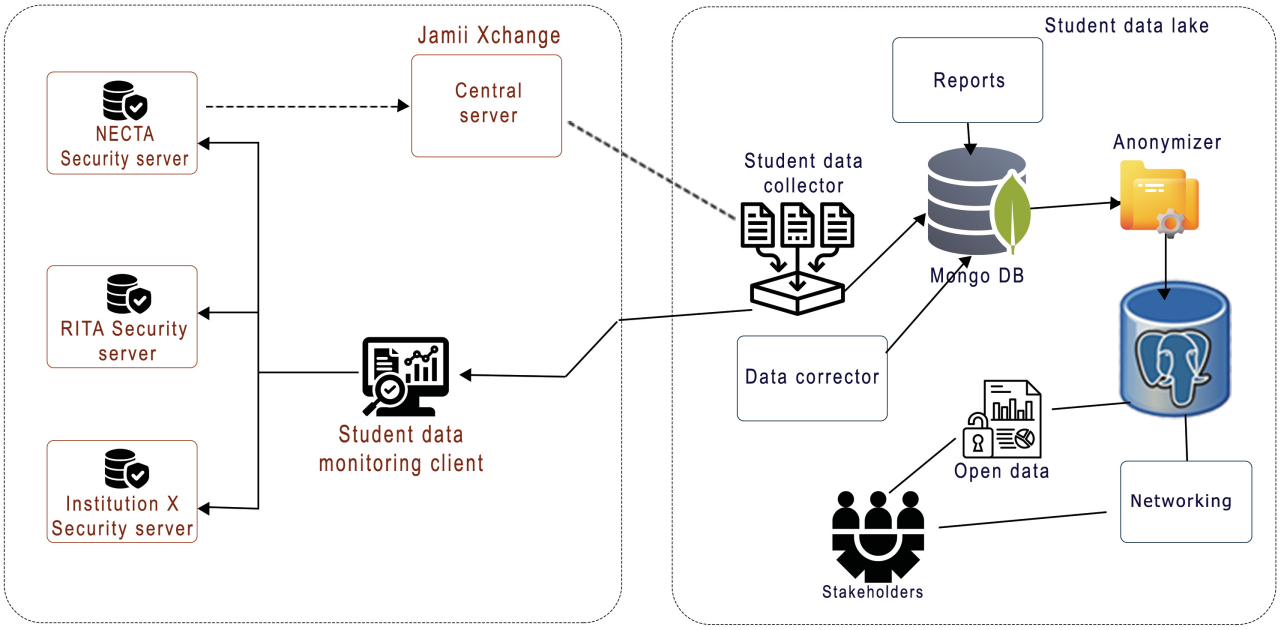

Figure 5. System architecture design for a Student Data Lake integrated with Jamii X-Change.

typewritten birth certificates. It converts them into authoritative, digitally verifiable identity records. Legacy birth certificate data can be digitised and ingested into the data lake through controlled capture processes. These processes standardise key attributes such as name, date of birth, registration number, place of birth, and issuing authority, and link them to student identifiers. Once ingested, the system can securely validate and reconcile these records against e-RITA registration and national identification systems. This reduces identity mismatches, duplicate records, and delays associated with manual verification during examination registration and results processing.

The architecture supports a phased deprecation approach, whereby legacy certificates remain accessible for audit purposes while new registrations rely on digitally authenticated identity data. This enables RITA to progressively replace typewritten certificates with trusted, interoperable digital records, thereby enhancing accuracy, efficiency, and integrity across Tanzania's educational assessment system. The approach minimises disruption to existing workflows, allowing institutions to maintain continuity during the transition. It also establishes a foundation for future integration with national digital identity systems, ensuring long-term scalability and compliance.

Furthermore, as data from additional institutions such as TCU, NACTVET, VETA, TIE, HESLB, MoEST, PMO-RALG, and higher learning institutions in Tanzania are integrated, a robust governance model ensures that data ownership remains with the originating institutions. At the same time, access rights and usage permissions are managed centrally through role-based access controls and comprehensive audit logging, which reinforces accountability, security, and compliance with national data policies, while supporting both day-to-day operational requirements and the long-term sustainability of the system.

6. Policy Implications of Integrating a Student Data Lake with Jamii X-Change for Educational Assessment in Tanzania

From a policy perspective, the adoption of a Jamii X-Change-enabled student data lake has significant implications for educational assessment in Tanzania. First, it provides an evidence-based foundation for targeted resource allocation, enabling the Ministry of Education to prioritise investments in ICT infrastructure for underserved regions, particularly rural and hard-to-reach schools where data gaps are most pronounced. Real-time, integrated data on enrolment, assessment outcomes, and teacher deployment can support more equitable distribution of learning materials, examination resources, and qualified educators, consistent with national goals for inclusive and equitable quality education [3] [21].

Second, the successful operationalisation of such a data lake necessitates deliberate policy commitment to digital literacy and capacity building, including continuous training for education officers, school administrators, and ICT personnel in data management, analytics, and information security. Empirical evidence suggests that investments in digital skills are critical for translating ICT infrastructure into improved institutional performance and data-driven decision-making [4] [22]. Without parallel investments in human capital, the benefits of advanced data architectures may remain underutilised. Finally, embedding the student data lake within national education policy frameworks would strengthen data governance, interoperability standards, and accountability mechanisms, ensuring sustainability and alignment with Tanzania's broader digital government and education transformation agendas, including e-Government and cross-sector data sharing initiatives [3].

7. Recommendations

1) *Phased Nationwide Rollout*: Expand Jamii X-Change and the student data lake through staged pilots, starting in regions with varying infrastructure and scaling up based on lessons learned.

2) *Capacity Building*: Implement ongoing digital literacy and technical training for educators, administrators, and IT staff.

3) *Infrastructure Investment*: Prioritize internet connectivity and device access, especially in rural and underserved areas.

4) *Robust Data Governance*: Develop and enforce clear policies on data privacy, security, and access control.

5) *Continuous Monitoring*: Use X-Road Metrics and regular audits to ensure system health, guide improvements, and maintain stakeholder trust.

6) *Stakeholder Engagement*: Involve all relevant parties in design, implementation, and evaluation to ensure buy-in and address context-specific needs.

8. Conclusion

This study demonstrates that Jamii X-Change technology, when integrated with a

well-designed student data lake, can significantly enhance the efficiency, accuracy, and accessibility of educational assessment data in Tanzania. The pilot results and stakeholder feedback confirm that interoperability, automation, and centralized analytics are key to overcoming persistent challenges. The proposed architecture offers a scalable, secure solution that aligns with international best practices and supports evidence-based decision-making. For successful nationwide adoption, sustained investment in infrastructure, comprehensive training, and phased implementation are essential. Future work should focus on expanding the data lake to include additional educational and social data sources, further enhancing analytics for policy and planning.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Organisation for Economic Co-Operation and Development (2024) Education at a Glance 2024: OECD Indicators. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/09/education-at-a-glance-2024_5ea68448/c00cad36-en.pdf

[2] Smith, J. and Jones, B. (2022) Data Security in Educational Systems. *International Journal of Cybersecurity*, **18,** 321-345.

[3] UNESCO (2023) Global Education Monitoring Report 2023: Technology in Education. UNESCO Publishing. https://www.unesco.org/gem-report/en/publication/technology

[4] Brown, A. (2023) Interoperability Frameworks for Education. *Journal of Educational Technology*, **45**, 123-145.

[5] World Bank (2021) Digital Transformation of Education Systems: Policy Recommendations for Developing Countries. World Bank Group. https://www.worldbank.org/en/topic/digital/overview

[6] Ministry of Communications and Information Technology (2025) Jamii X-Change Platform Overview. Government of Tanzania. https://ictc.go.tz/storage/44/01J5TJNDHDVTA3AHVQN9CR6BV7.pdf

[7] DataCamp (2025) Data Lakes and Their Applications. https://www.datacamp.com/blog/what-is-a-data-lake

[8] Qubole (2024) Data Lake Architecture and Best Practices. https://www.qubole.com/blog/2024-data-lake-trends-and-best-practices

[9] RisingWave (2024) Data Lake Definition and Characteristics. https://risingwave.com/glossary/data-lake

[10] Ohio State University (2025) Reporting and Analytics Environment (RAE). https://eari.osu.edu/

[11] New York City Department of Education (2022) Data Lake Initiatives. https://www.schools.nyc.gov/about-us/reports

[12] Warioba, C., Kiwonde, M. and Temu, S. (2022) Designing Scalable Educational Data Architectures in Tanzania. *Tanzanian Journal of ICT in Education*, **10**, 45-60. https://journals.udsm.ac.tz/index.php/jicts/issue/view/774

[13] Nordic Institute for Interoperability Solutions (2021) X-Road Metrics [GitHub Re-

pository]. https://github.com/nordic-institute/X-Road-Metrics

[14] Creswell, J.W. and Creswell, J.D. (2018) Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage.

[15] Patton, M. (2015) Qualitative Research and Evaluation Methods. 4th Edition, Sage Publications.

[16] Amazon Web Services (2024) Data Lakes on AWS: Architecture and Best Practices. https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/

[17] Braun, V. and Clarke, V. (2006) Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, **3**, 77-101. https://doi.org/10.1191/1478088706qp063oa

[18] e-Estonia (2023) Digital Transformation in Education: The Estonian Experience. https://e-estonia.com/solutions/education

[19] Kitsing, M. (2020) The Role of Digital Platforms in Public Administration: Lessons from Estonia. Springer. https://thedocs.worldbank.org/en/doc/165711456838073531-0050022016/original/WDR16BPEstonianeGovecosystemVassil.pdf

[20] Ministry of Communications and Information Technology (2025) Digital Tanzania Project Overview. Government of Tanzania. https://documents1.worldbank.org/curated/en/485771622426544909/pdf/Tanzania-Digital-Tanzania-Project.pdf

[21] Ministry of Education, Science and Technology (2025) Education Sector Performance Report 2020/2021. Government of Tanzania. https://www.researchgate.net/publication/392125922_Education_Performance_for_Results_Program_in_Tanzania

[22] World Bank (2022) Investing in Digital Skills for Education Systems. World Bank Group. https://www.worldbank.org/en/topic/edutech