Government of Pakistan
Project Coordination Unit
Ministry of Federal Education and Professional Training

PRE-BID MEETING
Hiring of TA Firm under ASPIRE TA DLI 11.3 for Horizontal Integration of Provincial Education Databases

The subject meeting was held on 07-05-2026 to address the queries and/or clarification sought by potential bidding firms. Following participants attended the meeting:

Participants:

1. Ms. Uzma Malik, Procurement Specialist, PCU
2. Mr. Anjum Afzal, MIS Specialist, PCU
3. Mr. Zubair Farooq, Deputy Director (MIS), PIE

Firms Representatives:

1. Mr. Ishfaq Kiyani, M/s Saphire Consultant
2. Mr. Inshaal Rajput, M/s Saphire Consultant
3. Mr. Umair Khan, M/s PTCL
4. Raja Anwaar Ali, M/s PTCL
5. Mr. Sami Ullah, M/s Antox (Pvt.) Ltd.
6. Mr. Bilal Waqas Khan, M/s Section Soft
7. Mr. Akhtar Ali, M/s MK Technology
8. Mr. M. Arfan, M/s Micro Merger
9. Mr. Muhammad Niaz, M/s Eiffelcem
10. Mr. Waqar Ali, M/s Betacodes (Online)

UZMA MALIK
Procurement Specialist
Project Coordination Unit
Ministry of Federal Education & Professional Training

Meeting Proceedings:

The meeting was initiated by the Procurement Specialist with brief of the activity, and technical details were provided by the MIS Specialist and Deputy Director (MIS). During the meeting participants sought various queries/clarifications regarding project scope, implementation modalities, technical requirements, and operational arrangements. The following clarifications were provided by the concerned officials:

Sr. No.	Question	Answer
1	Whether hardware is required as part of the project implementation.	Hardware provision is not part of the project scope, however, during implementation, if any hardware gaps are identified, the vendor may share the required specifications with the provincial ASPIRE teams for arrangement and provisioning. Configuration would be done by the vendor.
2	Whether departments already have systems in place and whether standardized APIs are available.	Required integration expectations are already defined in the relevant annexure. M/s Tabadlab mapping is again annexed for further clarification. Vendors will be responsible for bringing data into the landing zone. In cases where systems do not exist interfaces/platform for

		data entry backed by database would be developed accordingly.
3	Whether there is any preferred or recommended technology stack.	Participants were advised to review the proposal documents carefully. The proposed technologies should remain compatible with the existing project ecosystem and operational environment. The legacy system is developed on React, Node.Js, Express.js and SQL Server with Superset.
4	What level of development is required for the project.	It was clarified that frontend, backend, and tested systems are already available. Any realignment requirement arise during implementation will be part of it. Vendors will primarily be responsible for 1. Development of Landing Zones for all provinces 2. Data integration through APIs or automated process. 3. Develop the data modules for departments where no system exists. 4. Perform ETL for Landing Zones 5. Integrate Landing Zones with NODP.
5	How data sanitization and duplication handling will be managed, and whether provinces will be responsible for duplicate data.	It was clarified that the objective is to strengthen existing systems rather than replace them. ETL will be performed at Landing Zone and checks will be applied accordingly.
6	Whether hosting is allowed or who will provide hosting.	It was clarified that hosting infrastructure for the landing zone will be provided by the provinces.
7	Clarification regarding logistics, TADA arrangements, and TWG meetings.	TWG participants as per applicable government TADA rules will be reimbursed by the vendor. An appropriate venue would be selected with meeting facilities. One Participant from each province and 3/4 participants from PIE attend the meetings. Meeting are done on rotation basis in all provinces twice a year.
8	What are the project deployment and verification timelines.	Participants were informed that system deployment and verification activities are required to be completed by March 27, followed by an operational engagement/support period up to June 28. It was emphasized that timelines are linked with World Bank disbursement milestones and Government of Pakistan commitments; therefore, delays will not be acceptable.
9	Whether dedicated resources and training sessions will be required during implementation.	The technical team recommended deployment of dedicated/deputed resources to ensure rapid coordination and implementation progress. The implementation team will also provide training sessions to departmental representatives, with one representative from each department expected to act as technical lead/focal person. PIE may require deployment of resources on-premises till validation.
11	Whether Consortium or Joint Ventures (JV) are allowed under the project.	Participants were informed that Joint Ventures (JV) or Consortium are not allowed due to rapid development and deployment nature of the project.

Participants were advised to submit any further technical and procurement-related queries through official communication channels within the stipulated timeline mentioned in the RFP documents. It was further informed that written minutes of the pre-bid meeting and clarifications would be shared/uploaded accordingly.

****

Uzma Malik
Procurement Specialist,
PCU, M/oFE&PT

UZMA MALIK
Procurement Specialist
Project Coordination Unit
Ministry of Federal Education &
Professional Training

Plan of Actions for Horizontal Integration

Version 2.0

October 2025

Table of Contents

1 Introduction and Background 6
1.1 Data Standardisation Framework 6
1.2 Data Integration 6
2 Scope of Plan of Actions (PoAs) 6
3 Approach and Methodology 7
4 Horizontal Integration 7
4.1 Architecture of Horizontal Integration 7
4.2 Steps for Horizontal Integration 8
5 Punjab 9
5.1 Private Education Provider Registration and Information System (PEPRIS) 9
5.2 Punjab Education Foundation (PEF) 9
5.3 Punjab Education Initiative Management Authority (PEIMA) 9
5.4 Punjab Education, Curriculum, Training and Assessment Authority (PECTAA) 10
5.5 Special Education Department 11
5.6 Literacy & Non-Formal Basic Education Department (L&NFBED) 12
5.7 Higher Education/ Colleges Department (HECD) 12
5.8 Provincial EMIS: PMIU 12
6 Sindh 14
6.1 Sindh Education Foundation 14
6.2 Directorate of Inspection and Registration of Private Institutes Sindh (DIRPIS) 14
6.3 Sindh Teachers Education Development Authority (STEDA) 14
6.4 Department of Empowerment of Persons with Disabilities (DEPD) 14
6.5 Directorate of Literacy and Non-formal Education (DLNFE) 16
6.6 Higher Education/ Colleges Department (HECD) 16
6.7 Provincial EMIS: RSU 16
7 Khyber Pakhtunkhwa 17
7.1 Private Schools Regulatory Authority (PSRA) 17
7.2 Merged Areas Education Foundation (MaEF) 17
7.3 Elementary and Secondary Education Foundation (ESEF) 18
7.4 Directorate of Professional Development (DPD) 19
7.5 Directorate of Social Welfare, Special Education & Women Empowerment (SWSE&WE) 19
7.6 Education Monitoring Authority (EMA) 19
7.7 Higher Education Regulatory Authority (HERA) 19
7.8 Higher Education/ Colleges Department (HECD) 20
7.9 Provincial EMIS: i-EMIS 20
8 Balochistan 21
9 Training of Technical Staff 22
10 Vertical Integration 23
10.1 Provincial EMIS 23
10.2 National and Federal Entities 23
10.3 Pakistan Institute of Education (PIE) 24
11 Annexures 24
11.1 Annex A: Sindh Education Foundation 24
11.2 Annex B: Directorate of Literacy and Non-formal Education 25
11.3 Annex C: KP Private Schools Regulatory Authority 25
11.4 Annex D: KP Elementary and Secondary Education Foundation 26
11.5 Annex E: Requirement Details 27

Introduction and Background

The 18th Constitutional Amendment of 2010 was a reform that devolved education service delivery to Pakistan's provinces, allowing them to design localised solutions tailored to their unique contexts. This devolution was aimed at strengthening ownership, enabling provinces to pursue policy priorities that best served their population, and establishing their provincial education data systems that supported their causes. Since 2010, each province has progressed at its own pace, reflecting different levels of capacity and distinct policy agendas.

However, while provincial autonomy drove innovation, one drawback was the fragmentation of data systems across the country. With each province operating independently, the lack of common standards meant that it became difficult to generate a coherent national picture of progress for Sustainable Development Goal 4 (SDG-4). The absence of standardisation reduced comparability between provinces and made it harder to consolidate information for national and international reporting.

Data Standardisation Framework

Recognising this challenge, the Pakistan Institute of Education (PIE), supported by the World Bank's Data and Research in Education (DARE) programme, introduced the Data Standardisation Framework (DSF) in 2023. Conceived as a live document, the DSF originally comprised over 50+ indicators across eight thematic areas, carefully designed to harmonise data collection and reporting across provinces. Critically, the DSF was endorsed by all provinces in the Inter-Provincial Education Ministers Conference (IPEMC), signalling broad political ownership and a collective commitment to improved data governance. In August 2025, an updated version was shared with provincial representatives to address emerging challenges and maintain the framework's relevance. The DSF now forms the foundation of Pakistan's education data reform, enabling provinces to align with a shared national framework while still addressing their local needs.

Data Integration

The National Open Data Portal (NODP) is the centralised, government-owned repository for structured, standardised, and publicly accessible data across sectors. Its core objective is to ensure transparency, interoperability, and evidence-based policymaking by enabling federal, provincial, and regional datasets to converge into a single access point. By adhering to internationally recognised open data principles and the domestic Data Standardisation Framework (DSF), the NODP ensures that data is both accessible and comparable across time, geography, and institutions.

Developed by the Pakistan Institute of Education (PIE) with support from the ASPIRE programme, the NODP acts as the national hub where provincial datasets are periodically updated and consolidated. Sensitive data is masked to protect privacy, while aggregated information is made openly available to policymakers, researchers, and the public. This strengthens accountability and enables Pakistan to meet international reporting obligations, particularly for SDG-4 indicators.

Horizontal integration links provincial departments and agencies to their respective provincial Education Management Information Systems (EMIS). Through this mechanism, diverse education data streams from multiple entities are consolidated within the provincial EMIS, enabling a comprehensive provincial view of education performance. At this level, the exercise includes all departments under the School Education Department, as well as other entities that generate education data at the provincial level.

Vertical integration builds on horizontal integration by ensuring that data consolidated within provincial EMIS is pushed to the NODP via secure APIs, creating a unified national education data ecosystem. In addition, national and federal institutions that do not participate in horizontal integration are integrated directly into the NODP through regulated portal access.

Vertical integration is being operationalised through two streams:

1. Provincial EMIS Integration – Horizontally consolidated datasets, along with provincial EMIS data, are transferred to the NODP through automated, API-enabled protocols following required approvals.
2. National and Federal Integration – Entities operating at the federal level are directly integrated into the NODP via controlled portal access, ensuring comprehensive coverage.

Scope of Plan of Actions (PoAs)

PoAs are created to present the as-is status of departments, outline their requirements for horizontal and vertical integration, and recommend the procedures to be followed. The focus is on assessing the current state of departmental data management, the extent of existing data integration mechanisms (if any) with provincial EMIS, and the support needed for full integration. The document covers all entities through which PIE collects education data for national and international reporting.

The as-is review will determine whether departments maintain digital databases, partially digitised records, or rely on manual systems, and whether they are already connected to EMIS, exchange data intermittently, or operate in isolation. Based on this, integration will follow two main pathways: departments with established systems will connect directly through Database Connect or Application Programming Interface (API)-based exchange, while others will be provided a portal-based solution to submit data in a standardised format.

For integration to be sustainable, some departments may require infrastructure and capacity, including upgraded hardware, compatible software, and staff training in data entry and validation. In certain cases, dedicated personnel may need to be appointed to manage the process.

The plan of actions, therefore, sets out a phased and tailored approach: advanced departments will progress through direct integration, while those less developed will be enabled through portals and capacity-building measures. Together, these steps will allow provinces to consolidate data more efficiently, improve education planning, and contribute to a unified and reliable education data system.

Approach and Methodology

The methodology for collecting data from each department against the respective Plans of Action (PoAs) involved conducting individual or group interviews with representatives who were well-versed in the department's data and technology systems. These representatives served as the primary source of information, responding to a structured questionnaire, available in the DSF 2.0 document, which guided the discussion and ensured consistency across departments. Where representatives were not available, despite repeated reminders and follow-up attempts, the required information was obtained through prior experience and reliable online sources. In such cases, broad standardised assumptions and suggestions were applied uniformly to ensure that all departments were adequately represented within the overall analysis.

Horizontal Integration

Horizontal integration is the foundational step in creating a unified education data ecosystem in Pakistan. It focuses on ensuring that all education-related departments and entities within a province are systematically connected to the central Education Management Information System (EMIS) of that province. While each department historically maintained its own records and systems, often tailored to its specific functions, horizontal integration brings these diverse sources together, creating a single, coherent repository of provincial education data.

In practice, this means that information held by major education departments — whether operating as attached formations of the School Education Department (SED) or functioning as independent entities — is integrated into the provincial EMIS. The following departments in the provinces are understood to be the central education repository for education data in the province:

- **Punjab:** Programme Monitoring and Implementation Unit (PMIU) in Lahore
- **Sindh:** Reforms Support Unit (RSU) in Karachi
- **Balochistan:** Policy, Planning and Implementation Unit (PPIU) in Quetta
- **Khyber Pakhtunkhwa:** KP-EMIS (Civil Secretariat) in Peshawar

This process delivers several important benefits. First, it breaks down institutional silos, allowing departments to move away from fragmented record-keeping towards integrated, standardised data management. Second, it improves the quality and reliability of information by ensuring consistency across all contributing entities. Third, it enables provinces to use their data more effectively for planning, monitoring, and evidence-based decision-making, as education officials gain access to a consolidated system rather than navigating multiple disconnected datasets.

By consolidating records generated by departments and entities on students, teachers, institutions, and other relevant variables into a unified system, provinces will be able to generate a comprehensive and accurate picture of their education landscape. Moreover, it will assist the PIE team in consolidating comprehensive data from provinces and developing a complete picture of education data in Pakistan.

Architecture of Horizontal Integration

This integration architecture provides the framework for connecting data sources, transforming and then storing them into a consistent structure.

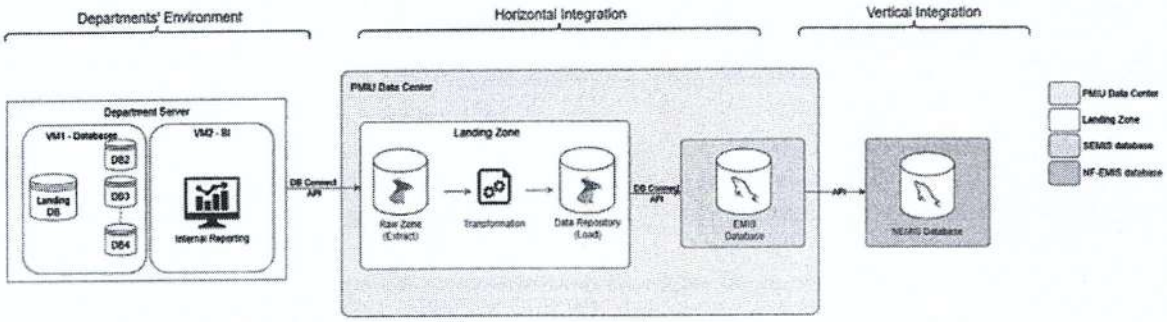


Figure 1: Architecture of HI

Steps for Horizontal Integration

A successful horizontal integration in the provinces is a 12-step process for entities targeted for horizontal integration. A brief description of each step is provided below:

1. **Live database**: A department needs to have a live database that is accessible and up to date with the latest data collected.
2. **Access to database**: A department must have full access to its databases and systems so that they can be integrated.
3. **Sufficient resources**: A department must have sufficient resources (hardware and software) to support the additional load of horizontal integration.
4. **Technical personnel**: A department must have technical personnel available to develop and extend APIs for horizontal integration.
5. **API development**: A department must have API(s) that can be used to share data with provincial EMIS or have the capability to develop them in case they don't exist.
6. **Data mapping**: The department's data should be mapped to conform to the data structures specified in Sapphire's API document that was developed for the vertical integration of data from provinces to NODP.
7. **Solution development**: A solution needs to be developed that performs data extraction, transformation and loading for horizontal integration.
8. **Testing**: The developed solution must be tested thoroughly to ensure that all modules are working as intended.
9. **Provisioning of landing zone**: Provincial EMIS to provision a landing zone that can be used as a development, testing and deployment environment for the solution.
10. **Deployment**: The developed solution must be deployed to the provisioned landing zone.
11. **Automation and scheduling**: Solution should be scheduled and automated to run on specified time intervals.
12. **Documentation**: The entire solution needs to be documented for smooth handover and continued usage.

Punjab

For the scope of this document, the following departments from Punjab are covered, segregated by their affiliation with the School Education Department (SED) in the table below:

Department	Abbreviation	Affiliation with SED
Private Education Provider Registration and Information System	PEPRIS	Attached formation
Punjab Education Foundation	PEF	Attached formation
Punjab Education Initiative Management Authority	PEIMA	Attached formation
Punjab Education, Curriculum, Training and Assessment Authority	PECTAA	Attached formation
Special Education Department	Special Ed.	Outside SED
Literacy & Non-Formal Basic Education Department	L&NFBED	Outside SED
Higher Education/ Colleges Department	HECD	Outside SED

Private Education Provider Registration and Information System (PEPRIS)

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	Provided by PMIU	Fully integrated	Portal access

The department is already horizontally integrated with PMIU.

Punjab Education Foundation (PEF)

Current Status

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	On Premises	Partially integrated	API

The department maintains its own dedicated database, which is internally managed by its technical team and securely hosted at its own premises. This arrangement provides PEF with autonomy in terms of data governance, security, and control over system maintenance. At present, PEF has already established a partial integration with the Programme Monitoring and Implementation Unit (PMIU). However, this integration was executed prior to the introduction of the Data Standardisation Framework (DSF). As a result, while certain variables are successfully exchanged between PEF and PMIU, there remains uncertainty regarding the DSF compliance of these variables. This raises the need for a systematic review to determine whether the existing dataset being shared is fully aligned with the new standards.

Requirements/ Aspirations

The department has not indicated any additional requirements in relation to hardware, software, or human resources for the intended horizontal integration.

Recommendations for HI

The department needs to systematically review the existing API to determine whether the existing dataset being shared with PMIU is fully aligned with the DSF 2.0 variables.

Punjab Education Initiative Management Authority (PEIMA)

Current Status

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	Provided by PITB	Not integrated	N/A

- **Servers:** The department has two databases and two systems hosted and managed by PITB. However, the details of these servers and systems in use were not provided. The contract with PITB will expire in December 2025.
- **Applications:** Two systems have been developed to meet the department's objectives. These systems are backed by two databases that are live and get directly populated from the sources.
- **Databases:** Two MySQL databases are currently in use that store the data collected by the department. This database contains approximately 3 years of data encompassing school, student, and teacher profiles. These databases hold information for 4,276 schools, 600,000 students and 20,000 teaching/non-teaching staff. However, the department is not aware of the volume per data collection activity or the current size of the database.
- **Technical Human Resource (HR):** The department currently has two dedicated technical personnel for data management. Moreover, the department has approved additional technical

positions to cater for the data requirements. Details of these approved positions were not shared.

- **Challenges:** Access to the relevant systems and databases was limited, and departmental representatives demonstrated only a partial understanding of the systems currently in use. Consequently, the representative was unable to specify precise hardware, software, and human resource requirements.

Requirements/ Aspirations

The representative showed interest in migrating their system to PMIU servers. However, should the department opt to establish in-house infrastructure for horizontal integration, the following requirements would tentatively apply based on the information provided:

Category	Item	Specification / Description	Units	Priority
Hardware	Standard Server	8 CPU cores, 64 GB RAM, 1TB SSD shared between Application, Database and Reporting VMs	1	Must
Hardware	UPS	As required	1	Optional
Software	Windows Server 2025 Standard	8-core license	1	Must
Software	PostgreSQL (Open source)	RDBMS, support costs apply	1	Must
Software	Python	To develop APIs and/or ETL workflows	1	Must
Personnel	Database Administrator	Manage & secure databases	1	Optional
Personnel	System Engineer	Maintenance and Upkeep of the server	1	Optional
Personnel	Data Engineer/ Developer	For data manipulation and APIs	1	Must

Annex E presents a detailed breakdown of the parameters used to determine the server specifications and outlines the limitations of the proposed infrastructure. Additionally, it includes the key requirements for each human resource role.

Recommendations for HI

In view of the limited access and control currently provided by PITB and considering that the department's existing contract is due to expire in December 2025, it is recommended that the department migrate its system to PMIU servers or to its own servers as prescribed above. Upon securing full access and administrative control, the department will be required to develop APIs or establish direct database connections with the landing zone provided by PMIU and ensure that all shared data is fully compliant with the variables prescribed under DSF 2.0.

Punjab Education, Curriculum, Training and Assessment Authority (PECTAA)

Current Status

Database	Hosting	Integration with PMIU	Integration Mechanism
Does not exist	N/A	N/A	N/A

PECTAA has been formed after a recent merger amongst three departments, including Punjab Education Commission (PEC), Punjab Curriculum and Textbooks Board (PCTB), and Quaid-e-Azam Academy for Educational Development (QAED). Currently, they do not have a live application or databases. PECTAA currently has nine technical resources who can be deployed to ensure seamless horizontal integration with PMIU.

Requirements/ Aspirations

The representatives showed interest in utilising the technical services of PMIU, including database development, hosting, and upkeep of applications and servers. However, should the department opt to establish in-house infrastructure for horizontal integration, the following requirements would apply based on the information provided:

Category	Item	Specification / Description	Units	Priority
Hardware	Standard Server	8 CPU cores, 64 GB RAM, 2TB SSD	1	Must
Hardware	UPS	As required	1	Optional
Software	Windows Server 2025 Standard	8-core license	1	Must

Category	Item	Specification / Description	Units	Priority
Software	PostgreSQL (Open source)	RDBMS, support costs apply	1	Must
Software	Python	To develop APIs and/or ETL workflows	1	Must
Personnel	Database Administrator	Manage & secure databases	1	Optional
Personnel	System Engineer	Maintenance and Upkeep of server	1	Optional
Personnel	Data Engineer/ Developer	For data manipulation and APIs	1	Must

Annex E presents a detailed breakdown of the parameters used to determine the server specifications and outlines the limitations of the proposed infrastructure. Additionally, it includes the key requirements for each human resource role.

Recommendations for HI

The department must set up its applications and databases. Once they have a live database available, the department must develop APIs or establish direct database connections with the landing zone provided by PMIU and ensure that all shared data is fully compliant with the variables prescribed under DSF 2.0.

Special Education Department

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	Provided by PITB	Not integrated	N/A

- **Servers:** The department has two databases and two systems hosted and managed by PITB. However, the details of these servers and systems in use were not provided. The contract with PITB has expired, but due to their strong political influence, the services provided by PITB are not discontinued.
- **Applications:** The department has developed applications to collect data and store it directly in the database.
- **Databases:** The database is developed in MySQL and contains one year of data encompassing school, student and teacher profiles. This database holds information for 130 schools, 40,000 students and 6,000 teaching/non-teaching staff. The department collects approximately 6.4 GB of data per data collection activity, but the total volume of data in the database is unknown.
- **Technical Human Resource (HR):** No dedicated technical personnel are available to manage the systems. However, 19 technical positions for hiring have been approved by the department.
- **Challenges:** Access to the relevant systems and databases was limited, and departmental representatives demonstrated only a partial understanding of the systems currently in use. Consequently, the representative was unable to specify precise hardware, software, and human resource requirements.

Requirements/ Aspirations

The representative showed interest in migrating their system to PMIU servers. However, should the department opt to establish in-house infrastructure for horizontal integration, the following requirements would apply based on the information provided:

Category	Item	Specification / Description	Units	Priority
Hardware	Standard Server	8 CPU cores, 64 GB RAM, 1TB SSD	1	Must
Hardware	UPS	As required	1	Optional
Software	Windows Server 2025 Standard	8-core license	1	Must
Software	PostgreSQL (Open source)	RDBMS	1	Must
Software	Python	To develop APIs and/or ETL workflows	1	Must
Personnel	Database Administrator	Manage & secure databases	1	Optional
Personnel	System Engineer	Maintenance and Upkeep of server	1	Optional

Category	Item	Specification / Description	Units	Priority
Personnel	Data Engineer/ Developer	For data manipulation and APIs	1	Must

Annex E presents a detailed breakdown of the parameters used to determine the server specifications and outlines the limitations of the proposed infrastructure. Additionally, it includes the key requirements for each human resource role.

Recommendations for HI

In view of the limited access and control currently provided by PITB and considering that the department's existing contract has already expired, it is recommended that the department migrate its system to PMIU servers or to its own servers as prescribed above. Upon securing full access and administrative control, the department will be required to develop APIs or establish direct database connections with the landing zone provided by PMIU and ensure that all shared data is fully compliant with the variables prescribed under DSF 2.0.

Literacy & Non-Formal Basic Education Department (L&NFBED)

Database	Hosting	Integration with PMIU	Integration Mechanism
No information	No information	Not integrated	N/A

No information could be obtained regarding L&NFBED, as the department neither attended any meeting nor shared any information regarding their database system.

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration. If the database exists, it should be integrated with PMIU using an API, otherwise, a standard data entry portal should be provided by PMIU to the L&NFBED, where the department can input the required data on regular intervals. In case the data entry portal is provided, the department will require one data entry operator.

Higher Education/ Colleges Department (HECD)

Database	Hosting	Integration with PMIU	Integration Mechanism
No information	No information	Not integrated	N/A

No information could be obtained regarding HECD, as the department neither attended any meeting nor shared any information regarding their database system.

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration. If the database exists, the database should be integrated with PMIU using an API, otherwise, a standard data entry portal should be provided by PMIU to the HECD, where the department can input the required data on regular intervals.

Provincial EMIS: PMIU

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	On Premises	N/A	N/A

Requirements for HI

Landing zone: PMIU needs to provision a landing zone that will be utilised to extract and transform data from other departments expected to be horizontally integrated. The recommended specifications of the landing zone are given below:

Type	Version
CPU	Multi-core x64 processor (≥ 8 cores)
OS	Windows 10 or above
RAM	32GB to 64GB
Storage	500GB to 1TB High-speed SSD
Database	SQL Server 2019 Developer Edition, PostgreSQL, or MySQL
ETL Tool	e.g., Talend Open Studio
IDE	e.g., SSMS, DBeaver
Python Environment	Anaconda

Development of Data Portal: A standardised data portal should be developed for departments that fall outside the purview of SED and lack APIs or direct database connectivity to share data with PMIU.

Sindh

For the scope of this document, the following departments from Punjab are covered, segregated by their affiliation with the School Education Department (SED) in the table below:

Department	Abbreviation	Affiliation with SED
Sindh Education Foundation	SEF	Attached formation
Directorate of Inspection and Registration of Private Institutes Sindh	DIRPIS	Attached formation
Sindh Teachers Education Development Authority	STEDA	Attached formation
Department of Empowerment of Persons with Disabilities	DEPD	Outside SED
Directorate of Literacy and Non-formal Education	DLNFE	Outside SED
Higher Education/ Colleges Department	HECD	Outside SED

Sindh Education Foundation

Database	Hosting	Integration with RSU	Integration Mechanism
Exists	On Premises	Integrated	Direct database connection

The department is already horizontally integrated with RSU.

Recommendations for HI

The department needs to perform a thorough testing of its system to mitigate the data quality issues. The details of such data quality issues can be found in Annex A.

Directorate of Inspection and Registration of Private Institutes Sindh (DIRPIS)

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	On cloud hosting	Not integrated	N/A

The database is managed by a third-party private firm, and the department does not require any additional resources or support for horizontal integration.

Recommendations for HI

The department needs to provide APIs to RSU and ensure that the shared data is fully aligned with the DSF 2.0 variables.

Sindh Teachers Education Development Authority (STEDA)

Database	Hosting	Integration with RSU	Integration Mechanism
Exists	No information	Not integrated	N/A

- **Servers:** The department utilises an online application for data collection. However, detailed information regarding this application has not been provided.
- **Applications:** The department utilises an online application for data collection. However, detailed information regarding this application has not been provided.
- **Databases:** The department maintains a live database containing approximately 240,000 records. Out of these, around 150,000 were collected through the application portal, while the remaining records were obtained via direct registrations
- **Technical Human Resource (HR):** The department has only one technical resource that manages its system and data requirements.
- **Challenges:** Since no technical personnel from the department attended the session, the department's current challenges and needs could not be fully assessed.

Requirements/ Aspirations

The department needs additional technical human resources to manage the process of API development and horizontal integration smoothly.

Recommendations for HI

- The provision of the following three additional personnel is required, with specialised profiles including a Data Engineer, a Database Administrator, and a Systems and Network Engineer.
- Develop APIs or provide direct database connection for sharing data with RSU, ensuring that all applicable variables are aligned with DSF 2.0.

Annex E provides key requirements for each human resource role.

Department of Empowerment of Persons with Disabilities (DEPD)

Database	Hosting	Integration with RSU	Integration Mechanism

No information	No information	Not integrated	N/A

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration. If the database exists, the database should be integrated with RSU using an API, otherwise, a standard data entry portal should be provided by RSU to the DEPD, where the department can input the required data on regular intervals. In case the data entry portal is provided, the department will require one data entry operator.

Directorate of Literacy and Non-formal Education (DLNFE)

Database	Hosting	Integration with RSU	Integration Mechanism
Exists	COMSATS, Islamabad	Integrated	API

The department is already horizontally integrated with RSU.

Recommendations for HI

The department needs to perform a thorough testing of its system to mitigate the data quality issues. The details of such data quality issues can be found in Annex B.

Higher Education/ Colleges Department (HECD)

Database	Hosting	Integration with PMIU	Integration Mechanism
No information	No information	Not integrated	N/A

No information could be obtained regarding HECD, as the department neither attended any meeting nor shared any information regarding their database system. If the database exists, the database should be integrated with PMIU using an API, otherwise, a standard data entry portal should be provided by PMIU to the HECD, where the department can input the required data on regular intervals.

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration.

Provincial EMIS: RSU

Database	Hosting	Integration with PMIU	Integration Mechanism
Exists	On Premises	N/A	N/A

Requirements for HI

Landing zone: RSU has already provisioned a landing zone with the specifications given below:

Type	Version
CPU	Multi-core x64 processor (≥ 8 cores)
OS	Windows 10 or above
RAM	32GB to 64GB
Storage	500GB to 1TB High-speed SSD
Database	SQL Server 2019 Developer Edition, PostgreSQL, or MySQL
ETL Tool	e.g., Talend Open Studio
IDE	e.g., SSMS, DBeaver
Python Environment	Anaconda

RSU has sufficient hardware and software resources to manage complete horizontal integration in the province, however, they will require additional human resources to ensure the seamless integration of these systems.

Annex E provides key requirements for each human resource role.

Khyber Pakhtunkhwa

For the scope of this document, the following departments from KP are covered, segregated by their affiliation with the School Education Department (SED) in the table below:

Department	Abbreviation	Affiliation with SED
Private Schools Regulatory Authority	PSRA	Attached formation
Merged Areas Education Foundation	MAEF	Attached formation
Elementary and Secondary Education Foundation	ESEF	Attached formation
Directorate of Professional Development	DPD	Attached formation
Education Monitoring Authority	EMA	Attached formation
Directorate of Social Welfare	DSW	Outside SED
Higher Education Regulatory Authority	HERA	Outside SED
Higher Education/ Colleges Department	HECD	Outside SED

Private Schools Regulatory Authority (PSRA)

Database	Hosting	Integration with RSU	Integration Mechanism
Exists	Provided by KPDC	Not integrated	API

- **Servers:** The department has a live database and an application hosted on servers provided by KPDC. However, the details of these servers and systems in use are not provided.
- **Applications:** An online application is developed to meet the department's data collection and reporting objectives. This application is connected directly to a database to store data.
- **Databases:** The department uses a MySQL database to store data related to schools, students, and teacher profiles. It collects approximately 5 gigabytes of data annually. However, the total size of the database is unknown.
- **Technical Human Resource (HR):** The department has only one technical resource that manages its system and data requirements.
- **Challenges:** Lack of technical personnel to support development and enhance systems.

Requirements/ Aspirations

The department requires additional technical personnel to effectively manage and enhance APIs, address data quality issues, and ensure seamless horizontal integration.

Recommendations for HI

- The provision of the following three additional personnel is required, with specialised profiles including a Data Engineer, a Database Administrator, and a Systems and Network Engineer. The profile details can be found in Annex E.
- Develop APIs or provide direct database connection for sharing data with KP i-EMIS, ensuring that all applicable variables are aligned with DSF 2.0.
- The department needs to perform a thorough testing of its system to identify data quality issues and take necessary actions to mitigate them. The details of such data quality issues can be found in Annex C.

Merged Areas Education Foundation (MaEF)

Database	Hosting	Integration with RSU	Integration Mechanism
Exists	Provided by KPDC	Not integrated	Direct database connection

- **Servers:** The department has a live database and an application hosted on servers provided by KPDC. However, the details of these servers and systems in use are not provided.
- **Applications:** An online application has been developed to support the department's data collection and reporting objectives. It is directly connected to a live database for data storage.
- **Databases:** The department uses a MySQL database to store data related to schools, students, and teacher profiles. It can collect up to 15 gigabytes of data per year and currently has approximately 20 gigabytes of data accumulated since 2022.
- **Technical Human Resource (HR):** The department has only one technical resource that manages its system and data requirements.
- **Challenges:** Lack of technical personnel to support development and enhance systems.

Requirements/ Aspirations

The department needs additional technical human resources to manage the process of API development and horizontal integration smoothly.

Recommendations for HI

- The provision of the following three additional personnel is required, with specialised profiles including a Data Engineer, a Database Administrator, and a Systems and Network Engineer. The profile details can be found in Annex E.
- Develop APIs or provide direct database connection for sharing data with KP i-EMIS, ensuring that all applicable variables are aligned with DSF 2.0.
- The department needs to provide appropriate permissions to access its database in read-only mode.

Annex E provides key requirements for each human resource role.

Elementary and Secondary Education Foundation (ESEF)

Database	Hosting	Integration with RSU	Integration Mechanism
Exists	Provided by KPDC	Not integrated	API

- **Servers:** The department has a live database and an application hosted on servers provided by KPDC.
- **Applications:** An online application has been developed to support the department's data collection and reporting objectives. It is directly connected to a live database for data storage.
- **Databases:** The department uses a MySQL database to store data related to schools, students, and teacher profiles. However, the volume of data generated per collection activity and the overall size of the database is unknown.
- **Technical Human Resource (HR):** The department has technical personnel available to develop APIs and/or provide direct database connections.
- **Challenges:** No information provided.

Requirements/ Aspirations

No information provided.

Recommendations for HI

- The provision of the following three additional personnel is required, with specialised profiles including a Data Engineer, a Database Administrator, and a Systems and Network Engineer.
- The department needs to perform a thorough testing of its system to identify data quality issues and take necessary actions to mitigate them. The details of such data quality issues can be found in Annex D.

Annex E provides key requirements for each human resource role.

Directorate of Professional Development (DPD)

Database	Hosting	Integration with i-EMIS	Integration Mechanism
Exists	Provided by KPDC	Partially integrated	API

DPD has already established a partial integration with the i-EMIS. However, this integration was executed prior to the introduction of the Data Standardisation Framework 2.0(DSF). As a result, while certain variables are successfully exchanged between DPD and i-EMIS, there remains uncertainty regarding the DSF compliance of these variables. This raises the need for a systematic review to determine whether the existing dataset being shared is fully aligned with the new standards.

Recommendations for HI

The department needs to systematically review the existing API to determine whether the existing dataset being shared with PMIU is fully aligned with the DSF 2.0 variables

Directorate of Social Welfare, Special Education & Women Empowerment (SWSE&WE)

Database	Hosting	Integration with i-EMIS	Integration Mechanism
No information	No information	Not integrated	N/A

No information could be obtained regarding SWSE&WE, as the department neither attended any meeting nor shared any information regarding their database system.

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration. If the database exists, the database should be integrated with i-EMIS using an API, otherwise, a standard data entry portal should be provided by i-EMIS to the SWSE&WE, where the department can input the required data on regular intervals. In case the data entry portal is provided, the department will require one data entry operator.

Education Monitoring Authority (EMA)

Database	Hosting	Integration with i-EMIS	Integration Mechanism
Exists	Provided by KPDC	Not integrated	N/A

- **Servers:** The department has a live database and multiple applications hosted on servers provided by KPDC. However, the details of these servers and systems in use are not provided.
- **Applications:** The department has developed an Android application that collects data from schools and stores it directly in the database.
- **Databases:** The department collects data on schools, students, and teacher profiles, which is directly stored in a Microsoft SQL Server database. Each data collection activity can yield up to 500 gigabytes of data per year. The database contains records spanning the past 11 years, with a total size of approximately 2,500 gigabytes (2.5 TB).
- **Technical Human Resource (HR):** The department has a dedicated technical team that can manage all its data and system-related requirements.
- **Challenges:** KPEMA has a mandate-related conflict, which has restricted the sharing of its data. However, should the department decide to make its data available, it already possesses sufficient resources and technical capacity to facilitate horizontal integration through the development of APIs.

Recommendations for HI

The department needs to develop or update its APIs to share data with i-EMIS that is fully aligned with the DSF 2.0 variables.

Higher Education Regulatory Authority (HERA)

Database	Hosting	Integration with i-EMIS	Integration Mechanism
No information	No information	Not integrated	N/A

No information could be obtained regarding HERA, as the department neither attended any meeting nor shared any information regarding their database system.

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration. If the database exists, the database should be integrated with i-

EMIS using an API, otherwise, a standard data entry portal should be provided by i-EMIS to the HERA, where the department can input the required data on regular intervals. In case the data entry portal is provided, the department will require one data entry operator.

Higher Education/ Colleges Department (HECD)

Database	Hosting	Integration with i-EMIS	Integration Mechanism
No information	No information	Not integrated	N/A

No information could be obtained regarding HECD, as the department neither attended any meeting nor shared any information regarding their database system.

Recommendations for HI

A comprehensive mapping of the existing system is required to establish a baseline for assessing readiness for horizontal integration. If the database exists, the database should be integrated with i-EMIS using an API, otherwise, a standard data entry portal should be provided by i-EMIS to the HECD, where the department can input the required data on regular intervals.

Provincial EMIS: i-EMIS

Database	Hosting	Integration with i-EMIS	Integration Mechanism
Exists	On Premises	N/A	N/A

Requirements for HI

Landing zone: i-EMIS needs to provision a landing zone that will be utilised to extract and transform data from other departments expected to be horizontally integrated. The recommended specifications of the landing zone are given below:

Type	Version
CPU	Multi-core x64 processor (≥ 8 cores)
OS	Windows 10 or above
RAM	32GB to 64GB
Storage	500GB to 1TB High-speed SSD
Database	SQL Server 2019 Developer Edition, PostgreSQL, or MySQL
ETL Tool	e.g., Talend Open Studio
IDE	e.g., SSMS, DBeaver
Python Environment	Anaconda

i-EMIS has sufficient hardware and software resources to manage complete horizontal integration in the province; however, it will require additional human resources to ensure the seamless integration of these systems.

Annex E provides key requirements for each human resource role.

Balochistan

For the scope of this document, the following departments from Balochistan are covered, segregated by their affiliation with the School Education Department (SED) in the table below:

Department	Abbreviation	Affiliation with SED
Balochistan Education Foundation	BEF	Attached formation
Private Educational Institutions Registration and Regulation Authority	PEIRA	Attached formation
Provincial Institute for Teacher Education	PITE	Attached formation
Social Welfare Department (Special and Non-Formal Education)	SWD	Outside SED
Higher Education/ Colleges Department	HECD	Outside SED

At present, BEF, PIERA, PITE, and SWD do not maintain live or integrated databases. Meanwhile, the BEMIS, managed by the PPIU, is undergoing a major transition as its legacy platform is being upgraded with support from UNICEF, with completion expected by December 2025. This upgrade aims to modernise data management processes and enhance interoperability across departments. Once the upgraded system becomes operational, the attached formations will be provided with dedicated portals to facilitate data entry. To support this, each department will require designated data entry personnel. In parallel, the PPIU will need to engage technical staff—such as data engineers—to ensure the system's smooth operation and sustain efficient performance. Finally, the entity responsible for the Horizontal Integration (HI) activity must ensure that departments outside the School Education Department (SED) are also granted access to these portals, enabling comprehensive data integration across the sector.

Training of Technical Staff

To ensure that all technical personnel across horizontally integrated departments remain aligned, capable, and responsive to evolving system requirements, a structured year-long training programme should be instituted. The objective is to foster coherence across teams by building a shared understanding of the digital systems, data structures, and tools that form the backbone of operations. This initiative should emphasise both skill enhancement and cross-departmental collaboration, ensuring that technical representatives are not only proficient in their domains but also attuned to the broader ecosystem they operate within.

The training framework should encompass key technical areas such as Horizontal Integration, Data Standardisation Framework (DSF), Python programming, SQL Server management, and API development and maintenance. These components are essential for sustaining a robust and interoperable system architecture. Participants should gain practical knowledge in system connectivity, automation, and data management—critical for maintaining seamless integration between departments. Regular sessions should focus on hands-on learning and real-world troubleshooting to build confidence and deepen expertise.

Given the importance of contextual awareness, an additional dimension of the programme should cover sectoral understanding, particularly in the education sector. This will enable technical staff to better appreciate the operational realities and policy contexts within which their systems function, thereby improving the relevance and responsiveness of their technical solutions.

Training sessions should be conducted quarterly, with each module combining workshops, system demonstrations, and feedback sessions. These should serve as platforms for technical representatives to raise concerns, share challenges, and collaboratively resolve issues. Over time, this continuous cycle of learning and engagement should help build a cohesive, informed, and proactive technical community. The initiative should also act as a conduit for familiarising staff with upcoming software and hardware updates, ensuring that every participant remains current with technological advancements.

By the end of the first year, this programme should establish not just technical proficiency, but a culture of collective problem-solving and systems thinking—laying the foundation for sustained digital transformation across all horizontally integrated departments.

Vertical Integration

Provincial EMIS

The following table outlines the list of provincial EMIS that are integrated with NODP via application programming interface (API).

S. No	Department	Integration Status	Notes
1.	PMIU	Integrated	At present, integration is confined to EMIS data only. To ensure data sharing of horizontally integrated datasets with NODP, the department must either consolidate integrated datasets into its EMIS database, thereby enabling the transfer of both EMIS data and horizontally integrated datasets, or extend the existing APIs to directly retrieve data from the landing zone.
2.	RSU	Integrated	
3.	i-EMIS	Integrated	
4.	PPIU	Integrated	

National and Federal Entities

The following table outlines the entities identified for vertical integration with the NODP, together with their respective integration status as of October 2025.

S. No	Sector	Department	Integration Status	Notes
1.	Military and Defence Institutions	Army Public Schools and Colleges System (APSACS)	Not integrated	
2.	Military and Defence Institutions	Bahria Education & Training Services (BEATS)	Not integrated	
3.	Military and Defence Institutions	Federal Government Educational Institutions (Cantts/Garrisons)	Not integrated	
4.	Military and Defence Institutions	Pakistan Air Force (PAF)	Not integrated	
5.	Military and Defence Institutions	Pakistan Navy (PN)	Not integrated	
6.	Military and Defence Institutions	Pakistan Ordnance Factories (POF)	Not integrated	
7.	Military and Defence Institutions	Cadet Colleges	Not integrated	
8.	Public Educational Networks	Daanish Schools and Centre of Excellence Authority	Not integrated	
9.	Public Educational Networks	Divisional Public Schools & Inter Colleges (DPS)	Not integrated	
10.	Corporate and Institutional Schools	Fauji Foundation (FF)	Not integrated	
11.	Corporate and Institutional Schools	Pakistan Atomic Energy Commission (PAEC)	Not integrated	
12.	Corporate and Institutional Schools	Pakistan Railways (PR)	Not integrated	
13.	Corporate and Institutional Schools	Pakistan Steel Mills	Not integrated	
14.	Corporate and Institutional Schools	Telecom Foundation	Not integrated	
15.	Corporate and Institutional Schools	Water and Power Development Authority (WAPDA)	Not integrated	
16.	Corporate and Institutional Schools	Mines & Minerals Department, National Police	Not integrated	
17.	Corporate and Institutional Schools	Foundation, Overseas Pakistanis Foundation (OPF)	Not integrated	
18.	Corporate and Institutional Schools	Pakistan Bait-ul-Mal	Not integrated	
19.	Madaris		Not integrated	
20.	Special Education	Directorate General of Special Education (DGSE), Islamabad	Not integrated	

S. No	Sector	Department	Integration Status	Notes
21.	Technical and Vocational Education	National Vocational & Technical Training Commission (NAVTTC)	Not integrated	
22.	Distance Learning	Allama Iqbal Open University (AIOU)	Not integrated	
23.	Public	Higher Education Commission (HEC)	Not integrated	
24.	Public	Federal Directorate of Education (FDE)	Not integrated	
25.	Public	Private Educational Institutions Regulatory Authority (PEIRA)	Not integrated	
26.	Public	Directorate General of Religious Education (DGRE)	Not integrated	
27.	Public	Basic Education Community Schools (BECS)	Integrated	User of NFEMIS
28.	Public	National Commission for Human Development (NCHD)	Integrated	User of NFEMIS
29.	Public	National Education Foundation (NEF)	Integrated	User of NFEMIS

Pakistan Institute of Education (PIE)

PIE serves as the custodian of the NODP, which will house integrated education data from across the country. As custodian, PIE will receive both horizontally and vertically integrated data, ensuring consistency and interoperability across systems. Over 150 entities are targeted for integration with the NODP, connecting provincial and federal data sources into a unified national platform.

Annexures

Annex A: Sindh Education Foundation

Data Quality Issues

Following are the data quality issues in Sindh Education Foundation's data.

1. Teacher Professional Qualification: This field is expected to give teachers' professional qualification such as 'M.Ed.', 'PTC', 'ECE' etc. On the other hand, the entries in this field are misspelled, invalid and missing. For example, 'Computer', 'B. Ad', 'Enter', 'N A' etc. making the data difficult to map accurately.
2. Staff Training: This field is supposed to have a 'Yes/No' indicators to show whether a teacher has received any pedagogical training or not. Whereas it contains various training names including missing and invalid values such as 'Head Teacher', 'N/A', '21st century' etc. which cannot be mapped to the expected binary format.
3. Staff Designation: This field is intended to contain teachers' designations such as 'Principal', 'Vice Principal', 'Subject Specialist (Physics)' etc. However, even after applying the segregation between teaching and nonteaching staff, several non-teaching designations such as 'Chowkidar', 'Sweeper', 'Peon' etc. still appear under the teaching staff category, indicating data misclassification.
4. Student Disability: This field contains students' disabilities such as 'Blind', 'Polio', 'Dumb' etc. Whereas it is supposed to give difficulty types as: 'Seeing, even when wearing glasses', 'Hearing, even if using a hearing aid', 'Walking or climbing steps' and 'Remembering or concentrating'. Hence, this field is not aligned with required classification framework.
5. School Contact Number: Around 70 schools do not have contact numbers including mobile numbers and landline numbers.

Missing Columns

Following columns are missing in Sindh Education Foundation's data.

Teacher Profile	Student Profile	School Profile
Census_Year	Studied_ECE_Kachi_Last_Year	Census_Year
BasicPayScale_Id	StudentStatus_Id	Village
Acquire_TrainingForSpecialChildren	DifficultyType_Id	Circle_Name
Acquire_ECETraining	DifficultyCategory_Id	City
Date_Of_Current_Posting	Stipend_Scholarship_Holder	PA_No
NatureOfService_Id		NA_No

Teacher Profile	Student Profile	School Profile
Current_Year_Training_Duration		Fax_No
DifficultyType_Id		Phone_No
DifficultyCategory_Id		Fax_No
Date_Of_Retirement		Tehsil_Id
		District_Id
		Location_Id
		FuncationalStatus_Id
		Management_Id
		Kind_Id
		SchoolCommittee_Id
		Altitude

Further details of Horizontal Integration process and Source mapping can be found in Horizontal Integration ETL Pipeline for SEF, Sindh and Source Matrix – SEF, Sindh, which are shared separately.

Annex B: Directorate of Literacy and Non-formal Education

Data Quality Issues

Following are the data quality issues in Directorate of Literacy and Non-formal Education, Sindh's Data and all other departments of Non-formal Education.

1. Student Disability: This field contains students' disabilities such as 'Cognition', 'Hearing', 'Self-Care' etc. Whereas it is supposed give difficulty types such as: 'Seeing, even when wearing glasses', 'Hearing, even if using a hearing aid', 'Walking or climbing steps' and 'Remembering or concentrating'. Hence, the values in the field are not aligned with the required classification framework.
2. District Name: Some of the districts have inconsistent spellings compared to their official names. For example, 'Lakki Marwat' is spelled as 'Lakki', 'Malakand' as 'Malakand Protected Area' etc. resulting in inconsistencies and difficulty in mapping.

Missing Columns

Following columns are missing in Directorate of Literacy and Non-formal Education, Sindh's Data and in the data of other non-formal education departments.

Teacher Profile	Student Profile	School Profile
Census_Year	Census_Year	Census_Year
BasicPayScale_Id	Studied_ECE_Kachi_Last_Year	Village
Acquire_TrainingForSpecialChildren	StudentStatus_Id	Circle_Name
Acquire_ECETraining	DifficultyType_Id	City
Date_Of_Current_Posting	DifficultyCategory_Id	PA_No
Designation_Id	Stipend_Scholarship_Holder	NA_No
NatureOfService_Id	Shift_Id	Fax_No
Current_Year_Training_Duration		Phone_No
DifficultyType_Id		Fax_No
DifficultyCategory_Id		Website
Date_Of_Retirement		Email
		Location_Id
		FuncationalStatus_Id
		Established_Date
		Management_Id
		Kind_Id
		Shift_Id
		SchoolCommittee_Id
		Sector_Id
		Altitude

Further details of Horizontal Integration process and Source mapping can be found in Horizontal Integration ETL Pipeline for NFE, Sindh and Source Matrix – NFE, Sindh, which are shared separately.

Annex C: KP Private Schools Regulatory Authority

Data Quality Issues

Following are the data quality issues in KP Private Schools Regulatory Authority's Data.

1. schoolStaffDesignation: This field is intended to contain teachers' designations such as 'Principal', 'Subject Specialist (Physics)' etc. However, even after applying the segregation between teaching and non-teaching staff, it has many non-teaching designations under teaching category such as 'Khala', 'Guard', 'S.GUARD' etc., indicating data misclassification.
2. schoolStaffQualificationAcademic: This field is supposed to give teachers' academic qualification such as 'Matric', 'F.Sc.', 'MA' etc. On the other hand, the entries in this field are misspelled, invalid and missing. For example, 'Metric', '???? ???????', 'N/A', 'BAy.shahadatulalamia' etc., making the data difficult to map and standardize.
3. schoolStaffQualificationProfessional: This field is supposed to give teachers' professional qualification such as 'M.Ed.', 'PTC', 'ECE' etc. However, the entries in this field are misspelled, invalid and missing. For example, 'BS (BIO)', '???? ???????', '10th', '8th Shahadataul Alamia' etc., making the data difficult to map and standardize.

Missing Columns

For PSRA, the student profile information is not available at the required level of detail and, therefore, cannot be utilized. In addition, the following data fields are missing from the KP Private Schools Regulatory Authority's dataset:

Teacher Profile	School Profile
Census_Year	Census_Year
BasicPayScale_Id	Village
Acquire_PedagogicalTraining	Circle_Name
Acquire_TrainingForSpecialChildren	City
Acquire_ECETraining	PA_No
Date_Of_Current_Posting	NA_No
Designation_Id	Fax_No
NatureOfService_Id	Phone_No
Current_Year_Training_Duration	Fax_No
DifficultyType_Id	Website
DifficultyCategory_Id	Email
Date_Of_Retirement	Location_Id
Date_Birth	FuncationalStatus_Id
	Established_Date
	Management_Id
	Kind_Id
	Shift_Id
	SchoolCommittee_Id
	Sector_Id
	Altitude
	UC_Name
	UC_No
	Tehsil_Id
	District_Id

Further details of Horizontal Integration process and Source mapping can be found in Horizontal Integration ETL Pipeline for PSRA, KP and Source Matrix – PSRA, KP, which are shared separately.

Annex D: KP Elementary and Secondary Education Foundation

Data Quality Issues

Following are the data quality issues in KP Elementary and Secondary Education Foundation.

1. Teacher Qualification: This field is supposed to give teachers' academic qualification such as 'Matric', 'F.Sc.', 'MA' etc. However, the entries in this field are invalid and non-uniform. For example, '?', 'B,A' etc., making the data difficult to map and standardize.
2. Teacher Professional Qualification: This field is supposed to give teachers' professional qualification such as 'M.Ed.', 'PTC', 'ECE' etc. Whereas the entries in this field are invalid, non-uniform and multivalued. For example, 'B.Ed ,M.Ed,CT,PST', 'TT', etc., making the data difficult to map and standardize.
3. Student disability_details: This field contains students' disabilities such as 'Left leg', 'eye', 'Artificial leg' etc. On the other hand, it is supposed to give difficulty types as: 'Seeing, even when wearing glasses', 'Hearing, even if using a hearing aid', 'Walking or climbing steps'

and 'Remembering or concentrating'. Hence, the values in this field are not aligned with the required classification framework.

4. Student disability: This field contains students' disabilities such as '1 hand is disable', 'Blindness', 'Brain' etc. However, it is supposed to give difficulty categories as: 'Not Reported', 'No Difficulty', 'Some Difficulty', 'Lot of difficulty' and 'Cannot do ay all'. Hence, the values in this field are not aligned with the required classification framework.
5. Schools y_cord/x_cord: These two fields are meant to represent latitude and longitude. Whereas it is not confirmed which field corresponds to latitude and which one to longitude, leading to potential ambiguity in location mapping.

Missing Columns

Following columns are missing in KP Private Schools Regulatory Authority's Data.

Teacher Profile	Student Profile	School Profile
Census_Year	Census_Year	Census_Year
BasicPayScale_Id	Student_Id	Village
Acquire_PedagogicalTraining	Studied_ECE_Kachi_Last_Year	Circle_Name
Acquire_TrainingForSpecialChildren	StudentStatus_Id	City
Acquire_ECETraining	DifficultyType_Id	PA_No
Date_Of_Current_Posting	DifficultyCategory_Id	NA_No
Designation_Id	Stipend_Scholarship_Holder	Fax_No
NatureOfService_Id	Shift_Id	Phone_No
Current_Year_Training_Duration		Fax_No
DifficultyType_Id		Website
DifficultyCategory_Id		Email
Date_Of_Retirement		Location_Id
		FuncationalStatus_Id
		Established_Date
		Management_Id
		Kind_Id
		Shift_Id
		SchoolCommittee_Id
		Sector_Id
		Altitude

Further details of the Horizontal Integration process and Source mapping can be found in Horizontal Integration for ESEF, KP and Source Matrix – ESEF, KP, which are shared separately.

Annex E: Requirement Details

Parameters

All recommendations for hardware and software are calculated based on the input provided by the departments. Following are the parameters used to make calculations.

Parameter	Description
Total Students	Total student count for the department.
Total Teachers	Total teacher count for the department.
Total Schools	Total school count for the department.
Student Growth Rate	By what percentage does the student count increase Year over Year.
Teacher Growth Rate	By what percentage does the teacher count increases Year over Year.
School Growth Rate	By what percentage does the school count increase Year over Year.
Student Profile Size (MB)	How many MBs data one student generate in its life cycle.
Teacher Profile Size (MB)	How many MBs data one teacher generate in its life cycle.
School Profile Size (MB)	How many MBs data one school generate in its life cycle.
RAM to Storage Ratio	32 GB of RAM for 1 TB SSD.
Data Processing per physical Core	200GB per physical Core

Limitations

ID	Category	Assumption / Limitation	Impact / Risk
1.	General	Proposed IT setup caters to Horizontal Integration needs only. Any other IT needs are not included in the dimensioning and design.	Each Department may assess additional needs and add on top
2.	DSF related data	Proposed IT setup is designed to cater for DSF data needs only (text and numeric data for student, teacher, building and similar profiles).	Any other type of needs e.g. Monitoring Data (pictures / videos / other file formats) are not considered, and additional components can be added to the setup if required.
3.	Started configurations	In most cases, started configurations have been proposed that are sufficient for the requirement at hand. This is done with the intent of making it easy to move forward and start the Horizontal Integration journey in a short timeframe.	If more advanced configurations are sought due to needs in addition to (1) and (2), Departments may choose to do so.
4.	Enterprise Level Infrastructure	As discussed above, the recommended configuration is sufficient to meet the requirements of horizontal integration. However, if the department intends to establish an enterprise-level infrastructure, the following components would be required: 1. Production Server 2. Development Server 3. Testing Server 4. Web Server 5. ETL Server 6. Backup Servers 7. VPN and Virtualisation Software The exact hardware specifications and software requirements will depend on the use cases, which are to be defined and provided by the department	
5.	Data Source(s)	The proposed infrastructure assumes that data will be sourced from internal and secure systems. Should external or unsecured data sources be utilized, the department would need to implement a hardware firewall to ensure enhanced data protection and mitigate the risk of data breaches	
6.	Inputs	Inputs including total student, teacher and student count provided is accurate	Any misquoted numbers or wrong numbers could affect the appropriate allocation for systems required
7.	Profile Size	Profile size for each student, teacher and school is provided by RSU. It is assumed that 1 MB size represents	Directly impacts resources and systems required

ID	Category	Assumption / Limitation	Impact / Risk
		entire life cycle in the system for each profile	
8.	Server Type	Tower servers chosen for cost-efficiency and low-density setups.	May not scale efficiently for larger clusters.
9.	CPU Allocation	Estimated 1 core per 200 GB of storage as a planning heuristic.	Can vary by workload (ETL jobs vs. analytics).
10.	RAM Sizing	A standard RAM to Storage ratio is used for the calculation i.e. 32 GB RAM for 1 TB of SSD. This ratio is optimal for a power user to perform HI tasks and analytics of medium complexity	For metadata-heavy use cases, actual needs can differ.
11.	Growth Assumption	Growth numbers were not provided by the departments. To calculate growth potential, the percentage growth data from RSU has been used. The following are the growth percentages provided by RSU: 1. Students: 9% (applied growth rate: 9%) 2. Teachers: 0.5–1% (applied growth rate: 9%, to ensure an appropriate teacher-pupil ratio) 3. Schools: <1% (applied growth rate: 1%)	Directly impacts scalability and capacity planning.
12.	Software Licensing	Windows Server assumed on each physical node	Affects licensing cost calculation.
13.	Workload	Designed for ETL, data storage, and push jobs	May need higher spec systems for analytics-intensive workload
14.	Scalability	All recommended hardware specification of server is adjusted to server the purpose of horizontal integration for up to three years.	Assuming the department grows by the assumed growth rate. Higher data growth rate (for various reasons) will require additional hardware.
15.	Backups	The provided specifications pertain to the active system and may also be utilized for storing backups. However, if a department intends to establish a Disaster Recovery (DR) setup, the specified requirements should be doubled accordingly	

Human Resource Profiles

Data Engineer

1. BS in Computer Science, Data Science or related field
2. Strong analytical skills with the ability to collect, organise, analyse, and disseminate significant amounts of information
3. 2+ years of experience in data engineering or a related role
4. Proficiency in data pipeline development, including ETL processes
5. Build processes supporting data transformation, data structures, metadata, dependency and workload management
6. Hands-on experience with Python is a must
7. Hands-on experience with SQL queries

System Administrator

1. BS in Computer Science, IT or related field
2. 2+ years of experience in system engineer and network adimistrator role
3. Responsible to monitor and manage the network and relevant devices (switches, routers, Firewalls, wireless etc.)
4. Proactively manage and maintain server, network, and firewall systems
5. Installation of new hardware and software (Physical Servers, VMs, computer workstations/laptops, Network Devices (Switches, Routers, Access Points, etc.)
6. Administer and support core Microsoft, Linux, Database Servers, Antivirus solutions, Hyper-V, and VMware technologies
7. Design and implement monitoring, configuration management and reporting functions that will make a hands-off environment
8. Responsible to manage the operating systems, network devices automatic security updates distribution to network, servers and systems
9. Maintain inventory and asset configuration documentation
10. Assist in procurement process of hardware, firewall, telecom and software vendor evaluation, recommendation, and negotiations
11. Monitoring of Ticket system and provide documented solutions
12. Provide Technical support to remote users, train staff on new systems, day to day admin and monitoring of network use

Database Administrator

1. 2+ years of experience in database administration, design, implementation, and support for mission-critical applications across diverse industries
2. Expertise in managing relational database systems such as Microsoft SQL Server, MySQL, and PostgreSQL
3. Skilled in database performance tuning, indexing, query optimisation, and troubleshooting to ensure optimal performance and minimal downtime
4. Proficient in database backup, restore, and disaster recovery strategies, ensuring data integrity and business continuity
5. Experienced in setting up high availability (HA) and replication solutions using Always On, Log Shipping, and Clustering (SQL Server) or Data Guard (Oracle)
6. Proficient in scripting and automation using T-SQL, PL/SQL, PowerShell, and Python to streamline routine DBA tasks
7. Adept at documenting database environments, procedures, and data flow diagrams to support audits and knowledge sharing
8. Successfully led data migration projects, including on-premises to cloud migrations using AWS RDS, Azure SQL Database, and Google Cloud SQL
9. Strong background in data security, including encryption, access control, and auditing to meet compliance requirements (e.g., GDPR, HIPAA)
10. Developed and maintained ETL pipelines and worked closely with development teams to support application database needs