Table of Contents

# 1 Data Requirements and Migration Specifications

## 1.1 Introduction

This document outlines the comprehensive data requirements and migration specifications for implementing the Signox LogX System. It serves as a guide for institutions to understand what data is needed, how it should be structured, and the process for migrating from existing systems.

## 1.2 Core Data Categories

### 1.2.1 1. Institutional Data

#### 1.2.1.1 Basic Institution Information

* **Institution Name**: Full legal name and display name
* **Institution Type**: University, Polytechnic, Technical College, etc.
* **Registration Numbers**: Government registration, accreditation numbers
* **Contact Information**: Address, phone, email, website
* **Academic Calendar**: Terms, semesters, holidays, examination periods
* **Operating Hours**: Normal operating hours, weekend schedules

#### 1.2.1.2 Organizational Structure

* **Departments**: Academic departments with codes, names, descriptions
* **Programmes**: Academic programmes offered by each department
* **Levels**: Year/semester levels within programmes
* **Specializations**: Specific tracks within programmes

**Sample Data Structure:**

{  
 "institution": {  
 "name": "Nyeri National Polytechnic",  
 "type": "Polytechnic",  
 "registration\_number": "REG/2023/001",  
 "address": "P.O. Box 1234, Nyeri, Kenya",  
 "phone": "+254-xxx-xxxxxx",  
 "email": "info@Signox .ac.ke",  
 "website": "https://www.Signox .ac.ke"  
 },  
 "departments": [  
 {  
 "code": "ENG",  
 "name": "Engineering Department",  
 "head": "user\_id\_123",  
 "description": "Department of Engineering Studies"  
 }  
 ]  
}

### 1.2.2 2. Academic Structure Data

#### 1.2.2.1 Programmes and Courses

* **Programme Information**: Code, name, duration, qualification type
* **Course Catalog**: Course codes, names, descriptions, prerequisites
* **Course Units**: Theory hours, practical hours, credit units
* **Assessment Structure**: CAT percentages, exam percentages
* **Curriculum Mapping**: Course relationships and sequences

#### 1.2.2.2 Academic Periods

* **Academic Years**: Start dates, end dates, status
* **Terms/Semesters**: Duration, examination periods, holidays
* **CAT Weeks**: Continuous Assessment Test periods
* **Special Periods**: Orientation, graduation, maintenance

**Sample Course Data:**

{  
 "course": {  
 "code": "ENG101",  
 "name": "Engineering Mathematics I",  
 "department": "ENG",  
 "programmes": ["MECH", "ELEC", "CIVIL"],  
 "theory\_hours\_per\_week": 4,  
 "practical\_hours\_per\_week": 2,  
 "credit\_units": 6,  
 "prerequisites": [],  
 "assessment": {  
 "cat\_percentage": 30,  
 "exam\_percentage": 70  
 }  
 }  
}

### 1.2.3 3. User Management Data

#### 1.2.3.1 Student/Trainee Data

* **Personal Information**: Names, ID numbers, contacts, addresses
* **Academic Information**: Programme, level, enrollment date, status
* **Financial Status**: Fee balance, payment history (if integrated)
* **Emergency Contacts**: Parent/guardian information
* **Biometric Data**: Facial recognition data, fingerprints (if used)

#### 1.2.3.2 Staff Data

* **Personal Information**: Names, employee IDs, contacts, qualifications
* **Employment Information**: Department, position, hire date, contract type
* **Academic Roles**: Teaching load, specializations, courses taught
* **System Roles**: Admin, HOD, trainer, etc.
* **Biometric Data**: Facial recognition data for authentication

**Sample User Data Structure:**

{  
 "trainee": {  
 "personal": {  
 "first\_name": "John",  
 "last\_name": "Doe",  
 "middle\_name": "Smith",  
 "student\_id": "Signox /2023/001",  
 "national\_id": "12345678",  
 "gender": "male",  
 "date\_of\_birth": "2000-01-15",  
 "phone": "+254-xxx-xxxxxx",  
 "email": "john.doe@student.Signox .ac.ke"  
 },  
 "academic": {  
 "programme": "MECH\_DIPLOMA",  
 "level": "Year 2",  
 "enrollment\_date": "2023-09-01",  
 "expected\_graduation": "2025-12-15",  
 "status": "active"  
 },  
 "emergency\_contact": {  
 "name": "Jane Doe",  
 "relationship": "Mother",  
 "phone": "+254-xxx-xxxxxx"  
 }  
 }  
}

### 1.2.4 4. Timetable and Scheduling Data

#### 1.2.4.1 Room and Facility Information

* **Physical Spaces**: Classrooms, labs, workshops, capacity
* **Equipment**: Available equipment, maintenance schedules
* **Location Data**: GPS coordinates for geolocation features
* **Availability**: Booking schedules, maintenance periods

#### 1.2.4.2 Timetable Data

* **Class Schedules**: Time slots, duration, frequency
* **Room Assignments**: Room bookings, equipment requirements
* **Trainer Assignments**: Trainer allocations, load balancing
* **Special Sessions**: Makeup classes, extra sessions, examinations

**Sample Timetable Data:**

{  
 "timetable\_entry": {  
 "id": "tt\_001",  
 "course": "ENG101",  
 "trainer": "trainer\_123",  
 "room": "room\_a101",  
 "class\_group": "MECH\_Y2\_A",  
 "day\_of\_week": "monday",  
 "start\_time": "08:00",  
 "end\_time": "10:00",  
 "session\_type": "theory",  
 "effective\_date": "2023-09-01",  
 "end\_date": "2023-12-15"  
 }  
}

### 1.2.5 5. Attendance Data

#### 1.2.5.1 Session Records

* **Attendance Sessions**: Date, time, duration, location
* **Session Status**: Planned, started, ended, cancelled
* **Topics Covered**: Curriculum topics taught in each session
* **Trainer Notes**: Session observations, issues, achievements

#### 1.2.5.2 Individual Attendance Records

* **Check-in Data**: Timestamp, location, method (facial, manual, etc.)
* **Attendance Status**: Present, absent, late, excused
* **Verification Data**: Facial recognition confidence, GPS accuracy
* **Feedback**: Student feedback on sessions (optional)

**Sample Attendance Data:**

{  
 "attendance\_session": {  
 "id": "session\_001",  
 "timetable\_entry": "tt\_001",  
 "actual\_start\_time": "2023-09-04T08:05:00Z",  
 "actual\_end\_time": "2023-09-04T10:00:00Z",  
 "location": {  
 "latitude": -0.4167,  
 "longitude": 36.9500  
 },  
 "topics\_covered": ["Linear Equations", "Matrix Operations"],  
 "status": "completed"  
 },  
 "attendance\_record": {  
 "session": "session\_001",  
 "trainee": "trainee\_001",  
 "check\_in\_time": "2023-09-04T08:10:00Z",  
 "status": "late",  
 "verification\_method": "facial\_recognition",  
 "confidence\_score": 0.95  
 }  
}

## 1.3 Data Quality Standards

### 1.3.1 Data Validation Rules

#### 1.3.1.1 User Data Validation

* **Email Addresses**: Valid format, unique within system
* **Phone Numbers**: Valid format with country code
* **ID Numbers**: Unique, proper format validation
* **Names**: Proper capitalization, special character handling
* **Dates**: Valid date ranges, logical sequences

#### 1.3.1.2 Academic Data Validation

* **Course Codes**: Unique, standardized format
* **Programme Codes**: Consistent naming convention
* **Time Slots**: No conflicts, valid time ranges
* **Room Assignments**: Capacity limits, equipment requirements

#### 1.3.1.3 Attendance Data Validation

* **Timestamps**: Chronological order, reasonable durations
* **Location Data**: Valid GPS coordinates, within campus bounds
* **Status Consistency**: Logical status transitions
* **Verification Scores**: Within acceptable confidence ranges

### 1.3.2 Data Integrity Constraints

#### 1.3.2.1 Referential Integrity

* **User References**: All records link to valid users
* **Course References**: Valid course and programme links
* **Room References**: Valid room and facility links
* **Time References**: Valid academic period links

#### 1.3.2.2 Business Logic Constraints

* **Enrollment Limits**: Class size restrictions
* **Schedule Conflicts**: No double bookings
* **Academic Progression**: Proper level sequences
* **Attendance Rules**: Consistent with institutional policies

## 1.4 Data Migration Planning

### 1.4.1 Pre-Migration Assessment

#### 1.4.1.1 Current System Analysis

* **Data Sources**: Identify all current data repositories
* **Data Formats**: Excel files, databases, paper records
* **Data Quality**: Assess completeness, accuracy, consistency
* **Data Volume**: Estimate total records and storage needs
* **Data Dependencies**: Map relationships between data sets

#### 1.4.1.2 Migration Complexity Assessment

* **Simple Migration**: Single source, standard format
* **Moderate Migration**: Multiple sources, some cleanup needed
* **Complex Migration**: Legacy systems, significant transformation
* **Custom Migration**: Unique formats, extensive data processing

### 1.4.2 Data Mapping and Transformation

#### 1.4.2.1 Field Mapping

Create comprehensive mapping between source and target fields:

| Source Field | Target Field | Transformation | Validation |
| --- | --- | --- | --- |
| Student\_ID | trainee\_id | Format: Signox /YYYY/### | Required, Unique |
| Full\_Name | first\_name, last\_name | Split by space | Required |
| Course\_Code | course.code | Uppercase | Required, Valid course |
| Attendance\_Date | session.date | Date format | Valid date range |

#### 1.4.2.2 Data Transformation Rules

* **Name Standardization**: Proper case formatting
* **Date Normalization**: Consistent date formats (ISO 8601)
* **Code Standardization**: Uppercase course/programme codes
* **Address Formatting**: Standardized address structure
* **Phone Formatting**: International format with country code

### 1.4.3 Migration Methodology

#### 1.4.3.1 Phase 1: Data Extraction

* Export data from source systems
* Create backup copies of original data
* Document extraction procedures
* Validate extracted data completeness

#### 1.4.3.2 Phase 2: Data Cleaning

* Remove duplicate records
* Standardize formats and values
* Fill missing required fields
* Resolve data inconsistencies

#### 1.4.3.3 Phase 3: Data Transformation

* Apply mapping rules
* Transform to target format
* Generate required IDs and keys
* Create relationship links

#### 1.4.3.4 Phase 4: Data Validation

* Validate against business rules
* Check referential integrity
* Verify data quality metrics
* Generate validation reports

#### 1.4.3.5 Phase 5: Data Loading

* Load data in dependency order
* Verify successful loading
* Reconcile record counts
* Generate loading reports

## 1.5 Historical Data Requirements

### 1.5.1 Academic Records

* **Student Enrollment History**: Previous programmes, transfers
* **Course Completion Records**: Grades, completion dates
* **Attendance History**: Historical attendance patterns
* **Assessment Records**: Previous CAT and examination results

### 1.5.2 Operational History

* **User Account History**: Previous logins, role changes
* **System Configuration**: Previous settings and rules
* **Timetable History**: Previous schedules and changes
* **Room Usage**: Historical room utilization data

### 1.5.3 Retention Policies

* **Student Records**: 7 years after graduation
* **Attendance Records**: 5 years from session date
* **System Logs**: 2 years for audit purposes
* **Configuration History**: 3 years for troubleshooting

## 1.6 Data Security and Privacy

### 1.6.1 Personal Data Protection

#### 1.6.1.1 Data Classification

* **Public**: Institution information, course catalogs
* **Internal**: Timetables, room assignments
* **Confidential**: Student records, staff information
* **Restricted**: Biometric data, financial information

#### 1.6.1.2 Access Controls

* **Role-Based Access**: Access based on user roles
* **Attribute-Based Access**: Fine-grained permissions
* **Data Masking**: Sensitive data protection
* **Audit Logging**: Complete access tracking

### 1.6.2 Compliance Requirements

#### 1.6.2.1 GDPR Compliance (if applicable)

* **Consent Management**: Track data processing consent
* **Right to Access**: Provide user data on request
* **Right to Rectification**: Allow data corrections
* **Right to Erasure**: Delete data on request
* **Data Portability**: Export user data in standard format

#### 1.6.2.2 Educational Privacy Laws

* **FERPA**: Protect student educational records
* **Local Laws**: Comply with national data protection laws
* **Institution Policies**: Adhere to internal policies
* **Audit Requirements**: Maintain compliance documentation

## 1.7 Data Integration Specifications

### 1.7.1 Real-Time Integration

#### 1.7.1.1 APIs for Live Data Sync

* **User Management**: Real-time user updates
* **Enrollment Changes**: Immediate enrollment updates
* **Timetable Changes**: Live schedule modifications
* **Attendance Data**: Real-time attendance recording

#### 1.7.1.2 Webhook Notifications

* **Enrollment Events**: New enrollments, withdrawals
* **Attendance Events**: Session start/end, low attendance
* **System Events**: User logins, configuration changes
* **Alert Events**: System errors, security incidents

### 1.7.2 Batch Integration

#### 1.7.2.1 Scheduled Data Updates

* **Daily**: Enrollment updates, timetable changes
* **Weekly**: Performance reports, usage statistics
* **Monthly**: Academic progress, attendance summaries
* **Termly**: Grade imports, programme completions

#### 1.7.2.2 File-Based Integration

* **Import Formats**: CSV, Excel, XML, JSON
* **Export Formats**: PDF reports, Excel spreadsheets
* **Validation**: Automated format checking
* **Error Handling**: Comprehensive error reporting

## 1.8 Data Backup and Recovery

### 1.8.1 Backup Strategy

#### 1.8.1.1 Backup Types

* **Full Backup**: Complete system backup (weekly)
* **Incremental Backup**: Changed data only (daily)
* **Transaction Log Backup**: Continuous transaction logging
* **Configuration Backup**: System settings (before changes)

#### 1.8.1.2 Backup Storage

* **On-Site**: Local backup for quick recovery
* **Off-Site**: Remote backup for disaster recovery
* **Cloud**: Cloud storage for additional security
* **Encryption**: All backups encrypted at rest

### 1.8.2 Recovery Procedures

#### 1.8.2.1 Recovery Time Objectives (RTO)

* **Critical Systems**: 2 hours maximum downtime
* **Standard Systems**: 8 hours maximum downtime
* **Non-Critical**: 24 hours maximum downtime

#### 1.8.2.2 Recovery Point Objectives (RPO)

* **Critical Data**: Maximum 15 minutes data loss
* **Standard Data**: Maximum 1 hour data loss
* **Historical Data**: Maximum 24 hours data loss

## 1.9 Data Reporting and Analytics

### 1.9.1 Standard Reports

#### 1.9.1.1 Administrative Reports

* **Enrollment Reports**: Student counts by programme
* **Attendance Reports**: Attendance rates and trends
* **Performance Reports**: Academic achievement metrics
* **Utilization Reports**: Resource usage statistics

#### 1.9.1.2 Academic Reports

* **Student Progress**: Individual and cohort progress
* **Course Analytics**: Course performance metrics
* **Trainer Reports**: Teaching load and performance
* **Programme Effectiveness**: Programme outcome analysis

### 1.9.2 Custom Analytics

#### 1.9.2.1 Data Visualization

* **Dashboards**: Real-time operational dashboards
* **Charts**: Trend analysis and comparisons
* **Maps**: Geographic distribution of students
* **Calendars**: Schedule and event visualization

#### 1.9.2.2 Predictive Analytics

* **Attendance Prediction**: Identify at-risk students
* **Performance Prediction**: Academic success indicators
* **Resource Planning**: Capacity planning analytics
* **Dropout Prevention**: Early warning systems

## 1.10 Implementation Checklist

### 1.10.1 Pre-Implementation

* Complete data inventory and assessment
* Define data quality standards
* Create data mapping documentation
* Develop migration scripts and procedures
* Set up test environment for validation

### 1.10.2 During Implementation

* Execute data migration in phases
* Validate data quality at each step
* Test system functionality with migrated data
* Train staff on data management procedures
* Document any data issues and resolutions

### 1.10.3 Post-Implementation

* Monitor data quality metrics
* Establish ongoing data maintenance procedures
* Schedule regular data backup testing
* Review and update data retention policies
* Plan for future data integration needs

## 1.11 Data Requirements by Institution Size

### 1.11.1 Small Institution (< 1,000 Users)

* **Users**: 500-1,000 user records
* **Courses**: 50-200 course records
* **Sessions**: 100-500 sessions per term
* **Attendance**: 10,000-50,000 attendance records per term
* **Storage**: 50-200 GB total data storage

### 1.11.2 Medium Institution (1,000-5,000 Users)

* **Users**: 1,000-5,000 user records
* **Courses**: 200-1,000 course records
* **Sessions**: 500-2,500 sessions per term
* **Attendance**: 50,000-250,000 attendance records per term
* **Storage**: 200-1,000 GB total data storage

### 1.11.3 Large Institution (5,000-15,000 Users)

* **Users**: 5,000-15,000 user records
* **Courses**: 1,000-3,000 course records
* **Sessions**: 2,500-7,500 sessions per term
* **Attendance**: 250,000-750,000 attendance records per term
* **Storage**: 1-5 TB total data storage

### 1.11.4 Enterprise Institution (> 15,000 Users)

* **Users**: 15,000+ user records
* **Courses**: 3,000+ course records
* **Sessions**: 7,500+ sessions per term
* **Attendance**: 750,000+ attendance records per term
* **Storage**: 5+ TB total data storage

## 1.12 Conclusion

Proper data planning and management is crucial for successful implementation of the Signox LogX System. This document provides the framework for understanding data requirements, ensuring data quality, and implementing robust data management practices.

### 1.12.1 Key Success Factors

1. **Comprehensive Planning**: Thorough data assessment and planning
2. **Quality Standards**: Consistent data quality standards
3. **Migration Testing**: Extensive testing of migration procedures
4. **User Training**: Proper training on data management
5. **Ongoing Monitoring**: Continuous data quality monitoring

**Document Version:** 1.0  
**Last Updated:** June 2025  
**Document Owner:** Signox LogX Data Management Team