Data Architecture

Data Architecture is intended to provide a mechanism for the various stakeholders at various levels of Government to identify, discover, describe, manage, protect, and share the data it has and reuse information consistently within and across ministry or division of directorate or for the entire Government of Bangladesh.
Data Architecture provides standards for accessing data for online analytical processing (OLAP), including executive information systems (EIS) and decision support systems (DSS).

Whole of Government: The ICT for Government of Bangladesh

To realize the business strategy as defined, and in the long term realize the Vision for Digital Bangladesh by 2021 – one of the key drivers for ICT is Data domain – the most critical and complex. The data architecture needs to be defined in a manner that addresses all the challenges of the Government and makes it flexible to adopt in a rapidly changing business environment

Data Architecture Principles

The principles listed below are key guidelines for the design data for systems across the Government of Bangladesh. It is imperative that the principles would be followed rigorously for all data services – acquisition, storage, retention, archival and consumption that would lead to a secure, collaborative and adaptive ICT for the realization of Digital Bangladesh Vision 2021

These principles are listed to achieve the following objectives:

- **Enable architecture review**: Any new system development would require architecture review, data architecture principles would provide the necessary review parameters as far as database design is concerned
- **Provide a guidance mechanism**: to database design team or data architect, on what are the criteria that defines the best the database design
- **Discover gaps in data security, plan for secured and adaptive data architecture**: Data architecture principles compliance depicts the loopholes in data security, data protection and overall data design

Name	DP1: Data access through defined business rules
Description	Data access to be based on business rules only, all data access to be made following defined and approved CRUD for all roles accessing the system
Scope	Government hosted systems, portals, reports and mobile apps
Implementation Steps	1. For new systems implementation – define CRUD and approve from data architect 2. Architecture review checklist to include CRUD review 3. Prohibit data access from ad-hoc query through DBA defined rules 4. Review and enhance existing system's CRUD
Benefit	• Security breaches frequently occur at data access using ad-hoc queries; use of CRUD would ensure security • Data governance becomes organized and eases data management • Data rights becomes streamlines

Name	DP2: Data is an asset, shared and governed
Description	• Data is an asset - Data to be cleaned, synchronized and preserved using central data management tools. Data to be archived as per data archival policy. • Data is shared – Data to follow data sharing rules as per data classification • Data is governed – Data stewards to maintain data throughout its life cycle
Scope	Data within the purview of Government of Bangladesh
Implementation Steps	1. Data Cleaning and Synchronization 2. Data to be cleaned and synchronized using data clean-up tools at extraction level in Master Data Management System 3. Implementation of Master Data Management System along with Data clean up tools 4. Data Archival Policy 5. Draft and publish data archival policy 6. Implementation of data archival system 7. Implementation of ETL tool 8. Architecture review of data archival for all systems 9. Data Classification 10. Finalize Data classification rules 11. Publish and enforce data classification adoption across ministries 12. Draft and publish centralized data sharing rules 13. Data Governance 14. Please follow below section for data governance details
Benefit	• Data is maintained as an asset to the Government of Bangladesh – secured, governed and shared • Data is preserved and is easily retrievable • Data is classified ensuring access to the right person

Name	DP3: Data design aligned to National Meta Data Standards
Description	Data type, length and uniqueness for key and common data entities are aligned to published National Meta Data Standards
Scope	• Core and Common data entity
Implementation Steps	1. Draft and Publish Meta Data Standards 2. Architecture review of Meta Data for new system implementation
Benefit	• Sharing of data becomes easy as there would not be any compatibility issues • Eases system development, API development effort

Name	DP4: Data synchronized with federated master data management
Description	Core data entity must have relationships as per data standards, and established mechanism to incorporate in the National Master Data Management System

Scope	• Core and Common data entity
Implementation Steps	1. Implement MDM platform 2. Establish mechanism for data extraction and load to master data management system 3. Review data cardinality
Benefit	• Ease data integration for effective consumption in reports and analytical tools • Checks Data cardinality as per national standard

Name	DP5: Core Data Access
Description	Core data entity must have established identifier to access, store and preserve.
Scope	• Core data entity
Implementation Steps	Establish Core Data identifier adoption across all levels of data transaction – access, store and preserve. Following are the identifier for core data: • Citizen – National ID (NID) • Business – Business Identification Number (BIN) • Employee – Government ID • Things – Asset ID • GIS – Geo Spatial ID
Benefit	• Ease data integration for effective consumption in reports and analytical tools

Name	DP6: Data security
Description	Data to be made available to citizens, business or other entities who require the information as part of their role. For secured data – proper encryption and security measures for data protection
Scope	• All Data
Implementation Steps	Establish data governance as illustrated in the report Establish Open Data Catalogue 1. On classified public data draft open data catalogue 2. Extract data in open data repository 3. Publish open data API, catalogue for public consumption
Benefit	• Secured and accessible data for all

Core Data Entity

The data entities that seldom changes or is slow to change, Core Data of all such entities and its associated properties may be created as a Single Source of Truth, to be used by all the ministries providing services relating to the Government of Bangladesh. A Unique ID may be created for each such core data and every property, on the lines of data identifier

Following are classified as the core data entities for Government of Bangladesh:

- ✓ **Citizen**

  Data related to citizens are classified as a core data entity; different citizen's different profiles are currently stored in different ministry owned systems. Connecting this data would enable a 360-degree view of the citizens. Please follow **Appendix** for Citizen Data Profile across different ministries

- ✓ **Employee**

  Another Core data entity that is critical to the Government of Bangladesh, employee profile is mostly maintained centrally across ministries but depending upon the type of employee the profile management may differ. A central management of employee data is critical to the success outcomes

- ✓ **Things**

  Another crucial data entity that is all ministry owned infrastructure such as Land, Offices, Machinery, etc. A centralized management of the assets from procurement to disposal would benefit greatly in terms of lower cost to Government

- ✓ **Business**

  Business entity can be maintained as a single source of truth, avoiding duplication across ministries to derive the best outcome both for the business and for Government

- ✓ **GIS**

  A geo spatial data when maintained as a single source of truth would enable many benefits to decision-making, logistics and cost

Structured, Un-structured and Semi-Structured Data Repository

A Digital Data Resource is a digital container of information. A Digital Data Resource may correspond to three types of data: "Structured Data Resource", "Semi-Structured Data Resource", and "Unstructured Data Resource".

1. **Structured Data Resource**: Structured Data Resource is a type of Digital Data Resource containing only structured data. A Data Schema is used to define/describe a Structured Data Resource.
2. **Semi-Structured Data Resource**: A Semi-Structured Data Resource is a Digital Data Resource containing semi-structured data. A Semi-Structured Data Resource contains partly structured and partly unstructured data.
3. **Un-structured Data Resource:** An Unstructured Data Resource is a type of Digital Data Resource that contains only unstructured data. Unstructured data is collection of data values that are likely to be processed only by specialized application programs.

Content Repository

Content repository would comprise of easy to retrieve, indexed documents, media files, web graphics and templates

Meta Data Repository

Meta data is data about data defines and describes data or information. It is used to manage data, information and knowledge. Metadata is the structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource

Once a National level data standard is established, a Meta data repository would be needed to manage and maintain the critical data of data

Data Models

A data model ensures that data is defined accurately so it is used in the manner intended by both end users and remote applications.

There are three types of data models –

- ✓ **Conceptual Model**

A conceptual data model identifies the highest-level relationships between the different entities. Features of conceptual data model include the important entities and the relationships among them.

✓ **Logical Model**
A logical data model describes the data in as much detail as possible, without regard to how they will be physical implemented in the database

✓ **Physical Model**
A physical database model shows all table structures, including column name, column data type, column constraints, primary key, foreign key, and relationships between tables.

Data modeling tools can evaluate an existing database structure and reverse engineer a data model. The reverse engineered data model can be used to capture valuable information about the existing database.

Data Governance
Data governance encompasses the strategies and technologies used to make sure Government of Bangladesh's data stays in compliance with regulations and policies. It is proposed to be a collection of processes, roles, policies, standards, and metrics to ensure the effective and efficient use of information in enabling Government of Bangladesh as a whole to achieve its goals.

The Roles – Data Steward
Data governance strategy would push a cultural movement away from data ownership to data stewardship. Since data is an asset of value to the entire enterprise, data stewards are made accountable for properly managing the data. Data governance would be exercised at the enterprise level with federated governance to individual ministries and directorates. It would be proactively exercised when a new process, application, repository or interface is introduced. Existing data is likely to be impacted.

The Process – Data Steward
Under each stage of data life cycle, data steward would have a critical role to play

Data Life Cycle Stage	Create
Data Steward Role	▪ Define the attributes of the identified core data & the relationship ▪ Identify the systems for data acquisition of core data ▪ Finalize Data classification and security requirements
Data Life Cycle Stage	Store
Data Steward Role	▪ Identify database repository and list the data that is stored ▪ Define and manage security requirement for data Storage ▪ Draft data storage & backup policy need to be enforced ▪ DR setup

Data Life Cycle Stage	Use
Data Steward Role	▪ Review and manage data models ▪ Review CRUD for data ▪ Based on classification of data – transfer to open data repository

Data Life Cycle Stage	Share

Data Role Steward	▪ Define data integration standards ▪ Approve and finalize CRUD ▪ Approve API documentations and requests ▪ Define security requirements for confidential data

Data Life Cycle Stage	Archive & Destroy
Data Role Steward	▪ Draft and create archival policy and dispose policy ▪ Implement archival – ETL tools/Integration platform ▪ Monitor archival process and disposal process

Master Data Management

Master Data refers to those commonly required data, which are agreed upon and shared across the Government. It may be a reference data such as a list of values to be used for a data element such as sectors in Government. Gartner defines Master Data as "Master data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts"

It is proposed that Government of Bangladesh follow a Virtual Master Data Management Architecture with the following capabilities:

Master Data Virtualization Service: The service would enable same data acquisition from multiple data source, yet maintaining a single master following Federated Architecture. As an example, each ministry follows its own codes and identifier for similar entities such as Districts, the virtualization service would enable a mapping with all those different entities for the same district, for example, Code 4 in Finance Ministry might represent Khulna while Khulna is represented as Code 6 in Social Welfare ministry. The virtualization service would enable the mapping of Khulna with all the ministries

MDM Repository: The repository to store and preserve the master data to enable single source of truth view

Data Synchronization: The data synchronization and clean up would enable cleaning of data entities from say free form text entry, this would also enable meta data standard compliance

Data Conflict Resolution: The tool/capability would display the data conflict among various sources of same master data to help resolve conflict and preserve the right data

Data Warehouse

A data warehouse is a collection of data designed to support decision-making and analytical processing. Data warehouses contain a wide variety of data, usually from multiple data sources, presenting a comprehensive view of a particular business environment. Due to the nature of the data stored in a data warehouse, the size of the data warehouse is usually very large, so it requires special design and planning.

A data mart is a subset of a data warehouse. Where data warehouses are designed to support many requirements for multiple business needs, data marts are designed to support specific requirements for specific decision support

applications (i.e., particular business needs). Although a data mart is a subset of a data warehouse, it is not necessarily smaller than a data warehouse. Specific decision support needs may still require large amounts of data. Data marts are typically considered a solution for distributed users who want exclusive control of the information required for their business need.

Data warehouse efforts should begin with a specific requirement for a specific decision support application, similar to the practices of a data mart design. For scalability, the tools and databases used should be designed to support a very large data warehouse, instead of using data mart specific products.

Future State Data Architecture Model

Data Sources
- User Input
- Flat Files
- Spreadsheets
- API Requests

Core Data Entity

Core Data Entities (databases): Citizen | Employee | Things | GIS | Business | Schemes & Policies

Election Commission
National ID Register

Education	BCC	Labor Ministry	Health
Student Profile	Citizen e-Profile	Labor Profile	Patient Profile

Land	Home	Social Welfare	Agriculture
Land Owner's Profile	Criminal's Profile	Beneficiary's Profile	Farmer's Profile

Foreign Affairs/ Expatriates	Textile, Fisheries…	Railways, Transport & Aviation	Information
NRB/Foreigner's Profile	MSME Worker's Profile	Passenger's Profile	ICT User's Profile

Planning
Asset Procurement ID

All Ministry	Road, Railways
Asset Register	Roads, Railway Tracks Information

Education, Social Welfare	BCC
Government ID	ICT Infrastructure

Health	Public Administration
Hospital Infrastructure Medicine Stock Register	Office Building

Commerce
Business Identification ID (BIN)

Commerce	Planning
Business Details, TAX, VAT Details	Vendor Details

Agriculture, Livestock & Textile	Commerce
MSME Details	Import, Export – Customs duty Details

Health	Industries
Medicine Supplier, Pharmaceuticals	Business Details

Data Tools
- Data Governance Tools
- ETL Tools
- Data Cleanup Tools
- Data Encryption Tools

Public Administration
Government ID

Education	All Ministry	Public Administration	All Ministry
Teacher's, Staff's Profile	Employee Profile	BCS Officer's Profile	Other Officer's Profile

Home	BCC and Other Ministry	Law & Justice	All Ministries
Security Officer's Profile	IT Consultant's Profile	Judge, Advocate, Judiciary Staff Profile	Field Worker's Profile

Foreign Affairs/ Expatriates
Foreign Worker's Profile

BCC
Geo Dash ID

All Ministry	Sanitation & Water Supply
Office Location	Sewerage, Water Pipes

Power	Railways, Transport & Aviation
Transmission Lines	Roads, Highways, Railway Track GIS

Health	Industries
Hospitals, Clinic Locator	Land Details (Industrial Estates)

Cabinet
Policy ID, Act Number, Rule Number, Law Code

All Ministry
Policy, Acts, Rules and SOP Details

Cabinet
Relevant Status

Law & Justice
Law Details Court Cases

Repositories
- Structured Data Repository System
- Content Repository System
- Open Data Repository
- Archive Database

- Meta Data Repository
- Data Models repository
- Data Modeling Tools & Data Modals
- Open Data Catalogue