

Terms of Reference (ToR)

Data Preparation (census, survey and other data sets) and Data Migration

Bangladesh Bureau of Statistics, the National Statistical Office (NSO) of the country, is in the process of enhancing its data infrastructure to ensure efficient and reliable data management by establishing an integrated statistical data warehouse. A critical component of this initiative is to prepare the existing data sets (census, survey, and other data sets) and eventually the Extract, Transform, Load (ETL) process within the data warehouse. The specific Terms of Reference (ToR) of the assignment are the followings:

A. Data Preparation (Census and survey and other data sets):

1. Discovery and Assessment of microdata

- **Identify Data Sources:** Prepare a list of the data sets and locate all relevant old census/survey microdata files, systems, and storage locations with the help of the officials of BBS.
- **Understand Data Context:** Review documentation, metadata, and historical records to understand the data structure and meaning
- **Assess Data Accessibility:** Determine how to access the data (e.g., file formats, database systems, physical media).
- **Evaluate Data Quality:** Conduct an initial assessment for completeness, consistency, and accuracy.

2. Metadata Preparation

- **Collect Metadata:** Gather information about the data fields, formats, units, and relationships.
- **Define Data Types:** Document field types (e.g., numeric, categorical, text) and constraints.
- **Document Legacy Standards:** Develop metadata for each data set by maintaining old coding schemes, classifications, and formats used in the data.

সৈয়দ মোহাম্মদ আবু নাসের
সহকারী পরিসংখ্যান কর্মকর্তা
১৫/১০/২০

মুহাম্মদ আরিফুল ইসলাম
উপপরিচালক
১৫/১০/১৫

মেটা ফার্মেক সোহেল
লিলিয়ার প্রেসার্যার
বাংলাদেশ পরিসংখ্যান ব্যৱস্থা
পরিসংখ্যান ও তথ্য ব্যবহারণ বিভাগ
প্রক্রিকজ্ঞান অক্ষয়ালয়

ফজলুর রহমান
সিস্টেম এনালিষ্ট
বাংলাদেশ পরিসংখ্যান ব্যৱস্থা
গণপ্রজাতন্ত্রী বাংলাদেশ সরকার
১৫/১০/২০

3. Data Extraction

- **Select Extraction Tools:** Identify tools or scripts needed to extract data from legacy systems or file formats.
- **Prepare Access Credentials:** Ensure proper permissions to access restricted or archived data.
- **Verify Data Availability:** Confirm that all required datasets are available and intact.

4. Data Cleaning

- **Identify Common Issues:** Anticipate typical problems like missing values, duplicates, or outdated codes.
- **Define Cleaning Rules:** Establish guidelines for handling missing data, inconsistencies, and invalid entries.
- **Compile Reference Data:** Gather external resources (e.g., codebooks, geographic mappings) to aid in cleaning.

5. Data Standardization

- **Unify File Formats:** Convert files to a standard format (e.g., CSV, Excel, or database-compatible).
- **Standardize Encoding:** Ensure all files use consistent character encoding (e.g., UTF-8). Normalize Field Names: Rename fields to adhere to a consistent naming convention.
- **Update Legacy Codes:** Identify and document mappings for outdated codes to modern equivalents.

6. Data Validation

- **Define Validation Rules:** Specify checks for data consistency, such as valid ranges or logical relationships (if needed).

7. Format of database to submit

- **Data format:** All data sets should be submitted to the Project authority in the following formats viz. Stata, SPSS, CSV, and DBF including metadata for each of the data sets.

সৈয়দ মোহাম্মদ আবু নাসের
সহকারী পরিসংখ্যান কর্মকর্তা

মুহাম্মদ আরিফুল ইসলাম
উপপরিচালক

মোট ফার্মেল স্লোভেন
সিনিয়র প্রেসিডেন্ট
বাংলাদেশ পরিসংখ্যান ব্যৱসা
পরিসংখ্যান ও তথ্য বিব্রাচনা বিভাগ
প্রকল্পসমূহ প্রতিপাদ্য

বাংলাদেশ পরিসংখ্যান ব্যৱসা
গণপ্রজাতন্ত্রী বাংলাদেশ সরকার

B. Data Migration and ETL Process:

1. ETL Design and Development and Load

- Design an ETL architecture tailored to the BBS's needs.
- Develop ETL workflows to extract data from various structured and unstructured sources.
- Implement data transformation rules to ensure data quality, consistency, and compliance.
- Load transformed data into the data warehouse with optimal performance.

2. Data Governance and Quality Assurance

- Implement data validation and error-handling mechanisms.
- Ensure compliance with BBS's data governance policies and international standards.

3. Testing and Deployment

- Conduct comprehensive testing, including unit, integration, and performance testing.
- Deploy the ETL processes in a production environment.

4. Documentation

- Provide detailed technical and user documentation.

C. Deliverables

The agency is expected to deliver the following:

- Inception report.
- Data validation and quality assurance reports.
- Report on Data Preparation.
- Inception Report (including the status of existing datasets).
- Standardized and cleaned datasets ready for migration (All data sets should be submitted to the Project authority in the following formats viz. Stata, SPSS, CSV, and DBF).
- Fully migrated datasets within the integrated data warehouse.
- Metadata documentation for all migrated datasets.
- Final Report.

সৈয়দ মোহাম্মদ আবু নাসের
সহকারী পরিসংখ্যান কর্মকর্তা
০৩/০১/১২

মুহাম্মদ আবিদুল ইসলাম
উপপরিচালক
০৩/০১/১২

মোট কর্মসূচক
সিলিগুর প্রেসার্যার
বাংলাদেশ পরিসংখ্যান বুরো
পরিসংখ্যান ও তথ্য প্রকল্প বিভাগ
পরিকল্পনা মন্ত্রণালয়

ফোন
০৩/০১/১২

যতন কুমার সাহা
সিলেক্ট এনালিষ্ট
বাংলাদেশ পরিসংখ্যান বুরো
গণপ্রজাতন্ত্রী বাংলাদেশ সরকার

D. Duration

Three years (or as per TAPP)

E. Quantity of data sets

A total of 189 data Sets (Tentative - including Census of Manufacturing Industries as Census data set)

F. Confidentiality

The agency must ensure the confidentiality and security of all data and adhere to applicable data protection laws.

স্বত্ত্বান্বিত ০৭/০১/২৮
সৈয়দ মোহাম্মদ আবু নাসের
বাংলাদেশ পরিসংখ্যান কর্মকর্তা
সহকারী পরিচালক

মুহাম্মদ আরিফুল ইসলাম
উপরিচালক
০৭/০১/২৮

মোঃ হারুন কে সোহেল
নিয়মীয়ার প্রেসার্বার
বাংলাদেশ পরিসংখ্যান ব্যৱৰ্তো
পরিসংখ্যান ও তথ্য নথি হাপনা বিভাগ
পরিকল্পনা মন্ত্রণালয়

ব্যতন কুমার সাহা
সিটেম এনালিষ
বাংলাদেশ পরিসংখ্যান ব্যৱৰ্তো
গণপ্রজাতন্ত্রী বাংলাদেশ সরকার