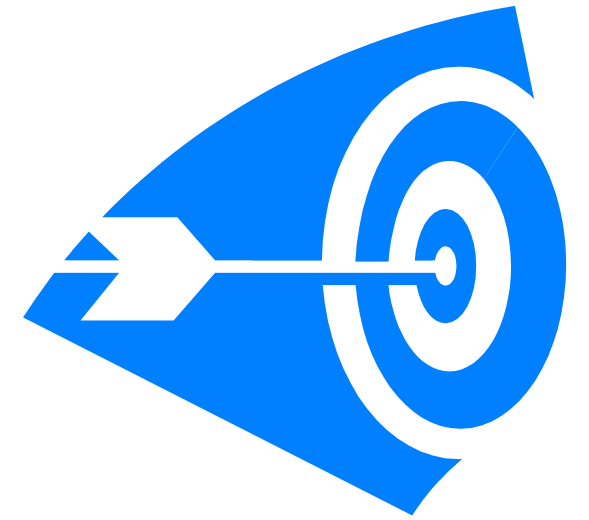


Course: e-Governance Project Lifecycle

Day 3: Session 2

Data Digitization & Quality Management

Agenda



- Need for Data Digitization in e-Governance Projects
- Data digitization methods
- Approach for data digitization and data quality management

Data in e-Governance Implementations

- IT changes constantly; business processes; applications; and the underlying technologies change, and are eventually replaced
- While all these changes are happening, an organization's data is the one aspect of business that must accurately reflect the past, present and future
- Data is a critical component of successful business operations and IT implementations
- Business demands a careful approach for managing data quality and IT change, techniques and approaches for a successful data management

Significance of 'Data' in e-Governance – Case Study

IT efforts of a Tax Department (Illustrative)

- 'abc' Tax Department initiated IT project and has implemented IT System for automation of tax functions including:
 - Registration of an entity
 - Issuing copies of registration certificates
 - Assessment of tax
 - Generation of Demand Notice
 - Collection of tax
- Department has digitized the manual data and migrated the data to the new system
- Department used the new system along with the data for online transactions including:
 - Serving demand notice through e-mail/online demand notices
 - Online payment of tax
 - Online facility for issuing copy of registration certificate

Significance of 'Data' in e-Governance – Case Study (contd..)

Planned Benefits from IT initiative:

- Reduction in administrative burden
- Improvement in convenience
- Improvement in revenue collections and monitoring
- Reduction in time for delivering services
- Enhancement of department image and perception in public.....

Significance of 'Data' in e-Governance – Case Study (contd..)

What has been achieved?

- Department has experience following unique challenges post IT system adoption
 - Drop in revenues/tax collections
 - Significant number of complaints from entities for:
 - corrections in the details in the certificate copies issued by the department
 - Errors in the demand notices served online or through e-mail
 - Online facility for issuing copy of registration certificate

Outcome from IT initiative:

- Revenue loss
- Increased administrative burden in addressing the complaints from citizens
- Reputation loss and negative publicity
- Reduced trust in 'IT' or e-Governance...

Significance of 'Data' in e-Governance – Case Study (contd..)

Problem Analysis

Department has analyzed the situation and identified following issues:

- Errors in the digitized data
 - Errors in the entity master records
 - Errors in the tax assessment records
 - In complete data for the records migrated into the IT system.....
- Data quality controls not applied during digitization and post digitization
- Digitized data not verified by the department for accuracy and completeness

Department initiated corrective measures for addressing data gaps/errors, but

DAMAGE IS ALREADY DONE!!

Significance of 'Data' in e-Governance – Case Study (contd..)

Having Optimized business processes, State of the Art IT systems and Skilled/trained employees will not help if underlying 'data' used for service delivery is incorrect.....

Understanding Data Quality

The existence of the right data in the right format at the right place and time to meet the needs of business processes

Dimensions of Data Quality

Accuracy - A measure of the correctness of the digitized data in comparison with the source data

Uniqueness - A measure of unwanted duplication existing within or across systems for a particular field, record, or data set based on the type data (master data/records)

Consistency and Synchronization - A measure of the equivalence of information stored or used in various data stores, applications, and systems, and the processes for making data equivalent

Timeliness and Availability - A measure of the degree to which data are current and available for use as specified and in the time frame in which they are expected

Completeness – Completeness of the data

Managing Data in e-Governance Implementations

- Understand what data is needed in digital format for delivering online services
- Define right approach for data digitization for
 - Ensuring security of master records
 - For ensuring quality and completeness of digitized records
 - For ensuring accountability of quality and completeness of digitized records
- Select right methods and tools for data digitization and data validation
- Have experienced and trusted service provider for data digitization and migration
- Ensure 100% validation of records for quality and completeness of digitized records
- Adopt fail proof approach for migration of digital data to the IT system
- Validate migrated data for quality and completeness
- Sustain data quality efforts through IT systems operations and management

Data in e-Governance Implementations

- Usage of IT Systems will require following data:
 - Master Data
 - Transaction data

Master Data

- Data essential for performing various business functions and transactions using the IT system
- Critical requirement for processing business transactions
- e.g.
 - Master data for HR Management System – Employee Records, Salary structure, leave records..
 - Master data for finance applications – Budget heads, budgeted amounts, DDOs data..
 - Master data for procurement application – Procurement items, procurement officers, vendor details..
- Supports transactional processes and operations and is often used/shared by multiple functions/system transactions

Transactional Data

- Data describing an event/transaction processed through the system
- Represents business transactions of the organization
- Generated through functions/transactions performed through the system using the master data
 - Transaction data from HR Management System – Payroll records, leave approvals, transfers.....
 - Master data for finance applications – Receipts, payments, deposits.....
 - Master data for procurement application – purchase of items, payments to vendors.....

Transactional Data

- Data describing an event/transaction processed through the system
- Represents business transactions of the organization
- Generated through functions/transactions performed through the system using the master data
 - Transaction data from HR Management System – Payroll records, leave approvals, transfers.....
 - Master data for finance applications – Receipts, payments, deposits.....
 - Master data for procurement application – purchase of items, payments to vendors.....

Need for Data Digitization in e-Governance

Data Digitization may be needed for :

- Issuing encumbrance certificate for properties/land parcels
- Issuing copies of certificates/ licenses (birth/death/company or dealer registration/building permissions....)
- Preserving the data from old manual records (land records/registration documents)

Data digitization may not be needed for:

- Online submission/filing of returns – may not need digitization of assesees master records
- Online payment of taxes – may not need demand notice/tax assessment data

Need for Data Digitization in e-Governance

Data Digitization may be needed for :

- Issuing encumbrance certificate for properties/land parcels
- Issuing copies of certificates/ licenses (birth/death/company or dealer registration/building permissions....)
- Preserving the data from old manual records (land records/registration documents)

Data digitization may not be needed for:

- Online submission/filing of returns – may not need digitization of assesse master records

Data digitization and data quality issues may lead to significant delays in IT adoption and departments can focus on non-data intensive services first during service prioritization

Data Conversion Methods

For Records

- Digitization/ Data entry
- OCR/ ICR/ OMR (Expand)
- Geo reference

For Documents

- Scanning
- Photographing
- Photocopy
- Microfilming

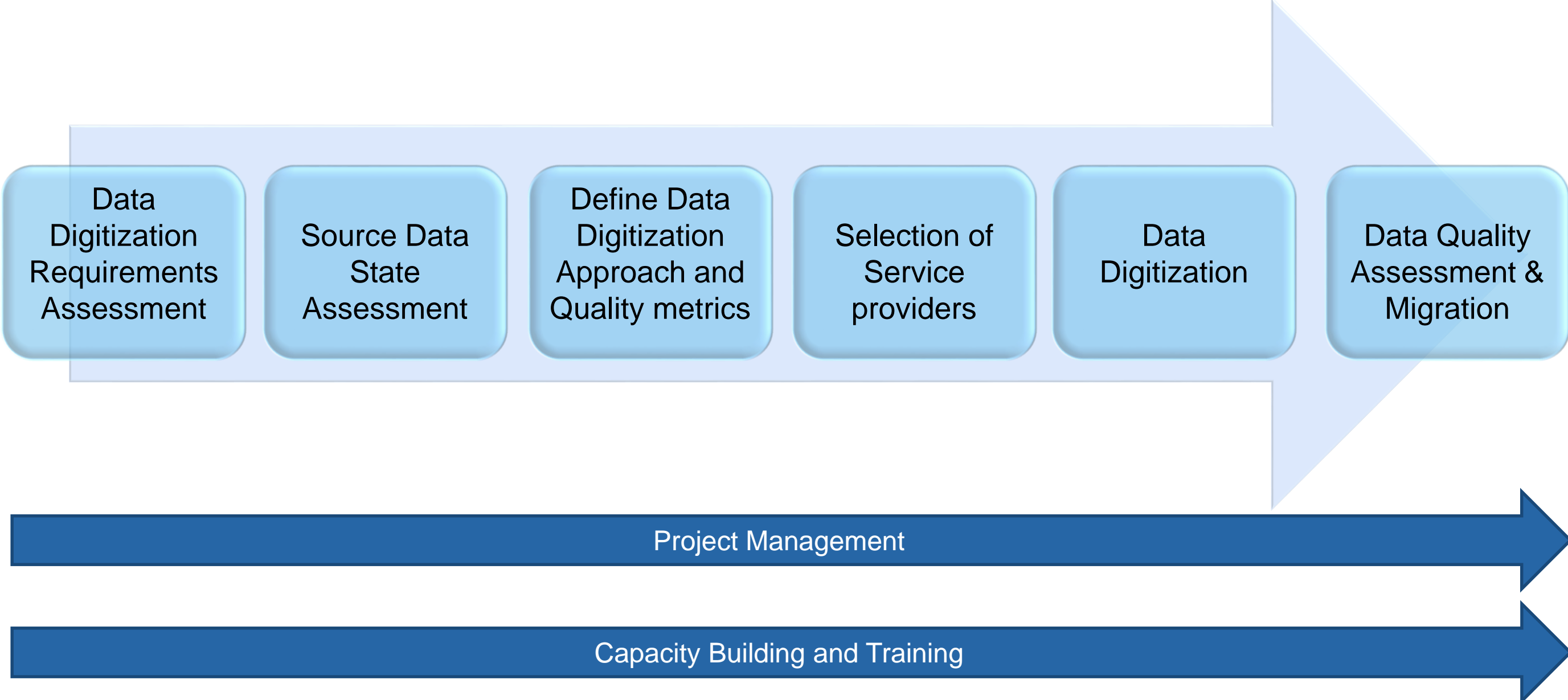
Challenges in Data Digitization

- Incomplete data
- Inaccurate data
- Data manipulation
- Invalid data
- Poor quality of scanning
- Size of scanned documents
- Incomplete scanning of document (missing pages)
- No standard/ common formats.....

Implications of Errors in Data Digitization

- Revenue loss
- Legal challenges
- Reputation loss
- Loss of important data
- Operational inefficiencies
- Conflicting reports, erroneous decisions that are made with outdated, inconsistent, and invalid data.

Data Digitization & Migration Approach



Data Digitization Requirements Assessment

- Identify the list of services prioritised for implementation under e-Governance
- Review the 'To-Be' business processes defined for identified services
- Review the 'Business' and 'Data Architecture' defined for the proposed system implementation
- Assess the data (master and transaction) needed for processing various activities/steps in the underlying processes related to service delivery
- Validate the feasibility of online service delivery without the digitized source data for the service, its risks and legal implications
- Finalise the data digitization requirements for proposed system and online services

Data Digitization Requirements Assessment (contd..)

Services identified for e-Governance Project	Data Requirements
Issuance of birth certificate	Elements in the birth records, which are printed in birth certificate
Online payment of property tax*	Tax assessment/ Demand notice data
Water tax collection*	Tax assessment/ Demand notice data
Issuance of death certificate	Elements in the death records, which are printed in death certificate
Issuance of copy of building permission	building permission certificate
Issuing encumbrance certificate	Extent and ownership details of land

* We may not need any data to be digitized for enabling online payment of taxes by citizens

Source Data State Assessment

- Digitizing incomplete/improper manual records will lead to inefficient online services
- Assess the quality and completeness of source data/records/documents
- Identify the gaps and approach for bridging the identified gaps
 - Surveying the filed for information collection – consumes long time and delays data digitization thus delaying system go-live (typically seen in land records/properties data)
 - Partial digitization with online data corrections/updation by citizens – requires stringent measures for validating data corrections made by citizens
 - Bridge the gaps based on data sources available in other departments/related agencies

Source Data State Assessment (contd..)

- Data Categories:
 - Textual records (employee data, assesse (tax) data, firm/company details..)
 - Documents (registration certificates, sale deed, licenses, photographs...)
 - Geo data (satellite images, ortho-photographs..)
- Data Formats:
 - Manual records (papers, files, books...)
 - Digital records (excel sheets, databases, digital images, satellite images...)

Define Data Digitization Approach and Quality metrics

- Selection of data digitization approach based on Data categories and data formats

Data Category	Data Format	Approach
Textual Records (employee data, assesse (tax) data, firm/company details..)	Manual	Data entry using data digitization application*/ business application Or OCR/ICR**
Textual Records (employee data, assesse (tax) data, firm/company details..)	Digital format	Data validation, digitize the gaps cleansing and migration
Documents (registration certificates, sale deed, licenses, photographs...)	Manual	Scanning
Documents (registration certificates, sale deed, licenses, photographs...)	Digital format	Validation and migration

Illustrative

Define Data Digitization Approach and Quality metrics (contd..)

- Define Data Quality expectations/requirements for each data element
- Data quality metrics/attributes vary from textual records and documents

Illustrative data quality requirements for a property tax record

Data element	Data Type	Mandatory (Yes/No)	Quality Definition/Criteria
Assessment_No	Text	Yes	<ol style="list-style-type: none"> 1. Should not be left blank 2. Should contain numerical values only 3. Should be a unique value for the properties in the location 4. Length of data shall be = 8 characters 5. Should not contain missing sequence numbers...
Reg_Date	Date/Time	Yes	<ol style="list-style-type: none"> 1. Should not be left blank 2. Should adhere to date and time format (mm/dd/yyyy) 3. Should not be greater than the current date
Owner_Name	Text	Yes	<ol style="list-style-type: none"> 1. Owner name should not be left blank 2. No salutations to be attached to names 3. Owner name should not contain numerical values or special characters
Property Tax	Numerical	Yes <i>Illustrative</i>	<ol style="list-style-type: none"> 1. Should not be left blank 2. Should not be a 'zero' value if property type is commercial 3. Should be equal to tax computed through formula and inputs such as area, unit rate etc..

Define Data Digitization Approach and Quality metrics (contd..)

Illustrative parameters for document data quality metrics/parameters

- Scanning resolution (DPI???)
- Scanning mode (8 bit gray scale, 24 bit color scale???)
- Output file format (PDF, TIF???)
- Expected data storage size per page

Define Data Digitization Approach and Quality metrics (contd..)

Definition of Roles, Responsibilities and liabilities for stakeholders in data digitization and migration project

Stakeholder	Responsibilities
Department <i>Illustrative</i>	<ol style="list-style-type: none">1. Quality and completeness of manual records2. Security of manual records3. Selection of service providers for data digitization and quality assessment4. Definition of quality expectations/requirements for digitized data5. Stock keeping and tracking of manual records6. Verification and validation of digitized data7. Verification and validation of migrated data8. Implementation of Security and controls on digitized data
Data digitization services provider	<ol style="list-style-type: none">1. Development of software for data digitization2. Provide tools and infrastructure for documents scanning3. Digitization of records4. Scanning of documents5. Provide print out of digitized records (based on number of iterations/validations needed)6. Proper upkeep and safety of manual/digital documents/data
Data Quality Assessment and Validation	<ol style="list-style-type: none">1. Definition of data quality criteria2. Data validation and verification post digitization and migration
System Integrator/ Software developer	<ol style="list-style-type: none">1. Migration of digitized data2. Data validation and verification post migration

Define Data Digitization Approach and Quality metrics (contd..)

- Definition of Data verification and validation approach
 - Tools and methods for data validation
 - % of records to be verified by the third party auditor (service provider) and employees of department
 - Templates for reporting the errors/gaps in the digitized data
 - Approach for providing feedback to the data digitization vendor for corrections
 - Number of cycles for data validation
- Define the data digitization timelines and synchronise the project go-live in line with the digitisation plan

Selection of Service Providers

Service Provider	Role	Key Selection Criteria
<p>Data digitization</p> <p><i>Illustrative</i></p>	<ol style="list-style-type: none"> 1. Digitization of data 2. Scanning of documents 3. Data corrections/ cleansing 4. Migration of data 	<ol style="list-style-type: none"> 1. Experience in large data digitization projects (based on size of department initiative) in similar context 2. Understanding of data quality requirements for department data 3. Tools and methods for digitization and migration 4. Experience of project management team proposed for engagement in similar context
<p>Data Quality Assessment and Validation</p>	<ol style="list-style-type: none"> 1. Data and Document Quality criteria definition 2. Conduct data validation and verification for digitized data based on defined quality criteria 	<ol style="list-style-type: none"> 1. Experience in data validation and verification in similar size and context 2. Strength and credentials of proposed tools for data validation and verification 3. Understanding of data quality criteria for department data digitization requirements 4. Experience of team proposed for conducting data quality assessment and validation 5. Approach for data quality assessment and validation

Data Digitization

- Application of Right software/ tools for digitization
 - For delivering quality and complete data, it requires quality systems and tools for digitization
 - The software used for data digitization should have appropriate built in process and security controls to address all the data quality requirements/metrics defined for each data field
 - The digitization software/tools must be thoroughly tested and audited for its functionality, inbuilt controls, security and audit trails
 - Selection of right equipment for scanning the documents

Data Digitization (contd..)

- Controls surrounding sharing of documents and data digitization
 - Define tracking approach and checklists for sharing the records and documents with service provider for digitization/scanning and reconciliation of the same
 - Define the information confidentiality and security requirements for the documents/ records
 - Sign Non-Disclosure Agreement with the service provider for ensuring data confidentiality and security
 - Implement physical security and monitoring (CCTV monitoring) controls to ensure documents are not stolen/destroyed/tampered (intentionally or unintentionally)
 - Security and controls surrounding IT systems used for data digitization/scanning
 - Restrict internet access
 - Restrict usage of USB drives/CD drives for copying and carrying the data
 - Frisking of operators during entry/exit for ensuring data/equipment is not carried out of secured premises

Data Quality Assessment & Migration

- Data validation and correction
 - To ensure accuracy and completeness of digitized data in comparison with the manual records
 - To identify the errors/gaps in manual records based on digitized data
 - e.g. variation in the total tax amount in the manual records and system computed tax amount (during digitization) based on the input parameters in manual records and the tax computation formula of the department
 - Digitized data can also be published to the citizens for confirming the validity and completeness
 - Process seeks inputs from citizens/businesses on the gaps in the digitized and published data and to bridge the gaps based on inputs provided by the citizens
 - Requires stringent controls and processes for validating the corrections requested by the citizens/businesses

Data Quality Assessment & Migration (contd..)

Data validation methods

- Verification of printed reports on digitized data and comparison with manual records by government employees and third party agencies
- Using tools for verifying gaps and errors in digitized data
- Published the digitized data to the citizens for confirming the validity and completeness
 - Seeks inputs from citizens/businesses on the gaps in the digitized and published data and to bridge the gaps based on inputs provided by the citizens
 - Requires stringent controls and processes for validating the corrections requested by the citizens/businesses

Data Quality Assessment & Migration (contd..)

Manual Verification/Validation	Using Tools
<ol style="list-style-type: none">1. Identify the officers responsible for manual validation of data2. Define % of data validation to be performed at each level in hierarchy (100%, 30%, 10%)...3. Define the report formats for printing the digitized records4. Obtain printed records for the digitized data from vendor5. Validate the printed records with manual data/records6. Identify the gaps and correct the errors in the manual records7. Sign each page of printed records for completeness and correctness (in case no errors are identified)8. Share the corrected reported/documents with service provider for addressing the identified errors in the digitized data9. Revalidation post data correction10. Sign-off on quality and completeness11. Preserve signed-off records	<ol style="list-style-type: none">1. Define the format/structure for sharing the digitized data2. Configure the data quality criteria/metrics defined for each data element in the tool3. Upload the digitized data4. Run the data quality tests as per pre-defined criteria5. Identify the gaps and share the gap/error report with the data entry service provider6. Obtain the corrected data and revalidate the data against the defined quality criteria7. Sign-off on quality and completeness8. Label and store the corrected data using appropriate media

Data Quality Assessment & Migration (contd..)

Data Migration

- Migrate the final version of digitized data into the testing and to the production environment
- Conduct the validation and completion tests (as per defined quality criteria) for the migrated data to ensure completeness and accuracy of the migrated data

End of Session