رؤية عُمان 2040
تعداد 2020
التعداد الإلكتروني للسكان والمساكن والمنشآت
Electronic Census of Population, Housings and Establishments
NATIONAL CENTRE FOR STATISTICS & INFORMATION
Enhancing Knowledge
SULTANATE OF OMAN

Electronic Census System for the E-Census for Population, Residences & Establishments 2020

E-Census Documents
Series 2020 | 07

His Majesty Sultan Haitham Bin Tarik

- may Allah protect him-

endorses E-Census 2020 results of Population, Residences and Establishments as of 12 December 2020. His Majesty the Sultan expressed his satisfaction with the outcome of this major project which was implemented as scheduled. His Majesty the Sultan underscored the significance of data and indicators provided by the Census in enhancing Oman Vision 2040, as well as the Census's important implications to developmental planning in all sectors of the Sultanate.

14 December 2020

Table of Contents

Introduction

In light of the change in statistical work tools in recent years and the future trends of most authoritiesrelying on electronic systems. The dependence on administrative records as a source for producing official statistics and statistical indicators has become the best reliable tool. The implementation of this census came in line with technical developments, through which the Sultanate presented a pioneering model by building a unified and integrated statistical information system on population, residences and establishments. It is based on the administrative records of government and private authorities in the Sultanate. The system is responsible for publishing statistical indicators and reports in accordance with international frameworks and standards, so that it acts as a tool to verify the validity of data and a means to integrate various data coming from different authorities in the Sultanate.

This system covers more than a thousand variables and is updated in a continuous way through data flow over electronic connection or traditional file uploading. The system ensures the continuity of data flow, checking its quality and completeness, and then applying the pass tables between administrative and statistical classifications. The scientific methodologies adopted in the electronic projects and enumeration units for census projects have been applied.

This document deals with the electronic census system starting from the third stages of implementation in 2018, passing through a comprehensive detail of structuring the system's inputs in terms of population, residences and establishment files, structuring its databases. In addition how the census system works and the documents and standards that have been implemented for the sake of data integrity and quality and integration.

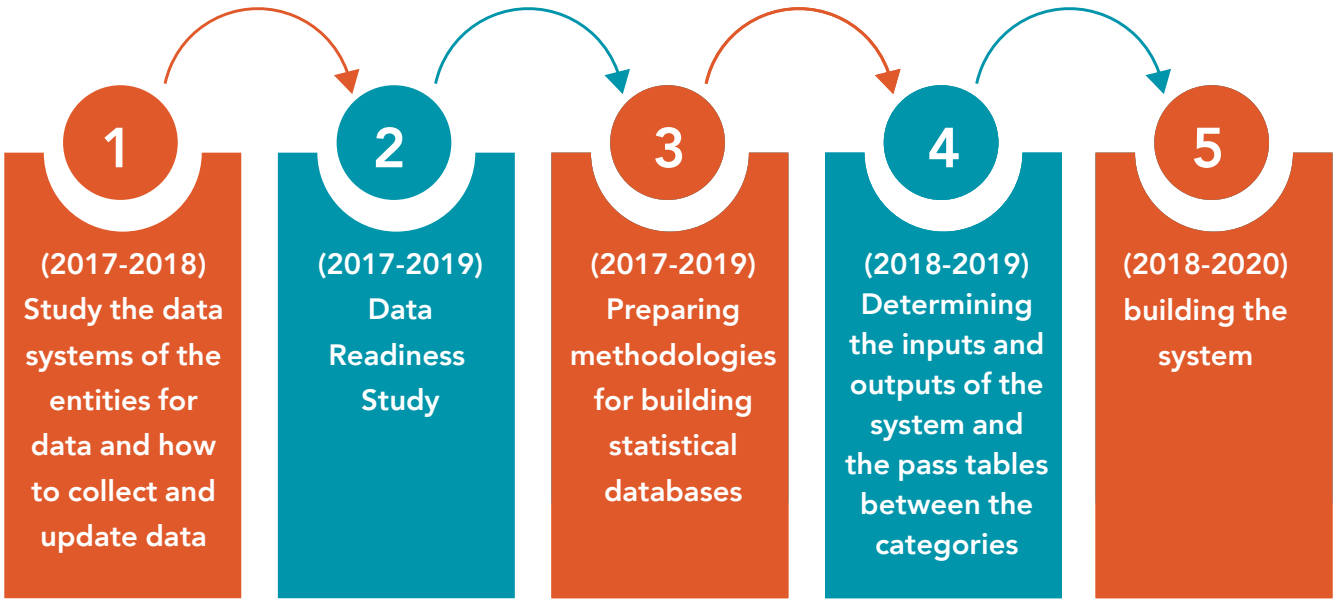
1 | Electronic Census System and Implementation Phases

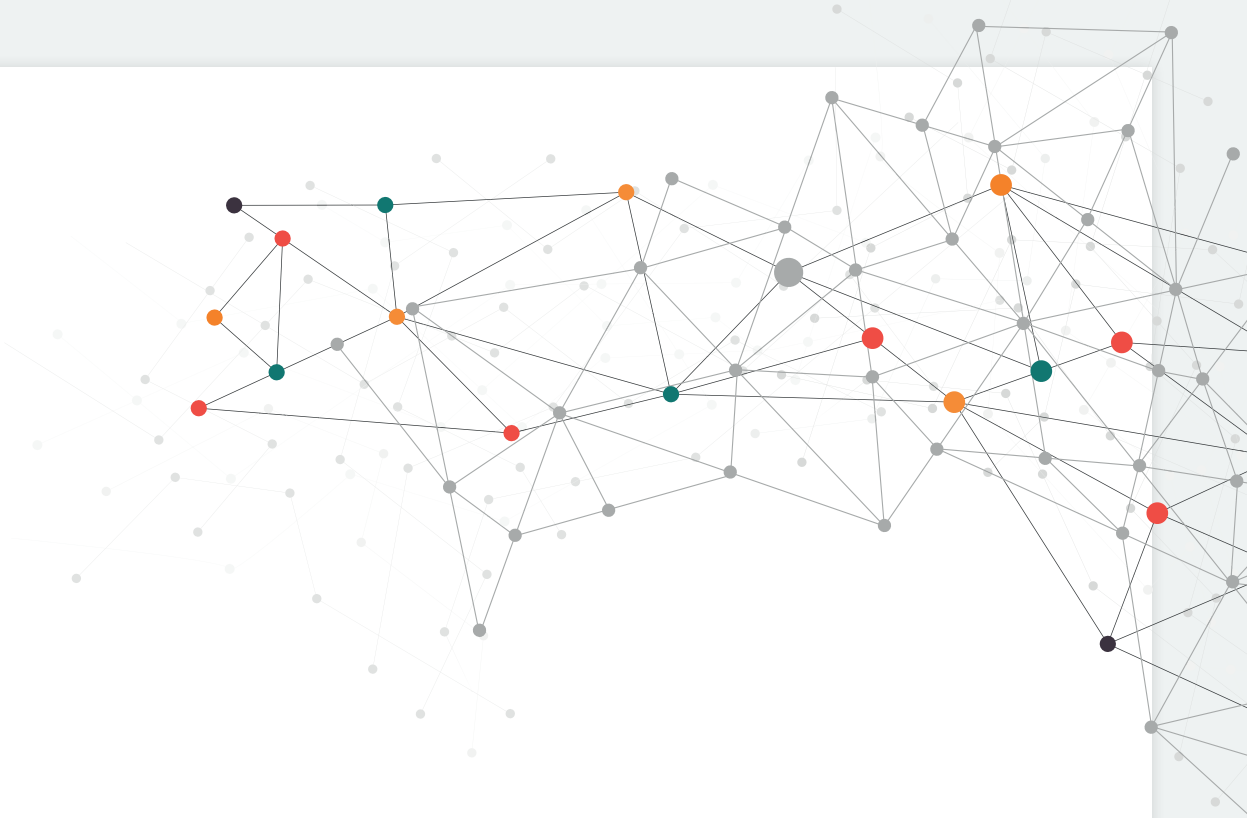
Electronic Census System:

It is a comprehensive and integrated national data system of high quality and electronically linked to databases from various sources. It aims at building up-to-date, audited and linked social and economic information with a spatial character through an electronic information system. It further aims at providing decision makers with real-time and updated data. This data carries, in particular, all indicators related to population, residences and establishments and their characteristics.

Phases of Implementing the Electronic Census System

The project's work started in mid-2017 and ended in December 2020 after approving the census results and handing over the electronic census system. This system will be able to conduct integrated future census projects in a short time. It will work to sustain the accurate electronic exchange of data from various relevant authorities, and build the required statistical indicators. Through the developed timetable, work has started on building the electronic census system in the third phase of the year (2018-2020). After the initial plans and methodologies were prepared in the first phase (2017-2018), data was collected. Then, correction and matching was carried out. The main designs of the databases were prepared through the establishment of special laboratories concerned with saving and analysing data in the second phase (2017-2019). Then the preparatory work and construction of the system began. Finally, the results were published according to the phases.





6

linking 2020 between entities

7

2020
Test the System and portal

8

12/12/2020
Completing the census

9

14/12/2020
Approving and dissemination of results

1.Study data systems of the data exporters and how to collect and update data.

To build a data system that is fed by the various authorities, it is necessary to study the databases and systems used by those authorities. This is in order the outputs of these authorities and their databases are compatible with the data system that the census wants to establish. Because the majority of authorities in the Sultanate have built their electronic systems and databases according to their needs, it necessary to modify, create and add many modifications within the structure of the databases and to develop some of them as well. In addition to the establishment of new systems and databases for authorities they do not have them. The base from which the census system is derived is able to fetch data from various exporting bases to ensure the existence of a common database for all authorities.

2.Data Readiness Study:

When reviewing the data bases and systems of the various authorities in the Sultanate, it is essential to lift up their quality and ensure their completeness and compliance with the reference databases before bringing that data into the census system. This is to ensure obtaining correct, accurate and complete data that reflects the actual situation of the population, residences and establishments in the Sultanate.

For that reason, the census team worked on studying all databases of the relevant authorities. Then, processing, correcting and completing these data in the authoritys' databases so that they are ready to feed the census system.

3.Preparing methodologies for building statistical databases:

Naturally, there are international standards and classifications used in classifying data issued by international organizations and bodies related to various topics. There is an international classification of education, employment, activities and professions. These standards shed light on the way of classifying data for each category, so an individual is classified as enrolled in education or not, and when he is classified as illiterate or not illiterate, and when he is classified as working in the manufacturing industries or other activity. There are measures for the rest of the data as well that should be classified in the correct way.

Consequently, the census project management has prepared calculation methodologies for the data it collects from the authorities and put the processes for classifying that data within the international classification standards.

A large set of conditions related to the logic and validity of the data were also applied, which will

be discussed later in this document.

4.Determining the inputs and outputs of the system and the pass tables between the categories:

The main census data basket relied on studying the administrative records in the various authorities from which the data will be collected. It mainly depends on international recommendations regarding census data.

It was also proposed to add another set of data that serves the Sultanate and decision makers in a way that supports the development of policies and development plans in the country. The greater the volume and quality of the available data, the greater chances of developing appropriate decisions and plans for development in the Sultanate, and distributing and providing services in line with those data.

Thus, the system expanded in the volume of data that is brought from the authorities with total commitment to the census data basket. In light of these data in the authorities upon various classifications and systems, a set of pass tables was developed. It is a means of converting the classifications in the authorities in the Sultanate to the international classifications used to display the data.

The fields and data that will be fetched from each authority were determined. The periodicity of updating the data between the systems of the authorities and the census system, and the format and arrangement of these data are set in line with the requirements of the system.

5.Building the system:

The construction stage is the most important stage in the census system, especially in determining the digital formula on which the system depends (programming). It also requires the fulfilment of all previous stages in terms of providing databases in which data is stored. The auditing standards are built and transferred in it as well, from the theoretical formula to the electronic format. How to integrate, arrange, and describe these data in the system, and the tables, images and shapes that the system forms and outputs, and how to disseminate that information.

The system has witnessed innumerable alterations in order to be developed and brought in its modern way. During the construction and testing phase, a large number of improvements appear that can be made to the system in order to speed up the system, raise the accuracy and validity of the data, and ensure that methodologies are applied correctly.

6.Linking with the Authorities:

At the beginning of the system's work, the process of entering data was done manually. The data was collected from different authorities. The technicians working on the system uploaded that data after being configured to the system's databases.

In order to provide the system with updated and continuously connected data, and in coordination with the Ministry of Transport, Communications and Information Technology, the electronic census system has been electronically linked to several authorities. The data is sent from these authorities to the system on a daily and continuous basis with different reference dates. The system receives this data in the first phase of the databases, where its quality is checked technically, and then inserted into the correct spreadsheets. Further, the data that contains errors are kept separately in error tables for the purpose of correcting them and re-inserting them into the system.

The number of authorities to which the system has been electronically connected is eight:

1- Royal Oman Police
2- Ministry of Health
3- Ministry of Education
4- Ministry of Higher Education, Scientific Research and Innovation
5- Ministry of Commerce, Industry and Investment Promotion
6- Ministry of Labor
7- Ministry of Social Development
8- Nama Group

7.Experiment of system and data portal:

This stage is one of the most essential stages in which the correct functioning of the system is verified and that there are no errors in the input and output process. Also, all standards, methodologies, pass tables, linking and calculation operations are done correctly.

This was done by following up the volume of data which is sent through electronic linkage and the data that was manually collected from the authorities databases. Then, assuring their validity, number and accuracy through the processes of the system and between the census laboratories whose main objective was to correct and update the data and ensure the validity of data and outputs.

As a result, this phase worked on correcting a large number of operations on the census system until it was confirmed that the system was working properly.

In order to ensure the readiness of the system, the census project management worked to conduct several trial censuses prior to the actual counting day to determine the system's working period and the time period for approving the results and publishing the data.

8.Completing the Census 12/12/2020:

After a series of experimental censuses, in which the validity and accuracy of the data flowing from different authorities were confirmed, the data of the system was updated by taking a picture of the data on 12/12/2020, which is the reference day approved by the electronic census project 2020.

9.Approving and dissemination of results:

His Majesty Sultan Haitham bin Tariq - may Allah protect him - approved the results of the census on 12/14/2020. In a press conference held on the same day, the results of the census and the most important statistical indicators were announced. The census data portal was launched that contains all the data and assembled indicators.

2 | Structuring the inputs of the system:

The data is uploaded to the system through the data files that were brought from the various authorities or through direct linking with the authorities. The files collected outside the linkage system are manually verified before feeding them to the system. Sometimes they are verified by the system of the authorities that sent data through the electronic link:

Here we explain the two ways of entering the data into the census system:

1. Data collected from authorities outside the electronic link:

- Data file must be in CSV
- The properties file contains:

1. File name
2. Authority Name
3. Reference Date
4. Number of Records
5. Categories used in the file

- For the geographical data, it is in (gdb) format, and a properties file is also attached to it. After collecting the data, it is verified through the statistical program (SAS), and then those files are uploaded to the census system over (FTP Wins cap). Then, the verification, processing, merging and linking phases with the rest of the other files are completed.

2. Automated data link:

The system has been directly linked to several authorities in order to receive the updated data faster through the web service using (SOAP) program. The system receives daily records from authorities that are sent electronically in the same CSV file format. This method has the advantage of being faster in receiving updated data.

3 | Structure of the databases:

The average number of records processed has reached more than 74 million records located in 38 files received from authorities. It includes 30 million records for the identification data of the population, 4 million records for manpower data, 3 million for education, 6 million for marital status, 1 million for births and deaths, 12 million for establishments and 19 million records of unit data. As a result of this huge amount of data that flows through the electronic linkage of the parties, it was necessary to put in place an organized mechanism through which the data would pass several phases.

4 | Building basic databases

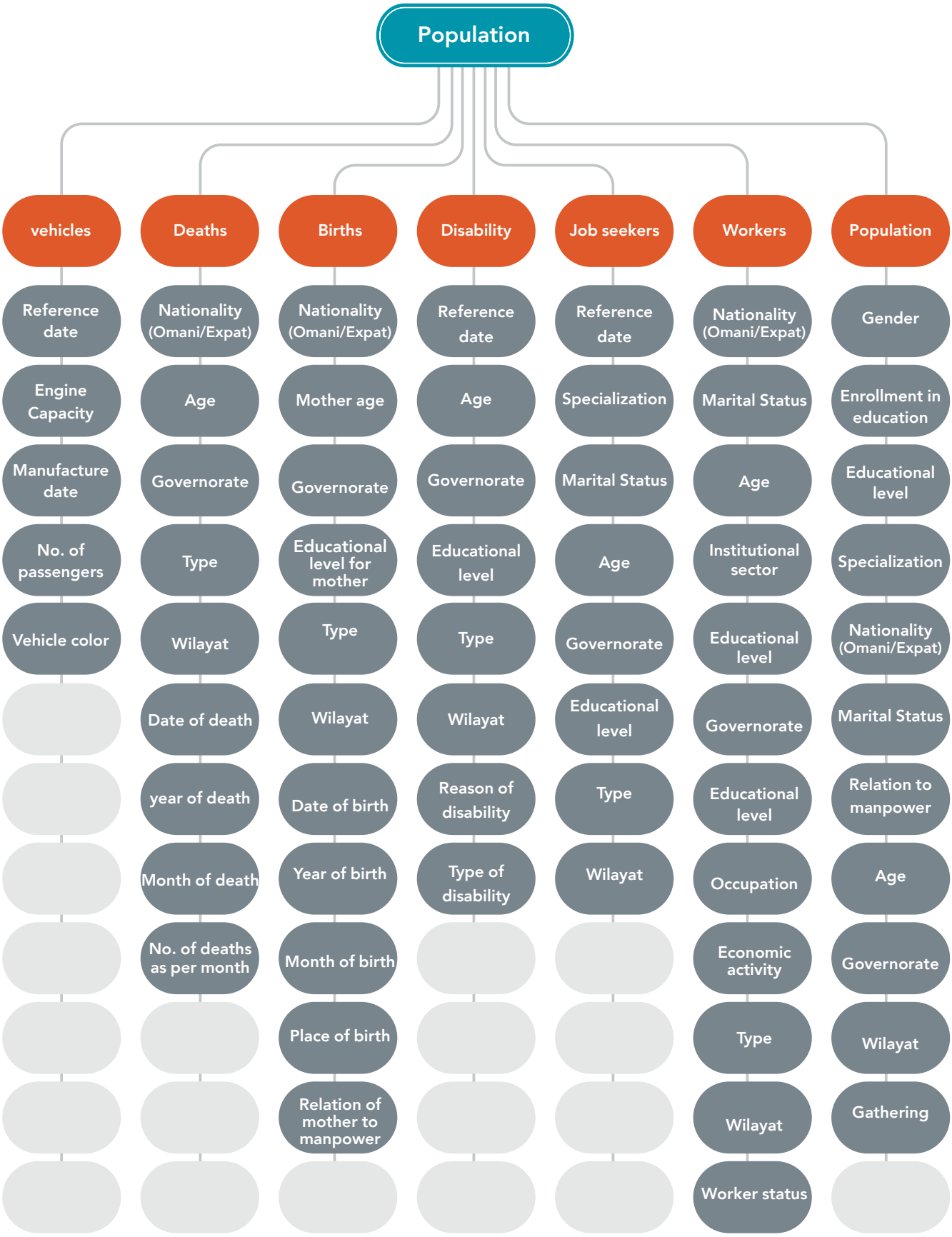
The electronic census databases were built to suit the basic objectives of the statistical system. It contains three main rules: the rules of population, residences and establishments. They form the artery through which the updated and connected data flows with each other. This stage has been completed, the part of the system has been examined and the rules for checking and correcting data has been finalized. This stage includes manual data download without activating the electronic connection so data will flow automatically.

Structure of population files

All individual data from government authorities, which are basic sources for forming a population database are:

- Royal Oman Police
- Ministry of Education
- Ministry of Higher Education, Research and Innovation
- Council of the Administrative Affairs for the Judiciary
- Ministry of Social Development
- Ministry of Health
- Ministry of Labor

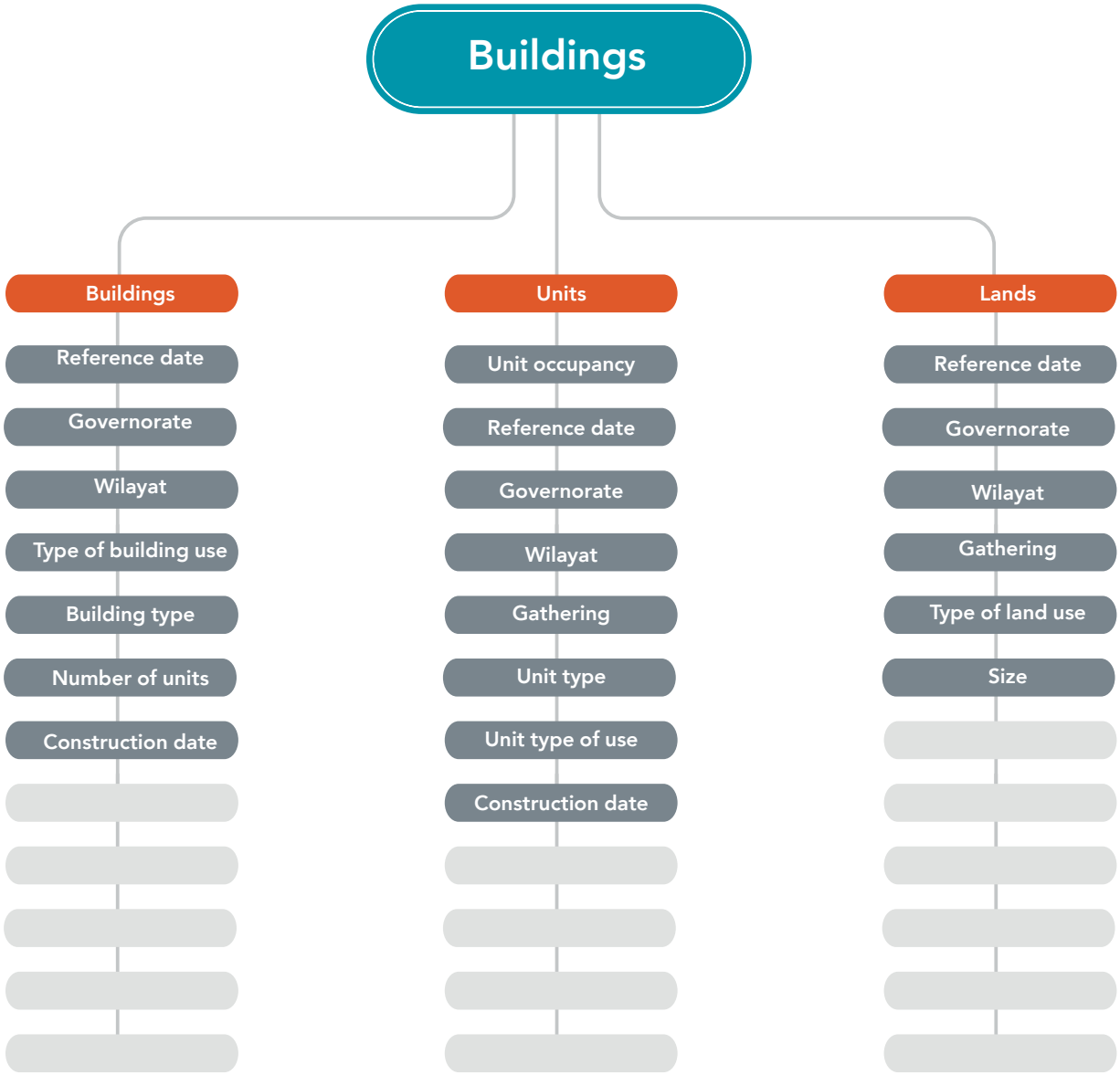
The population database consists of the following sections:



Structured unit files:

- All individual data from government authorities, which are basic sources for forming a housing database, which are:
- Ministry of Housing and Urban Planning
- National Center for Statistics and Information
- Municipalities
- Namaa Group

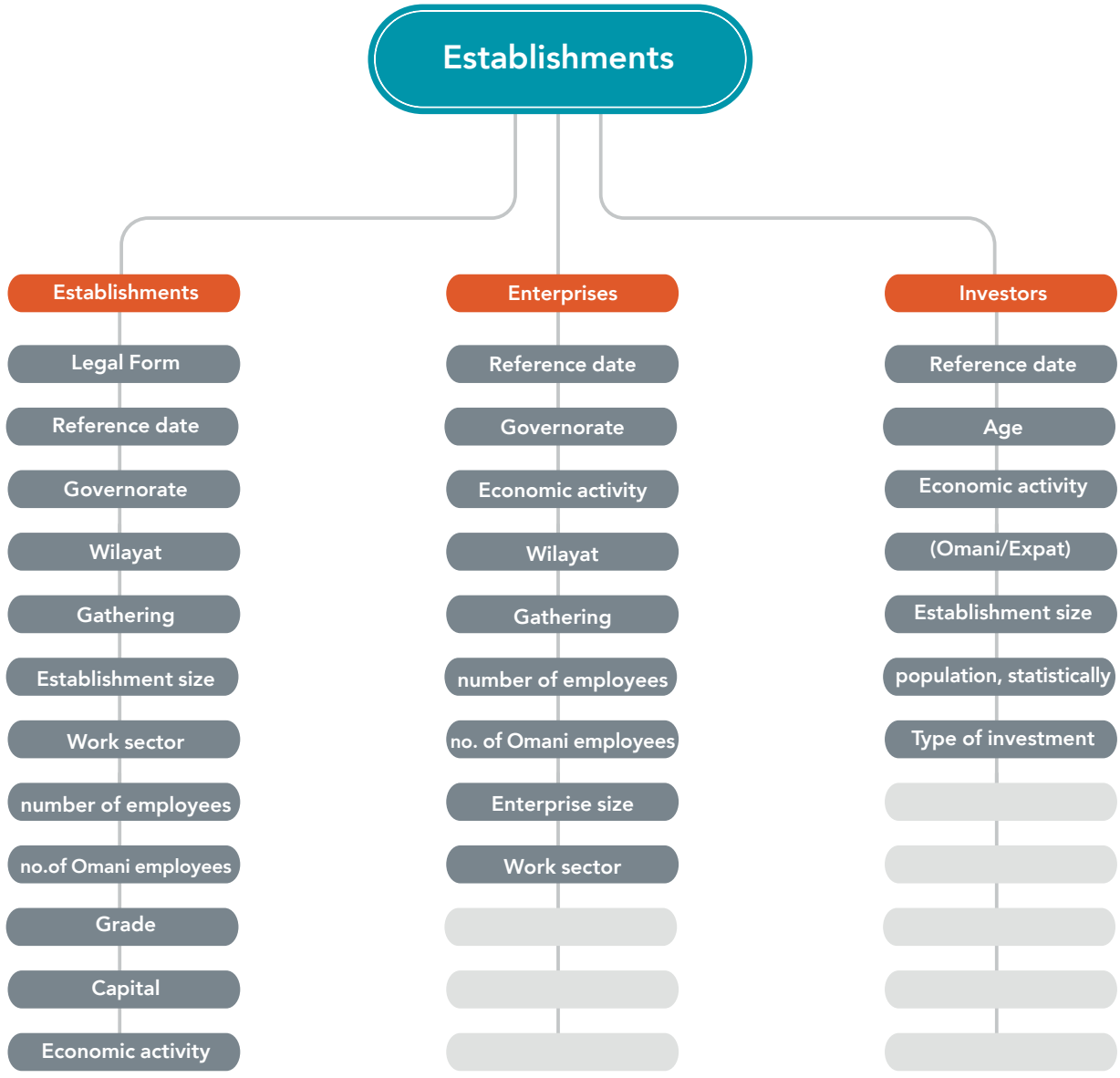
The housing database consists of the sections specified below



Structure of Establishments files

- All individual data from government authorities, which are basic sources for the formation of the establishments database are:
- Ministry of Commerce, Industry and Investment Promotion
- Ministry of Labor
- Municipalities

The authorities database consists of the sections specified below:



The following table shows the data received from the official authorities and the number of variables for each authority:

S	Authority Name	Number of Variables	Link Services
1	Royal Oman Police	36	Individuals
		28	Addresses
		11	Family Tree
		9	Visas
2	Ministry of Health	59	Births
		23	Deaths
		43	Daily visits to patients
3	Ministry of Education	33	School Pupils
4	Ministry of Higher Education, Scientific Research and Innovation	40	college and university students
5	Ministry of Trade, Industry and Investment Promotion	5	Establishments
		27	Establishments Addresses
		8	Establishments Capital
		8	Investors
		4	Legal form of Establishments
		4	Establishment Status
		3	Enterprises
		7	Enterprises activity
		28	Enterprises Addresses
		5	Enterprises Status
6	Ministry of Labour	37	Omani Employees
		37	Expat. Employees
		13	Job Seekers
		15	Retirees
7	Ministry of Social Development	30	People with Special Needs
8	Nama Group	25	Electricity Accounts
		6	Monthly Concumption of accounts
		9	Geographical location

The following table also shows the data sources and the most important uses for e-linking process:

Authority	File	The most important uses
Royal Oman Police	Individuals	Population characteristics (gender, age, marital status, (nationality
	Addresses	Residence address, family, connection with the housing unit
	Family Tree	Family, relationship to the reference person, marital status
	Visas	(Operation (Supplementary Resources
	Vehicles	
Ministry of Education	e-Portal	Education characteristics (enrollment, type of educa-(tion, educational level, specialization Education address
Ministry of Higher Education, Scientific Research and Innovation	ASAS	
Ministry of Labour	Workers	Employment characteristics (relationship to the labor force, profession, economic activity, work destination, (and work status
	Job Seekers	
	Retirees	
Ministry of Social Development	People with Special Needs	(Disability characteristics (type, cause, degree
Ministry of Health	Births	(Newborn Characteristics (Supplementary Data
	Deaths	(Cause of death (Supplementary data
	River of Healing	healing address
The Council of Administrative Affairs for the Judiciary	marriage and divorce documents	Marital status: a source for comparison
Namaa Holding Company	Accounts	Characteristics of units and buildings (occupations, (type and use Unit address
	Account sites	
	Consumption	
National Structure	land layer	Geographical matching between land, buildings and units
	building layer	
Ministry of Housing	Lands	(Land characteristics (ownership, use and area
Ministry of Commerce, Industry and Investment Promotin	Establishments	Characteristics of authoritiesand establishments (legal form, economic activity, capital, address, year of es-(tablishment and sector
	Enterprises	
	Activities	
	Addresses	
Chamber of Commerce and Industry	Establishments	Degree
Tax Authority	Establishments	Tax registration, sales number and profits
Ministry of labour	Establishments	Enterprise/Establishment size The main activity of the Enterprise / establishment
Municipalities/ Governorates Affairs	Establishments	The status of Enterprise / establishment activity
	activity	

5 | Geographical information in the census system:

It is information designed to process, analyse and manage all geo-referenced data that is entered into the system. It helps to know the spatial boundaries and administrative divisions of the governorates, Walayats and villages in the Sultanate.

Geographical information analysis helps in answering a lot of statistical information, such as:

- Population in the Governorates
- Population in the Walayats
- Population in the Villages

6 | Technical and Functional Requirements Document PRD:

We mentioned earlier in the stages of the census system that the stage of building the system is one of the most important stages in the census project. It is the core driver of data; starting from data entry and ending with its output and between the two processes of correction, classification, review, arrangement and output.

It was necessary to define the most important technical and functional requirements for the operation of this system. Therefore, the census project proceeded to make a document of technical and functional requirements, which allows understanding the functions, processes and software that the system is supposed to perform and how it should work. It contains the following requirements:

1. **Functional requirements:** Through which the system will work to link databases of population, residences and establishments over data of government authorities and flowing in real time through ETL operations.
2. **Technical requirements:** It is the requirements for security, network, platform and integration through the electronic census portal, which provides statistical data to all members of society. Here are some examples of the technical requirements:

- Work devices in the electronic census project

Below are the minimum hardware specifications to work on the electronic census system:

- Processor: Intel Core i5 or higher
- RAM: 16 GB or higher
- Storage: 100 GB or higher
- Network: 1 Gbps

Programs:

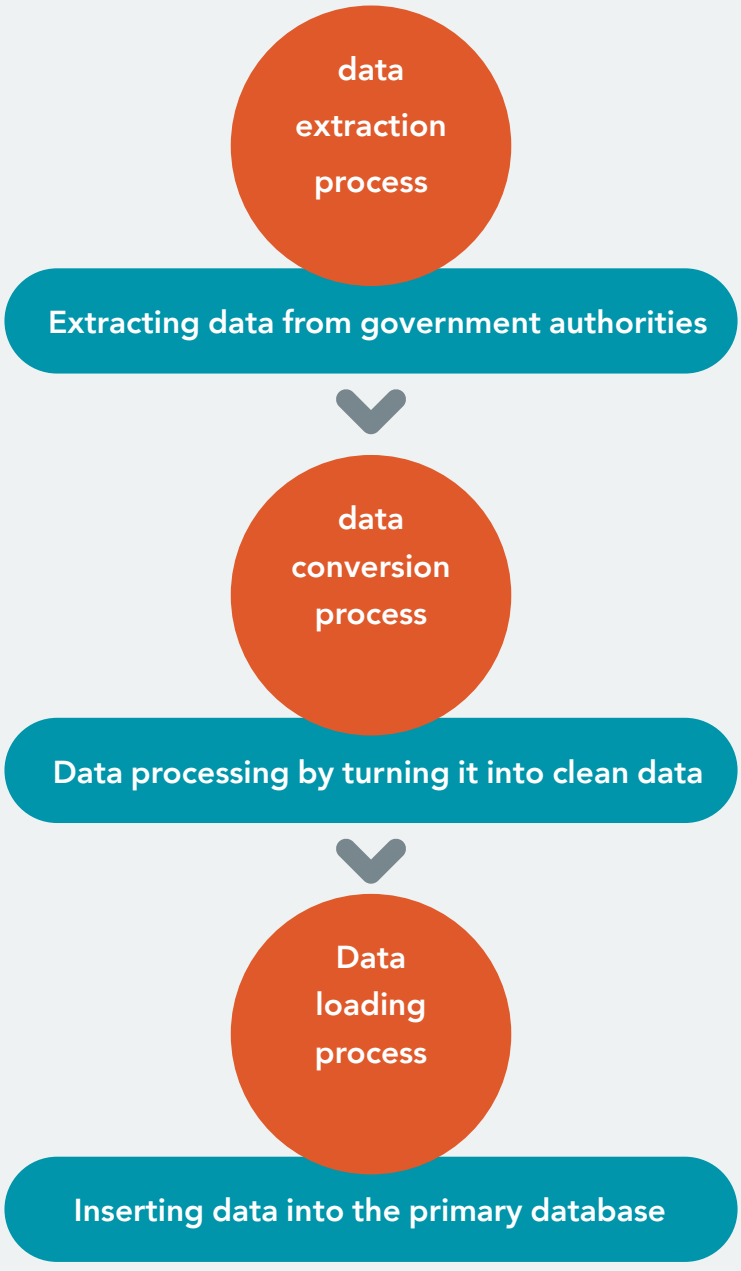
- PENTAHO DATA INTEGRATION (PDI) is the primary tool used in the design of ETLS
- Operating System: Microsoft Windows 10 (x64 bit)
- FTP Client: Any standard FTP client or WinSCP version 5.15 (or higher).
- Toad Client: SQL for version Oracle DBA or Oracle SQL Developer version 19 (or higher) or any RDBMS client tool
- Web browser: Google Chrome (recommended), Microsoft Edge, Mozilla Firefox

- Various tools to improve efficiency: Microsoft Excel 2016 or higher; Microsoft Access 2016 or higher, Notepad++, EmEditor, Adobe Reader.

Data Quality Assurance Criteria

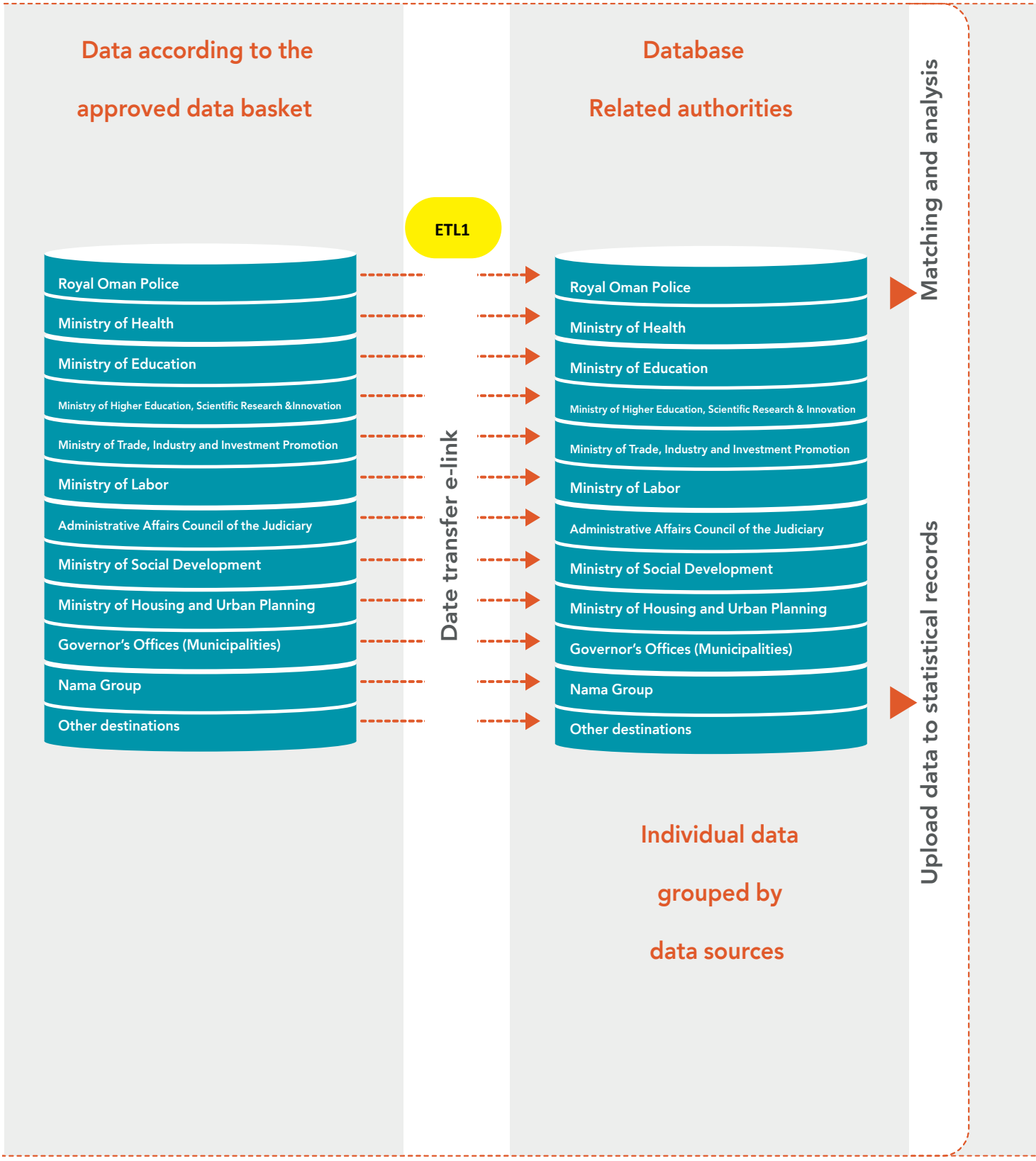
The success of statistical systems depends on their ability to analyze the available data and extract statistical data and indicators. Extract, Transform, Load (ETL) operations are the best operations used to verify the quality of data. ETL works on (extracting, transforming and loading) the data flowing from the main sources (Governmental authorities).

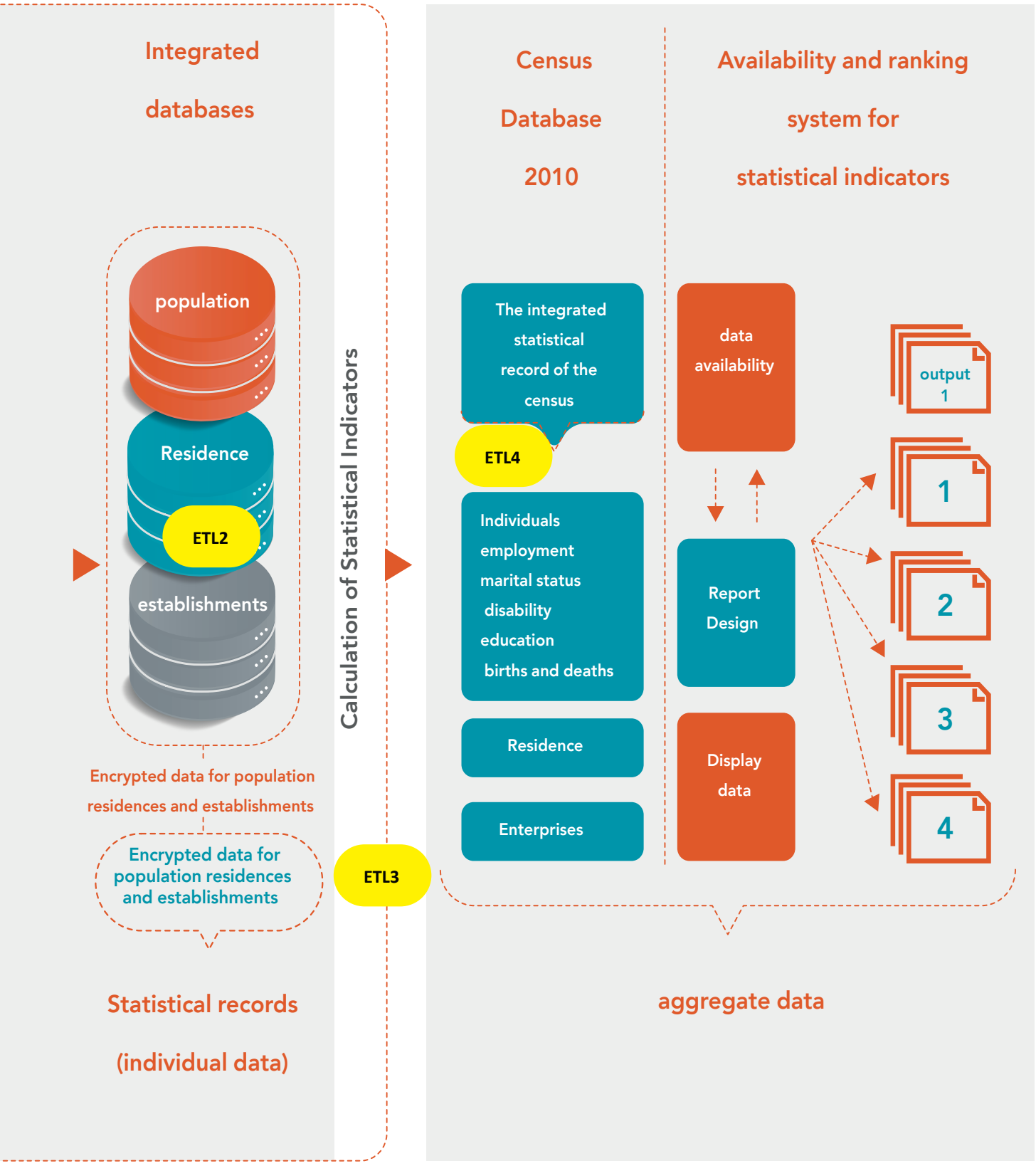
The (ETL) document transfers data through three operations during the electronic connection, as shown in the figure below:



A primary data warehouse has been built by Census Labs, and more than one data warehouse ETL1,ETL2,ETL3,ETL4 has been created to be applied in each database in order to clean up the data and improve its quality.

The following shape shows the components of ETL operations used in the enumeration.




Integrated databases
population
Residence
ETL2
establishments
Encrypted data for population residences and establishments
Encrypted data for population residences and establishments
Statistical records (individual data)
Calculation of Statistical Indicators
ETL3
Census Database 2010
The integrated statistical record of the census
ETL4
Individuals
employment
marital status
disability
education
births and deaths
Residence
Enterprises
Availability and ranking system for statistical indicators
data availability
Report Design
Display data
output 1
1
2
3
4
aggregate data

From the shape it is obvious that there are four stages of ETL operations that can be explained according to the following:

- **The first stage:** Uploading the data of the administrative records received from the concerned authorities. It is in turn checks the technical quality of these records, treats them technically, and indicates errors, repetitions and availability of mandatory fields. This stage also supervises the comparison of fields classifications that were received from the authorities with the classifications officially approved in the system. The schedules for this stage are divided.

Error reports issued by the system at the end of this stage constitute a basic source for knowing the quality of data received from authorities. These reports can be extracted and shared with the concerned authorities for the purpose of correcting errors.

This stage is also based on the processing of land data, buildings and accounts, which make up the geographical data from the system, and linking them to the respective servers through site positions.

- **The second phase:**This stage is concerned with linking data from an administrative point of view. It links the data and creates new tables in which the shared data from several authorities are linked, such as education data from Ministry of Education and Ministry of Higher Education, Research and Innovation. They constitute the main source of education data for Omanis and education data from Ministry of Labor for expatriates. This stage is further concerned with linking individuals' data with authorities, such as linking individuals to families and linking employees to authorities and establishments.

This stage also examines the data logically and generates customized reports to display these errors as an example of individuals' errors of educational age who are not included in the education records.

- **Third Stage:** One of the most important objectives of this stage is to encrypt data and everything that symbolizes identity to maintain the confidentiality and privacy of data. Data encryption includes all civil numbers, names, commercial records numbers and make them appear in an encrypted form.
- **The fourth stage:** The objective of the system is to provide statistical data and indicators. This stage processes data derived from the previous stages in a statistical manner, connects the dimensions with facts, and forms the part of the related statistical indicators that are presented in the electronic census portal and final reports. The data is filled in the tables based on data reference dates that are assigned at the beginning of the stage.

The four stages of the system's work depended on many of the necessary criteria to ensure the accuracy and completeness of the data at the beginning. Then, tabulated, collected, encrypted and building rules. Finally, it will be published, which are the stages we referred to it earlier. Therefore, each stage has a set of standards and conditions that were constructed in the system. It ensures that data is produced correctly. The following is a description and clarification of these criteria:

Technical validation criteria:

1. Verify the name of the data source: The file name of the data source will be checked. If the file name is not the agreed name of the data source, the system will not be able to identify the file so it will not load it into the suitable administrative record for the divisions of the electronic census system.

2. Mandatory information validation: Mandatory fields will be checked in the data source file, if the mandatory fields are not available, the record will be stored in the error table until it is corrected,

- Example of mandatory data Gender (Gender) information is mandatory for all individuals' records

3. Data Format Validation: The system is designed to predict data in a specific format from data sources.

4. The system is designed to predict the date of birth in the form of DD/MM/YYYY. For example, if the date of birth is presented in the format 'MM-DD-YY', the system will interpret that specific date of birth as an unknown date format and store the problematic history in error tables.

5. Checking the classifications against the added classifications in the system: The classifications are a set of possible values for a given field and here are some examples of classifiers:

- Simple is: Gender Classifier: male or female
- Marital status: single, married, divorced, widowed
- Legal Status of Establishment: Sole Proprietorship, Limited Company, Joint Stock Company

6. Validations of business keys, as each census topic has a primary data source:

for population; it is Royal Oman Police, and for authorities; it is the Ministry of Commerce and Industry.

Logical validation criteria:

Logical coherence in statistical data is the degree of consistency between data. Logical coherence is considered one of the most important criteria that indicates the extent of quality in logical statistical data. The census project was keen to achieve the logical data interconnection criterion with high earnestness by following a number of technical methods and procedures that achieve this dimension.

Logical auditing is examining the consistency of data internally within the database, so that the congruence between the nature of the field and the type of answers or data is monitored, as well as the consistency between the various data or answers. See the table (Ahmed is 4 years old and holds a PhD, and Mohamed is 8 years old and his marital status is married).

An example to illustrate some logical errors			
	Age	Education Level	Marital Status
Ahmed	4	PhD	Single
Mohammed	8	(Basic Education (Cycle 1	Married

As a result, census laboratories have been established to check the data and ensure its consistency between various data. The data audit goes through several stages, the most important of which are:

2. Manual desk audit: It uses the traditional approach represented in manual examination by a specialized auditing team that reviews all data in the audit laboratories.

3. Automated auditing scope: The data in this method is checked and reviewed in one go after being entered electronically, through systems and programs in the computers. They include the auditing rules that have been identified and applied to the data set and applied to them, to detect errors or to identify unacceptable formulas. The audit teams submit reports related to errors that require correction. Those data will be processed. Reports related to any of the recurring errors, will be sent to those concerned to avoid these errors in the future.

The data audit methodology in the electronic census 2020 is based on the following approach:

- Defining a detailed, comprehensive and specific data basket.
- Gathering data from approved sources
- Verifying the accuracy of the data and its conformity with each other.
- Correction and completion of data
- Data Consolidation

After those stages, a large set of logical criteria was produced that guarantees the validity and accuracy of the comprehensiveness of the data entered into the system. It has exceeded 100 bases classified into groups as follows:

- General Standards
- Standards related to family relations
- Standards related to marriage and divorce
- age-related Standards
- Standards related to education
- Employment-related standards
- Standards related to establishments
- Standards related to units

The logical criteria are divided according to the rules into two parts:

Warnings:

It means that there is suspicion about the record or the event rather than a direct breach of the rules. A typical example of the severity type "Warning" if a person over 100 years old. The mentioned situation is not a violation of rule but the registration or event needs special attention.

Errors:

In this case, either important information is missing in the system registry or there is a direct violation of the rules. For example, there can be multiple reasons or explanations for a situation in which a man has five active marriages in the system and they are as follows:

- There is missing information in the system about divorce from an ex-wife
- There is missing information in the system about the death of the wife
- There is wrong information about the marriage (the civil number is wrong)
- Violation of the actual law if the spouse has five active marriages

Accordingly, the data source is contacted and errors are checked and fixed. There are two types of reports related to the logical criteria:

- A report showing the summarized results of examining databases classified into groups with the number of records that failed in certain rules.

- A detailed report shows the records which failed in certain rules

Conclusion:

The electronic census system was able to form a giant national database. It connects with a group of authorities that are considered basic sources of data in the Sultanate, whether those related to the census or other additional data that contribute to the development of policies, programs, plans and support the decision maker with real-time information and data.

Thus, it has become necessary for the data to proceed flowing periodically and continuously from the various authorities to the electronic census system and to work with them to improve the data and expand it in the future which might be of interest to decision makers.

It is very important that the other authorities to be connected through the electronic link with the system. This is to form a large database that enrich the authorities from inquiring from different entities. It helps in facilitating the administrative work in these authorities without the need for identification papers or verifying the validity of certain data, since this system includes everything that may be required by various authorities to complete their work.

E-Census Documents
Series 2020