Standards

For Data quality in

State institutions systems

Document version	Date
10.00	18.01.2016

Content:

1. Glossary

Term	Description
Data	The word "data", if it refers to this document, will mean basic facts for entities (subjects and objects) collected and stored by the institutions, and from which by means of processing and analysis can "result" into information within the frames of the administrative procedures.
Register	Register means the inventory list, inventory book, catalog, in a certain order, chronologically, by numbers etc, facilitates the work not only in organizational terms and provides accuracy in recording the data. The register has the role of a public record.
Records	Records means a review of operations, inventory of subjects and objects that records are kept for, etc., and the same does not have to be a public ledger. "Records" means recording of certain basic and specific data, structured data set established by law, whether centralized, decentralized or dispersed on a functional or geographical basis, and regardless of whether it consists of individual or aggregated (aggregate) data[1]; Data collection or data whose degree of publicity determines the institution that collects the data.
Data collection	Data collection means registers set up by law, records specified by law or general acts based on the law, databases and data kept on the basis of public and other authorities.
Database	Database means collection of data stipulated by certain rules and standards with determined ways of running, updating and terms of storage by the holder of the base[2]. Database is a structured collection of records or data held in electronic form[3].
Classifications	The classifications are set of discrete, detailed observations that are mutually exclusive and that can be assigned to one or more variables in order to compare and present data
Nomenclatures	Nomenclatures are systematic naming of things according to certain principles and criteria that represent the systems of names
Public ledger	Collection of data with public character or the person who has an interest and is able to get data from the public record personally or electronically.

---

[1] Law on health records.
[2] Law on State Statistics. Law on health records.
[3] Law on State Statistics. Law on health records.

Term	Description
Entry	Any documented information, irrespective of its features, medium, physical form as well as the way it is stored. Entry is written text with specific contents. Document / entry is recorded information created or received in the initiation, execution or completion of the activity of a body, legal or physical entity and includes sufficient content, context and structure, so that It gives evidence of that activity regardless of the form or medium[4].
Entity	An entity is an object that is followed by a set of attributes, attributes are data that characterize the entity they belong to a particular data type (text, number, date ...)

---

[4] Law on archive material

2. Introduction

The functioning of state institutions, especially in the part of service providing to citizens and businesses also covers the process of entering, storing, processing and delivering data. The quality of the service state institutions provide largely depends on the quality of data they possess. With new information technologies where data storage is done in databases or other types of electronic forms for the purposes of data exchange**, question arises regarding the quality** of the data in electronic form. This issue should be reviewed continuously in order to create policies and measures properly and carry out activities that would improve the quality. In administrative proceedings, there can be punitive or other damages to the citizens, dissatisfaction with the services received by citizens, businesses or institutions due to poor data. Therefore the provision of services to citizens and businesses as well as the exchange of data and documents between state institutions imposes the necessity of possession, processing and exchange of **data with high quality.** Ways of carrying out processes which include entry, validation, verification, storage and processing of the data up to this point are allowed to take place independently in the institutions.

In practice, during the electronic processing of data, there can be high quality data but also data with quality that is not at the required level and should be improved. Moreover, there are data where there are errors such as misspelled letters or signs, lack of letters, signs or numbers, obsolete and out of date data, and more. There are also problems in semantics, structure and format of the store, there are open questions that refer to the manner of correction of data, difficulties in clear determination the physical or legal entities, open questions about the existence of unique identifiers etc.. There are also difficulties in the exchange of data and documents between the institutions, where there the unambiguous determination of the required entity could lead to **inability** or difficulty in sharing or linking data.

Quality data support decision-making process, and to this end it is necessary to develop standards whose implementation will ensure the quality of data.

Standards for data quality are framework for governance and accountability for data quality, with continued commitment to providing high quality data in any organization.

The quality of the data recorded in the registers is provided via adoption of appropriate standards and procedures for their implementation. Furthermore, it is necessary to implement systems and processes that ensure data quality as part of ordinary business activities. Also, the state bodies\ employees who are part of the systems and processes that ensure quality of data should have appropriate knowledge, skills and capacity to achieve the desired goal.

2.1. Need for quality data

This paper aims to identify and describe the criteria for data quality and identify a process for creating and maintaining quality data, because delivery of quality data is not an end in itself, but is an important part of providing quality and timely services to citizens, businesses and other institutions.

No determination and no resolving of gaps in data quality can lead to making wrong decisions and solutions.

The quality standards are necessary to:

- help determine the source of quality problems;
- set the framework for management and accountability for data quality;
- provide guidelines to eliminate problems;
- propose measures to prevent the occurrence of problems;
- assist in determining the priority of corrective measures;
- make decisions based on facts;
- realistically assess whether the solutions they achieve or exceed the intended purposes;
- allow easier exchange of data between data owners -public authorities;
- and other purposes.

2.2. Benefits

The benefits of quality data are related and affect the operation of the institution, other institutions, citizens and businesses:
- Obtaining timely and quality service, via short administrative procedures
- An effective and quality data exchange
- Effective use of resources in the provision of services, with single entering and verifying data
- Ability to quickly obtain relevant reports
- Possession of the data that important information for better decisions would be extracted from, cost effective investments and taking action
- Satisfied customers, partners and employees
- Good reputation of institutions
- Decreased number of complaints proceedings, and wrongly made decisions

3. Scope of standards

The quality standards of data contain mandatory guidelines that should be applied in the creation of new electronic registers by state institutions. Existing electronic records must comply with the standards within three years after they take effect. Although the standards for quality of data are compulsory only for the government, they can be applied by other owners of electronic records.

The standards primarily refer to the data necessary for administrative services and procedures that are conducted and given to citizens and businesses, but the scope of their application also includes cases of keeping electronic records intended for purposes other than administrative services, such as providing: correct economic policy, public security, parliamentary / local elections etc., i.e. records kept in the proceedings of general public interest, registers in the business sector and so on.

Responsibility for the implementation of standards has top management of the institution/ organization together with the heads of IT sectors, primarily for prevention, through: training, monitoring and good planning procurement, detection and discovery, through analysis and reporting.

4. Categories of stored data, definitions and descriptions

In the process of entering and processing data, the competent institution may use data from its own records and / or data from other institutions' records. Each registry consists of a series of records for entities, and the following information is kept for the same:
- Name of register
- Storage (only electronic, just paper and electronic and paper)
- A legal basis for storage (statutory act, Number of the Official Gazette, article)
- Period of data storage

Each entry consists of a collection of data specified in the law, bylaw or other act, and can be divided into three categories:
1. The data used to identify the individual entity
2. The data for which accuracy, diligence and integrity is responsible and competent given institution
3. Additional data used to further define the entity and issuing full service

Besides the three above mentioned groups of data, institutions can also hold data on a history of changes for individual entity / entities. If the history of data is regulated in the primary legislation that defines the service and is part of the registry or the institution, as a Service it gives information on previous condition, then the data has the same quality requirements as the data from Category 2 (or Category 1) . History data can be stored to meet other legal requirements (such as records of access to data under the Law on Data Protection) and in this case, it is governed by the legislation.

The rules for the practical application of the standards are set out in Appendix 1 of this document.

4.1. Data used to identify individual entity/ subject

In order reference and quickly access the requested record, the records contain necessary data in a unique way to identify the entity for which data is collected and stored. For this it is necessary to clearly label the identifiers and ensure their uniqueness.

These data may be one data, in cases where a unique and unrepeatable, but can also be set or a collection of several data, but they must be defined in advance in the event of data exchange between institutions in an official capacity.

Examples: ID number; CRN; registration number of the vehicle; cadastral parcel; and a set / collection: name, surname, address and phone mobile number.

4.2. Data where an institution is accountable and responsible for its accuracy, diligence and integrity

These data are data for whose generation, entry, editing and storage is responsible one institution, and where the parties need to turn when they need them. The same will be requested only by the institution as responsible for them in case of disputable situations.

For example, data for reporting employee is solely responsible Employment Agency of RM; for granted citizenship of the Republic of Macedonia is solely responsible Ministry of interior; etc.

For this, it is necessary to establish a clear and unambiguous authority for all data from the legal point of view, in the direction where one data cannot be under responsibility of more than one institution, or is not the responsibility of any institution.

4.3. Additional data used to further define the entity and issuing full service

These data are not required as the data of first and second category, but help the institutions in providing better services.

For example, the address of the physical / legal entity, to be used for sending correspondence or any statistics and analyzes by municipalities; address of the electronic mailbox (email) and phone number for further communication; etc.

5. Nomenclatures and classifications, method of application and referencing

National classifications, by decision of the Government shall be adopted as a national standard and as such must be used by all state institutions and authorities in Macedonia that would also enable comparability of statistics at a national level.

Classifications can be defined as a set of discrete, detailed observations that are mutually exclusive and that can be assigned to one or more variables in order to compare and present data.

On the other hand, nomenclatures can be defined as the systematic naming of things according to certain principles and criteria that represent systems of names. In the classification, the nomenclature represents systematic naming of categories. The terms of classification and nomenclature are often replaced with one another even though the definition for classification is broader in terms of the definition of nomenclature. Nomenclature is essentially a convention for describing observations, whereas classification is a system / standard where according to certain criteria (features, level of value), objects, phenomena, concepts etc. are organized or allocated and coded by classes, groups, divisions, departments (classifies and codifies the observations).

The classifications are used for statistical purposes and represent the process of organizing information into categories or classes, so that data can clearly be analyzed or understood. They represent one of the most basic tools in the process of data collection. Using the unique standardized classifications and nomenclatures provides a common and unified approach in the collection, classification and presentation of statistical data, and also provides one of the main goals of Statistics of the Republic of Macedonia, which is to produce comparable statistics as of European and international level. Therefore, the State Statistical Office of the Republic of Macedonia as an institution responsible for producing official statistics and coordinator of the statistical system in the country uses a series of national classifications which are harmonized with the standards and classifications in EU as well as with the international classifications.

The production of comparable statistical data with other countries or on international level inevitably requires the review of the existing classification system and harmonization of the statistical system of the country with the statistical systems of countries of the European Union as well as regularly reviewing and adapting the classifications to constantly changing social and economic conditions.

6. Quality data features

The following eight characteristics of quality data are determined for the purposes of the scope of these standards:

(Accuracy):
Data should be sufficiently accurate for the purpose they are intended for and they should be collected only once, although they can be used for multiple purposes.
Data should be collected at the point of activity, that is, at the source of the data, with precise and proven methodologies / techniques.
The system should contain basic data whenever it's possible, such as the numerator and denominator, and the result should be calculated from the system. This will eliminate errors in calculation and will provide contextual information to users.

(Validity):
The data should be recorded and used in accordance with the relevant requirements, including the correct application of all rules or definitions.

(Reliability):
Data should be a product of stable and consistent process of data collection by the collection points in time. Progress towards the goals should reflect real changes, not variations resulting in different approaches or methods for data collection.
The sources of data within the institution (category 2) must be clearly identified and easily accessible, whether the source is electronic or paper.
Whenever the data is received from third parties (category 3) there should be protocols by which it is regulated.

(Timeliness):
Data should be collected as soon as possible after the event or activity and should be available for the purpose for which they are intended in a reasonable period of time.
Data should be available quickly and often enough to provide the necessary information and support for management decisions. This will improve access to information and will eliminate the delay in the publication of information.
For instance, reports with quarterly data should be available no later than one month after the end of the quarter to prepare summary reports and plans for the next quarter.

(Completeness):
Data requirements should be clearly specified based on the information needs of the institution.
The process for collecting data should also match these requirements.
Regular inspection for missing values or for those which are in an unusable state is also required.
Although in some cases missing data is unimportant, but when missing data are critical for certain processes, completeness becomes significant.

(Relevance):
Collected data should be appropriate for the purposes they are intended for. This requires a periodic review of requirements to comply with changes in the needs.

(Consistency):

In the case of two different requirements with a unique identifier, for example by two different systems, it is necessary to obtain the same information on the same basic entity. This requires periodic checkups.

(Conformity):

In some situations, the values of the data should be in accordance with specific standards or formats for metadata. This requires periodic verification of the consistency of the data with formats / standards.

7.The process of managing data quality within an organization:

The process of managing data quality is achieved through iterative repetition of the four main stages, and describes the procedures and identifies the persons responsible for the processes for managing data quality within an organization.

The dynamics of the phases of the cycle and frequency of their repetition is determined by each institution separately, according to their needs.

The stages in the process are:
1. **Define**: Define the characteristics of your data
2. **Measure**: Measure the quality of your data before determining metrics
3. **Analyze**: Analyze and investigate the causes of the problems
4. **Improve**: Identify key areas for improvement and implement measures/ steps for improvement
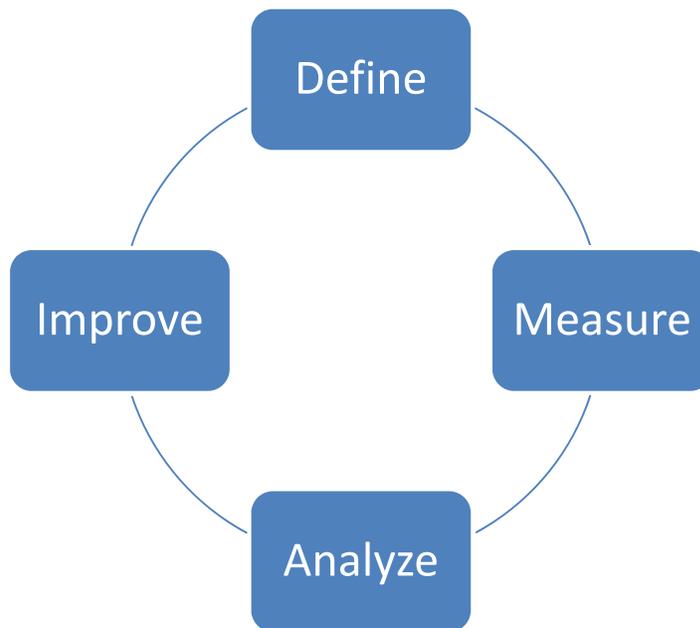
Some of the mechanisms for measuring, improving and maintaining quality data are listed in Appendix 3 of this document.

Appendix 1: Data

Data title	Authorized institution	Data type	Allowed signs	Formats and/or Controls[5] of the data itself[6]
Name	Civil Registry Office	Text, maximum 255 characters	А..Ш а..ш - '	No rules (Law on personal name)
Surname	Civil Registry Office	Text, maximum 255 characters	А..Ш а..ш - '	No rules (Law on personal name)
Unique personal ID (ЕМБГ)	Ministry of Internal Affairs	13 digit number	0..9	13 digits are grouped into six groups: - day of birth (2 digits) - month of birth (2 digits) - year of birth (3 digits) - registration area (2 digits) - combination of gender and the ordinal number for persons born on the same date (3 digits) for men from 000 to 499, and for women from 500 to 999 and - control number (1 digit) the control digit is according to Module 11
Company name	Central Registrar of RM	Text, maximum 2000 characters	no restrictions, including numbers, punctuation marks and other special characters and Latin letters	there cannot be two subjects with same title
Gender	Civil Registry Office	3 values		male gender (0) female gender (1) undefined (2)
Data regarding Addresses, in compliance with the Law on	Central Registrar of RM	Structure comprised of: (name of infrastructure facility (street, square, bridge) or other infrastructure	The name of the street, squares, bridges and other infrastructural facility is written in Macedonian	For the name of the infrastructure facility: In one settlement, there cannot be two or more streets with the same name. If the street name is determined only by number

---

[5]Code list
[6]Not on the entry, as for ex: in case of date, is it after or before something else;

Data title	Authorized institution	Data type	Allowed signs	Formats and/or Controls[5] of the data itself[6]
determining street, square, bridges names and other infrastructure facility		facility (shopping center, underground passage, ground passage, park and fountain) - text with maximum 100 characters, identification number of the infrastructure facility (integer), number of object (house / building) - (text with maximum 25 characters), identification number of the facility (integer), City and Municipality are written to those data (to determine the origin of the infrastructure facility)	language and its Cyrillic alphabet. In municipalities where at least 20% of citizens use an official language other than Macedonian language, the name of the street, squares, bridges and other infrastructure facility, in addition to Macedonian language and its Cyrillic alphabet, is written in the language and alphabet used by at least 20% of citizens in that municipality. The name can be determined only by a number. Number of object characters allowed [0-9] [А-Ш] For Personal identification numbers - numbers [0-9], the first digit 5 (five) for personal identification number of the infrastructure facility	there cannot be two or more streets with the same number. The street name cannot have more than four words. For the numbers of facilities: two or more objects on the same street cannot be numbered with the same number. For the personal registration numbers: last digit according to Module 11 of the previous seven (7) that are increases by 1 with each new generation
Telephone number	ITU standard[7]	Text	0..9 ( ) + празно место	National numbers: (xx) xxx xxxx Internaltional numbers: +xxx xx xxx xxxx
Unique tax number (ЕДБ)	Public Revenue Office	13 digit number	0..9	Control digit is according to Module 11
Unique Personal health care insurance	Health Insurance Fund	text with 9 characters	0..9	Luhn algorithm

---

[7] https://www.itu.int/rec/T-REC-E.123-200102-I/en, 30.12.2015

Data title	Authorized institution	Data type	Allowed signs	Formats and/or Controls[5] of the data itself[6]
number (Е3БО)				
Unique company ID (ЕМБС)	Central Registrar of RM	Integer	0..9	The last digit in Module 11 of the previous six (6) that increases by 1 with each new generation
City	State Statistical Office	text	text	Nomenclature of Territorial Units - NUTS is based on the territorial organization of local self-government in the Republic of Macedonia and is in accordance with the classification of the European Union Nomenclature of Territorial Units for Statistics - NUTS, established by Regulation 1059/2003.
Municipality	State Statistical Office	text	text	Nomenclature of Territorial Units - NUTS is based on the territorial organization of local self-government in the Republic of Macedonia and is in accordance with the classification of the European Union Nomenclature of Territorial Units for Statistics - NUTS, established by Regulation 1059/2003.
Citizenship	Ministry of Internal Affairs	text	text	logical controls Law on Citizenship of the Republic of Macedonia
Education	State Statistical Office	text	text	International Standard Classification of Education - MSKOB (ISCED 2011) is a national standard, adopted by a decision of the Government (Decision on reduction of system data on education in the Republic of Macedonia with the International Standard Classification of Education - ISCED 2011 "Official Gazette of RM "No.177 / 2013).
Activity	State Statistical Office	code - 4 digit number	0..9	National Classification of Activities - NKD - Rev. 2 was adopted as a national

Data title	Authorized institution	Data type	Allowed signs	Formats and/or Controls[5] of the data itself[6]
				standard based on the decision of the Government of the Republic of Macedonia ("Official Gazette of RM" No. 147/08) and the Law on One-Stop Shop System and the Commercial Register and Register for other legal entities ("Official Gazette of RM" No.84/05)
Settlement	State Statistical Office	code - 6 digit number	0..9	Nomenclature of Territorial Units - NUTS is based on the territorial organization of local self-government in the Republic of Macedonia and in accordance with the classification of the European Union Nomenclature of Territorial Units for Statistics - NUTS, established by Regulation 1059/2003.

Appendix 2: Encoding and transliteration

Suggested encoding for data is UTF-8 or UTF-16 (Unicode). Data recorded using YUSCII fonts (such as MAC C Times, Mac C Swiss, MacedonianTimes etc.) will be considered poor-quality data.

If some of the data is recorded with two alphabets (Cyrillic and Latin), recommended transliteration is the one used by the Ministry of Interior when issuing personal documents. It is shown in Table 1 below:

Table 1: Transliteration:

Nr.	Macedonian Cyrillic	ICAO 9303
1	А	A
2	Б	B
3	Ц	C
4	Д	D
5	Е	E
6	Ф	F
7	Г	G
8	Х	H
9	И	I
10	Ј	J
11	К	K
12	Л	L
13	М	M
14	Н	N
15	О	O
16	П	P
17	Р	R
18	С	S
19	Т	T
20	У	U
21	В	V
22	З	Z
23	Џ	DJ
24	Ќ	KJ
25	Ч	CH
26	Ѓ	GJ
27	Ш	SH
28	Њ	NJ
29	Љ	LJ
30	Ж	ZH
31	Ѕ	DZ

Appendix 3: Mechanisms

In order to treat each data as quality data, the same should go through several documents so it can be validated and verified.

Validation represents the first filter that checks predefined logical rules and whether the data fulfill their specification. Rules can apply to the format, length or content of the input data. The verification aims at determining the truthfulness of the data, which includes operations on the external or internal databases. A face can also be turned on In the process of verification, as opposed to strict validation that is executed by a computer system. In client-server architecture, validation can also be executed on the client side, while the verification of data should always be done by the server.

In addition, for increasing and maintaining the level of quality of data, mechanisms for identification and correction of errors and the method of reporting the same should also be defined. In the same direction, but from a security perspective, we need to define restrictions and restricting access to the data if necessary.

A) Validation during imputing data

Validation of data is the process of checking whether they are in compliance with their specification. It is the first process that takes input data. Examples of checks that can be performed are: the number of characters and type of characters, data type, range of values, consistency between the data and the other one in the same record and more. It requires the existence of procedures and procedures for entering or updating data by limiting the input format and validation. The validation method during imputing data needs to be documented. The main goal is the first level of control and accuracy of verification data.

Mechanisms for verification and quality assurance of the data entered depend on the situation. Examples of mechanisms:

- Automatic validation of text data: check the maximum number of characters, required data (must have value NULL), acceptable or mandatory signs (Address of the electronic mailbox (email), password, etc.), delete unnecessary blanks at the beginning or end of the text, use upper and lower case;
- Automatic dates checking: check if date entered is not active before the date required, obligatory field, checking the date based on business logic, checking the "from" and "to" dates;
- Automatic checking of numerous data: Mandatory data acceptable values between the smallest and largest defined value, verification of all or a decimal number, the sign of the numerical value (+/-)
- Automatic data validation module for 11 (ID number, registration number, TIN)
- Using national classifications of static data. Example address books (municipalities, settlements, infrastructure facilities), organizational forms;
- Use of the lists for selection of values in the case of input data belonging to finally set of known values;
- Multiple input (retype) and comparison of equality.

Validation procedures must be documented with main focus and insight into the following factors:
- Who is responsible for the validation procedure?
- Which information is subject to validation?
- What work practices, mechanisms are used for validation?
- In which part of the data managing process validation is done?

B) Verification

Verification of data is a process wherein the different types of data are checked for accuracy as a result of checking the data against the other data source or data set. Verification helps to determine whether the data is correctly transferred from one source to another. The main objective is to verify and determine the correct data entry.

Mechanisms for implementation of the verification process are:

- Benchmark with:
  - other electronic databases (internal or external), paper documents,
  - an on-site insight;
- Manual checking of second instance -
  - person responsible for checking has not been involved in data entry (principle of 4 eyes)
- Print a copy of the input data which is given to the client who gave the data, to check the data before entering the system;

The verification of the data is usually determined through review and verification of the following factors:

- Is there a need for verification?
- Who is responsible for verification procedures?
- What data should be subject to verification?
- On which data verification is done on?
- What methodology is used for verification?

C) Identification and reporting errors

There is a need for processes and procedures in the process of quality control data enabling identification of errors, and specifying the method of reporting errors. The general idea of achieving the identification and correction of errors is to add redundancy in the sense that recipients can check the consistency of the data they receive. Reporting errors can be by the staff, other government institutions, business sector and citizens. Bug reporting occurs after the identification of errors, prior to correction of errors, which finally allows the reconstruction of the original input data.

Developing awareness of errors is key factor to predict with high probability where an error can occur through collecting experiences on where and how errors occur as well as through analysis of the issues.

Mechanisms for identifying and reporting errors are:
- Reporting errors in electronic form, paper or call center (with or without a Form for reporting errors);
- Automatic (system) error detection;
- Manual (Manual) testing of errors;
- Data audit to detect anomalies, their features and contradictions.

The process of identification and reporting of errors should be documented and is usually determined through review and verification of the following factors:
- Are possible errors provided and classified (in terms of data / user dependent / independent of context[8]?
- Who can identify and report bugs: citizen, business, and officer of the institution or another institution, system?
- How and where to report bugs?
- Who is responsible for receiving reports about errors and handling them?
- When you report errors, while ensuring that the parties concerned will be notified of the action taken?

D) Correction of errors

There should be procedures according to which all identified and reported errors will be corrected. The process of error correction is a process enabling correction of data which do not satisfy the criteria or standard for quality data.

However the correction and filtering of data in some cases may be severely limited because of the phase of standardization in case they cannot amend records, the only way is if they are erased and that can lead to loss of unwanted entries.

Identification, reporting and correction of errors as processes are part of the process of improving the quality of data, which requires organizational approach, only by authorized persons.

Mechanisms for correcting errors are:
- Redirect to the source of the error;
- Delete records containing errors with parallel archiving, then allowing the new data to be added (insert);
- Directly update the data;
- De-duplication of data, or deleting the duplicate values, leaving only one value.

The process of correction of errors is usually determined through review and verification of the following factors:
- Who makes the correction of errors?
- Is there an authorization of persons responsible for correcting errors?
- How will you make the correction of errors?
- When will you make the correction of errors?
- When and how stakeholders will be notified of adjustments made?

E) Control of access to data

Access control, entering and updating data is a basic functionality each system for data processing must possess. The main goal is to protect data from unauthorized operations of reading and writing.

It is necessary to keep data that allows monitoring of changes to individual entity / entity, i.e. history.

Mechanisms used for controlling access to data are:
- Customized orders for access;
- Access with biometric data;
- Authorization for the "user roles". Example: "shop worker", "Web - external user", "Responsible person", "Administrator", and others;
- Creating log review and change on level of the application and on level of the base.

Access to data is usually determined through review and verification of the following factors:
- Who and why you should have access to data?
- What activities should and can be performed on the data?
- What information should be made available to users?
- How can the data be accessed?
- At what time of day or week should they be available?