International Journal of Computer Engineering and Sciences Research

VOL. 04, NO. 03, May-June 2022
Pages 01–13 (ISSN: 2581-8481)
Available online at: www.ijcesr.com

E-Services Evaluation And Delivery Model Using Data Cleansing Logical Constraints – The Case Of Kosovo Portal

Genc Hamzaj
Faculty of Contemporary Sciences and Technologies
South East European University
Tetovo, Republic of North Macedonia

Zamir Dika
Faculty of Contemporary Sciences and Technologies
South East European University
Tetovo, Republic of North Macedonia

Isak Shabani
Faculty of Electrical and Computer Engineering
University of Prishtina "Hasan Prishtina"
Prishtina, Republic of Kosovo

Abstract: Evaluating e-Services in Kosovo Government portal from the perspective of data quality as an important precondition for quality of services represents the main objective in this study.After the war in 1999, Republic of Kosovo created new electronic registers for registering data of citizens. These data were registered without any criteria or dimensions concerning quality of the data. Assessing and improving the quality of these datasets before delivering e-Services becomes difficult task because of having big volume of data and the diversity of data sources each having different structure and with low quality.Through this paper we are evaluating e-Services provided by Kosovo government portal with the prerequisite of defining optimal data quality dimensions, for the datasets of personal and vehicle registration data. We will analyze quality of data in specified datasets using data quality dimensions such are Accuracy, Completeness, Consistency, Uniqueness and Timeliness.Through this paper we will also assess the impact of implementation of data quality dimensions and standards for data quality for providing qualitative data, easy to use in a single government portal with the aim to give better G2C (Government to Citizens) services. Also, we will show a model for integrating of e-services using microservice architecture in Kosovo government portal.

Keywords—*e-Services; e-Kosovo Portal; Data Quality Dimensions; Domain Driven Design;microservices*

I. INTRODUCTION

Achieving high data quality represents a prerequisite in effective administration with e-Services in Government portals that retrieve data from multiple sources of data. Also, by achieving high data quality, it affects the government to define a very effective strategy and business model for decision making in order to deliver services

in one stop shop. Through Government e-Services it is possible to offer online channel for providing public services in real time 24 hours a day, 7 days a week. Not achieving to provide adequate and qualitative data for the services provided by government can negatively affect in citizen trust and be reflected with a lack of citizens satisfaction.

E-Services within the public sector are applications which enable citizens to consume online available services [1].

The integrated e-Services process, is focused on joining records from disparate data sources describing the same real world entity and offering services through Government Portal.

With aim to secure efficiency of public services, implementing e-Government for offering services to citizens has been justified by its transparency, speed and accountability [2].

Achieving high level of quality of the data can be defined as data fit for use by citizens by accessing e-Services in Government Portals [3]. Proposing to use electronic services in public sector institutions is currently significant theme in research on information systems and public administration [4].

With the rapid development of information technology, governments try to enable citizens and enterprises to access government services and information efficiently and effectively, as well as improving transparency, accountability [5].

E-Government services include access to government information, access to government forms and services, release of policy information, employment and business opportunities and election information [6].

At worldwide level, some successful models concerning e-government were adopted in terms of their ability to fulfill citizen's satisfaction by reducing bureaucratic transactions within the public departments as well as saving costs and resources [7].

The main focus of this paper will be in assessing and improving:

- Data quality through defining dimensions.
- System quality through improving electronic systems.
- E-Services quality through Government portal.

Poor e-Services because of different factors including poor data quality and system quality results in customer dissatisfaction, reduced financial income and higher costs associated with the time and additional resources needed to reconcile the service.

In this research we aim to provide a model of Government e-Services delivery with improved data quality of multiple data sources by using appropriate data quality dimensions, microservices and adequate method to integrate services in Kosovo Government Portal e-Kosovo.

Through this approach, we will give a guide how to choose appropriate model for integrating and delivering e-Services through Government Portal for services like applying for personal documents and vehicle registration service.

II. METHODOLOGY AND CHALLENGES IN DATA QUALITY IMPROVEMENT IN IMPOSING E-SERVICES FOR KOSOVO GOVERNMENT INSTITUTIONS

Data quality is a measure indicating the level to which data is suitable for use by data consumers. Assessment of quality of the data represents the process of evaluating the data in with the aim to decide whether it meets the qualities required for the purpose of business processes. Measuring the quality of data is like a prerequisite for offering better e-Services in Government institutions.

The number of obstacles and issues that may restrict the full implementation of governmental e-services are listed below [8]:

- Speed of transformation - transformation requires passing a number of phases and usually the speed of transformation from one phase to another is slow for many reasons such as the employees' resistance for change,

inefficient budget, legislations and legal issues, moral issues, and citizens concern regarding coasts, subscriptions' services, awareness and training.

- Implementing Government to Citizens (G2C) transactions - this category of e- services includes more complicated and sizable number of transactions compared to employees, business and government's categories.

- Privacy and Security issues - governmental departments should adequately consider such issues that entail personal information (for example the medical history of patients) by using effective techniques such as coding, firewalls, and electronic signatures.

A. Logical Constrains To Ensure Data Quality In The Personal Documents Register (PDR) And Vehicle Register (VR)

Logical constraints also called data quality dimensions represent attributes of data quality of the data that, if measured and used correctly, can indicate the overall level of data quality. The process of identifying the relevant data quality dimensions presents an important step when attempting to offer e-Services in Government Institutions [9]. Due to the highly dependence of the dimensions on their context, their relevance and importance may vary between organizations and data types.

After the war 1999 Kosovo government started creating source register from the beginning with the aim to register all requests made by citizens in different platforms such are Microsoft Excel, Access, without concerning for quality of data.

The data stored in personal documents register (PDR) and Vehicle Register (VR) are core part of many business operations of others institutions and agencies. Therefore, the quality management of the PDR and VR will determine the quality of the services that public and business organizations provide to their customers.

Data quality assurance is a process of analyzing the data to discover errors and other anomalies in the data and to perform the activities in order to improve data quality. It is very important that the application system for the registration of personal data and vehicle data prevent the input and storage of the incorrect data in the PDR and VR. Therefore, it is recommended that the rules proposed below be incorporated into application systems as logical constraints.

The main focus of this paper will be in assessing and improving:

- Accuracy
- Completeness
- Consistency
- Uniqueness
- Timeliness

The accuracy is the extent to which data are correct, reliable and certified [10].The accuracy is a measurement of the precision of data showing how the information stored in the database corresponds to reality. It specifies whether the data values stored in databases for a specific object are the correct values.

In order to ensure the accuracy of personal documents data and vehicle registration data, the PDR and VR must regularly, once a month, check the database for the persons and vehicles for below data quality indicators:

- List of persons whose date of marriage is later than the date of death
- List of persons whose age difference of children with their parents is less than 13 years
- List of persons who have several spouses
- List of persons who have more than 20 children
- List of persons who have more than two parents in the parents list
- List of persons whose children have been assigned to other parents

- List of persons whose parents have been assigned other children
- List of registrations where the registration date is later than the expiration date
- List of registrations where the validity of the registration is more than one year
- List of owners who have more than 10 vehicles
- List of vehicles that have more than one active owner
- List of vehicles that do not have an active owner.

When such persons have been identified, actions must be taken to correct their data in the PDR database.

The completeness is defined as the degree to which a given data collection includes data describing the corresponding set of real-world objects [11]. The completeness dimension for data quality is to measure the existence of required attributes of the data in the records of the database. Completeness of data refers to whether there are any missing data that should be collected.

In order to ensure the completeness of personal documents data in the PDR and vehicle data in VR, the database should include regular, once a month checks for the persons and vehicles for below data quality indicators:

- List of persons without personal name and/or surname
- List of persons without personal identification number
- List of persons without date of the birth
- List of persons without gender
- List of persons without citizenship
- List of persons without municipality of dwelling
- List of vehicles that do not have license plate numbers
- List of vehicles that do not have a chassis number
- List of personal or business owners that do not have unique ID
- List of vehicles that do not have a registration or expiration date
- List of vehicles that do not have completed the model and country of origin
- List of vehicles that have been registered without municipality of registration.

When such persons are identified, actions must be taken to correct their data in the PDR and VR databases.

Data consistency refers to whether the logical relationship between correlated data is correct and complete [12]. The consistency of data ensures that all application systems equally understand and interpret the meaning of the data, there are no internal conflicts between the meanings of the same data. It ensures that data will be stored in one country-wide agreed format and definitions.

In order to ensure the consistency of personal documents data in the PDR and vehicle data in VR, the database should include regular, once a month checks for the incorrect format or structure of the data for below data quality indicators:

- Names and/or surnames contain invalid characters such as: figures, distinctive or arithmetic signs, etc.
- Invalid format of the date of the birth, death, married or dwelling – the date is stored in a different format than is intended use in Kosovo
- Data of the date of birth, death, marriage or dwelling refers to the non-existing time
- The personal number does not comply with the established rules

- The data of the place of birth or residence does not comply with the established rules or refers to the non-existing municipality, city, street or address
- The code of gender, citizenship or marital status does not comply with the established rules applied
- The ID of municipality of birth, the ID of the place of birth and the data of municipality and place of birth refer to different municipalities and addresses
- List of records that have invalid characters in the Tag Number and chassis Number fields such as: figures, distinctive or arithmetic signs
- List of records that have invalid formats in the Tag Number field
- List of registrations that have invalid formats on the date of registration, expiration
- List of registrations which refer to the municipality, type of registration, owner with non-existent ID.

When such data are identified, the actions must be taken to correct this data in the PDR and VR databases.

Uniqueness dimension of data refers to a singularity of data or records. The same facts cannot be recorded and saved multiple times. The database cannot contain any duplicated records - a data record with specific details appears only once in the database.

In order to ensure the uniqueness of civil personal documents in the PDR and vehicle data in VR, the database should include regular, once a month, checks for the valid (non-historical) registration for below data quality indicators:

- List of several persons with the same personal number
- List of several places and/or dates of the birth of the person
- List of several places of the resident of the person
- List of several vehicles with the same chassis number or active license plates
- List of owners have registered the same car at the same time
- List of owners or businesses registered several times and have active vehicles.

When such records are identified, the actions must be taken to correct those data in the PDR and VR databases.

The timeliness reflects how well data in the database represents the current situation and whether the data is functionally available when needed. Late data entries and updates may negatively affect data analysis and reporting, as well as decision making processes.

In order to ensure the completeness of data in the PDR and vehicle data in VR, the databases should include regular, once a month, and checks for the persons for below data quality indicators:

- List of persons who are more than 110 years old
- List of persons who have invalid personal documents and the death was not registered
- List of persons without the registered place of residence
- List of vehicles that have not been registered in the last 10 years
- List of vehicles where the owners are over 90 years old
- List of vehicles that are older than 30 years and have not been registered as oldtimers.

Where such persons are identified, actions must be taken to correct their data in the PDR and VR databases.

B. Outcomes From Assessment And Improvement Of Data Quality On Source Electronic Registers

Outcomes from assessment and improvement of data quality are generated using a specific assessment methodology that consists of a sequence of activities for assess and improve of quality of data in the dataset.

Automatic quality assessment and improvement of data as a prerequisite is used. Different software queries over the datasets are executed in order to assess data that fulfils 5 most important dimensions.

The purpose of the automatic Data Quality Assessment and improvement is to measure the quality of data collections along most important data quality dimensions.

In order to assess and improve the accuracy, completeness and consistency, uniqueness and timeliness of the data kept in the critical database fields that were identified as important, a number of program scripts (SQL queries) were developed and execute on the datasets, both PDR and VR.

The identified critical database fields as well as the results from the scripts that were executed are given in table 1 with the aim to assess quality of data through data quality dimensions like accuracy, completeness, consistency, uniqueness and timeliness.

The findings from executed scripts in PDR and VR regarding to accuracy dimension are shown in the following table:

TABLE I. FINDINGS FROM EXECUTED SCRIPTS (ACCURACY).

The accuracy of the data	**Value**
List of persons whose date of marriage is later than the date of death	21
List of persons whose age difference with their parents is less than 13 years	226
List of persons who have several spouses	0
List of persons who have more than 20 children	1
List of persons who have more than two parents	0
List of persons whose children have been assigned other parents	10
List of persons whose parents have been assigned other children	0
List of registrations where the registration date is later than the expiration date	100
List of registrations where the validity of the registration is more than one year	88
List of physical owners who have more than 10 vehicles	92
List of vehicles that have more than one active owner	58
List of vehicles that do not have an active owner	32

The findings from executed scripts in PDR and VR regarding to completeness dimension are shown in the following table:

TABLE II. FINDINGS FROM EXECUTED SCRIPTS (COMPLETENESS).

The completeness of the data	**Value**
List of persons without personal name	74
List of persons without personal surname	100
List of persons without personal identification number	0
List of persons without date of the birth	0
List of persons without gender	10

The completeness of the data	Value
List of persons without citizenship	2
List of persons without municipality of dwelling	482
List of vehicles that do not have license plate numbers	100
List of vehicles that do not have a chassis number	58
List of personal or business owners that do not have unique ID	71
List of vehicles that do not have a registration or expiration date	125
List of vehicles that do not have completed the model and country of origin	78
List of vehicles that have been registered without municipality of registration	42

The findings from executed scripts in PDR and VR regarding to consistency dimension are shown in the following table:

TABLE III. FINDINGS FROM EXECUTED SCRIPTS (CONSISTENCY).

The consistency of the data	Value
Names and/or surnames contain invalid characters such as: figures, distinctive or arithmetic signs etc	42
Invalid format of the date of the birth, death, married or dwelling –the date is stored in a different format than is intended use	0
Data of the date of birth, death, marriage or dwelling refers to the non-existing time	37
The personal number does not comply with the established rules	25
The data of the place of birth or residence does not comply with the established rules or refers to the non-existing municipality, city, street or address	62
The code of gender, citizenship or marital status does not comply with the established rules	0
The ID of municipality of birth, the ID of the place of birth and the data of municipality and place of birth refer to different municipalities and addresses	10
List of records that have invalid characters in the Tag Number and chassis Number fields such as.; '	125
List of records that have invalid formats in the Tag Number field	10
List of registrations that have invalid formats on the date of registration, expiration	38
List of registrations which refer to the municipality, type of registration, owner with non-existent ID	47

The findings from executed scripts in PDR and VR regarding to uniqueness dimension are shown in the following table:

TABLE IV. FINDINGS FROM EXECUTED SCRIPTS (UNIQUENESS).

The uniqueness of the data	Value
List of several persons with the same personal number	21
List of several palaces and/or dates of the birth of the person	40
List of several palaces of the resident of the person	12
List of several vehicles with the same chassis number or active license plates	17
List of owners that have registered the same car at the same time	25
List of owners or businesses registered several times and have active vehicles	0

The findings from executed scripts in PDR and VR regarding to timeliness dimension are shown in the following table:

TABLE V. FINDINGS FROM EXECUTED SCRIPTS (TIMELINESS).

The timeliness of the data	Value
List of persons who are more than 110 years old	559
List of persons who have invalid personal documents and the death was not registered	410
List of persons without the registered place of residence	28
List of vehicles that have not been registered in the last 10 years	25
List of vehicles where the owners are over 90 years old	49
List of vehicles that are older than 30 years and have not been registered as oldtimers	32

Data quality assessment and improvement is done in datasets with millions of records each.

As we can see in the results table, the automatic software scripts identified a number of records in dataset with issues related to 5 most important data quality dimensions.

III. A MODEL FOR INTEGRATING E-SERVICES IN ELECTRONIC PLATFORMS – E-KOSOVO MODEL

When the process of assessing and improvement of the quality of data in the datasets, the aim or the goal is to use the appropriate model for integration developed e-Services in electronic platforms such is government portal of Kosovo.

Software architecture is the fundamental structure of a software that defines technical and operational requirements. This is the reason that choosing of appropriate architecture is very important in the primary phase of software development [13].

In nowadays the most popular solution regarding to Google Trends in using Microservice Architecture [14].

E-Kosovo is developed based on three-tier architecture with the latest methodologies and tools for developing complex systems. Performance is one of the most important parts of architecture. Those methodologies are used by big companies like Microsoft and Google.

The methodology that is used for developing this platform is the so-called “Domain Driven Design” (DDD) shown in figure 1 and the technique for developing applications in web “Model View Controller” (MVC). This architecture is an advanced version of “Multi-Tier Architecture”, which has been expanded to adapt with the new developing techniques, separation of components like Microservices and software testing.

Microservices are replacing monolithic architecture systems, narrowly focused on the distributed system while having an isolated service [15].

This architecture consists of the latest techniques/ properties like:

- Dependency Injection
- Repositories
- Interfaces
- Services

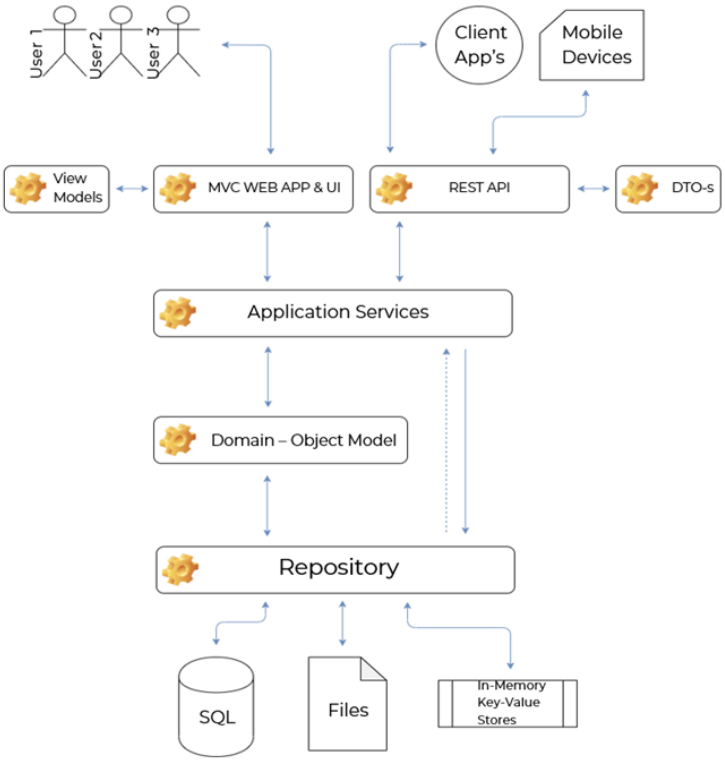

Fig. 1. Domain Driven Design

The technique/model (MVC) presented in figure 2 shows the model of building of web-based application in .NET Core/.NET Framework based in layers such are model, view and controller.

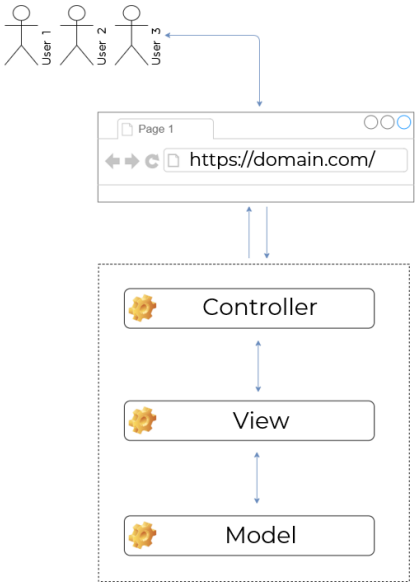

Fig. 2. Model View Controller

The physical structure of the integration of issuance and verification of documents from the electronic platform such is government portal of Kosovo is shown in figure 3 as below.

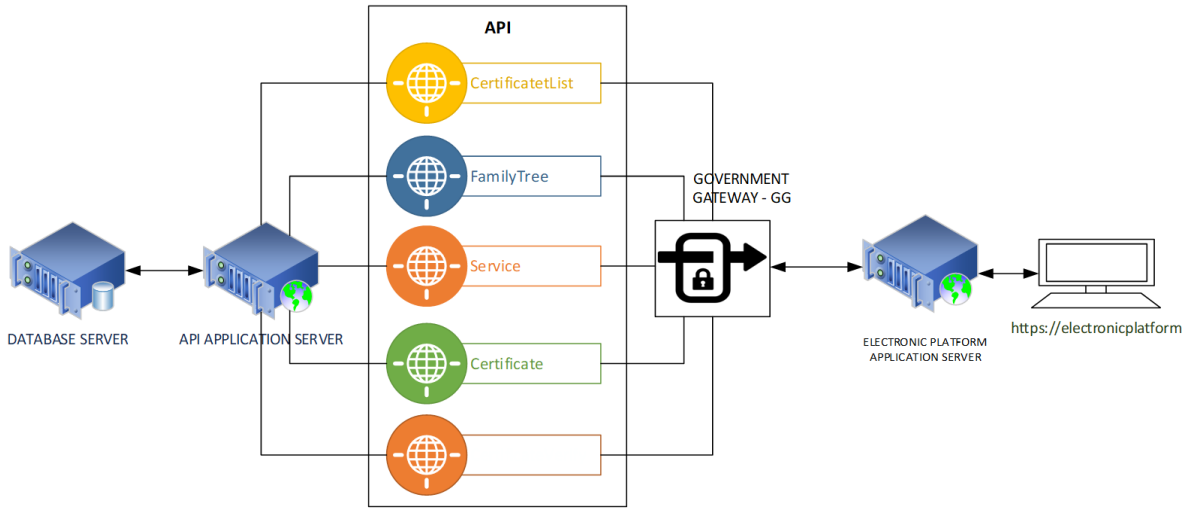

Fig. 3. Physical structure of the integration of issuance and verification services.

For the purposes of integrating of e-Services through government portals, several diagrams are created and functionalized.

We developed below diagram which we consider the most appropriate diagram when we deal with integrating services that treat issuing personal documents through government portals.

In figure 4 below is shown a model of flowchart for integration of e-Services in the electronic platform.

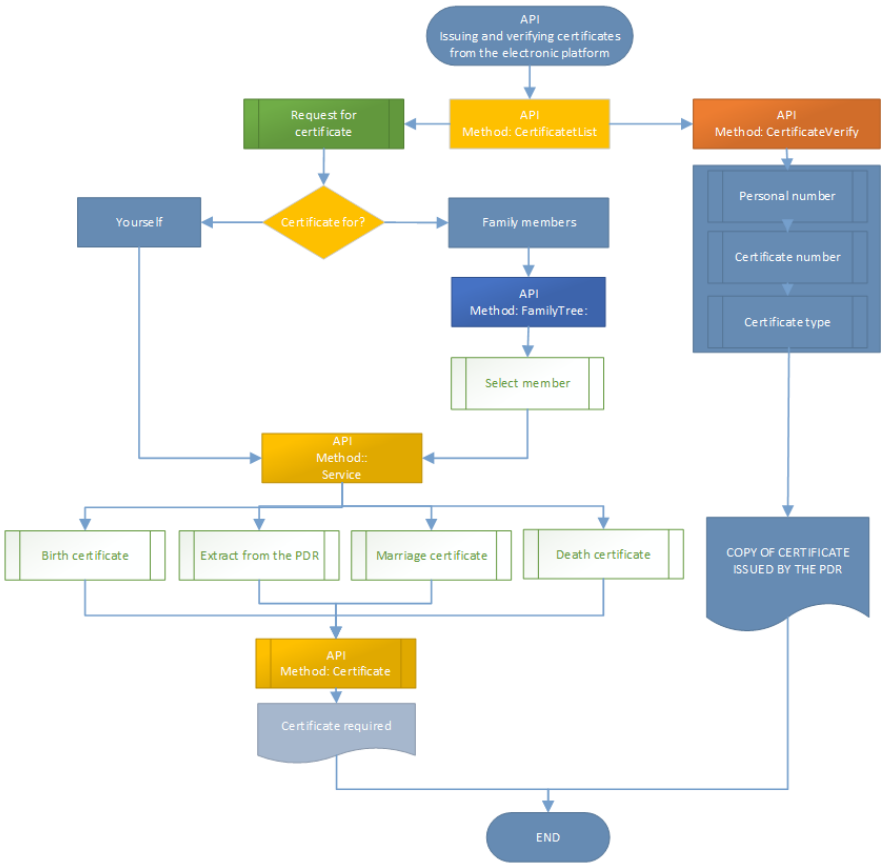

Fig. 4. Flowchart for integration of e-Services in the electronic platform.

Authentication through Token Based Authentication is based on a Signed Token that is sent from the Server through API to the Client in every potential request.

The standard that is used to implement Token Based Authentication is JWT.

This architecture consists of the latest techniques/ properties like:

- The client requests authentication in the endpoint of the API of the platform by entering User and Password
- API for authentication through Token Based Authentication does the verification of received credentials from the client by comparing them to those that are in the database
- In case of successful authentication in step #2 then API will return the token to the client. Token will have the needed information to identify the client and the validity time of the token. These information before being transmitted are signed and then through HTTPS are transmitted to the client
- The client after receiving the token from the server through API will store it in their device, all the future requests are done through the token, in this case the opportunity of user and password being exposed is avoided
- The server after receiving the token verifies that the token is valid and returns the needed resources, otherwise the client will have to continue from step #1.

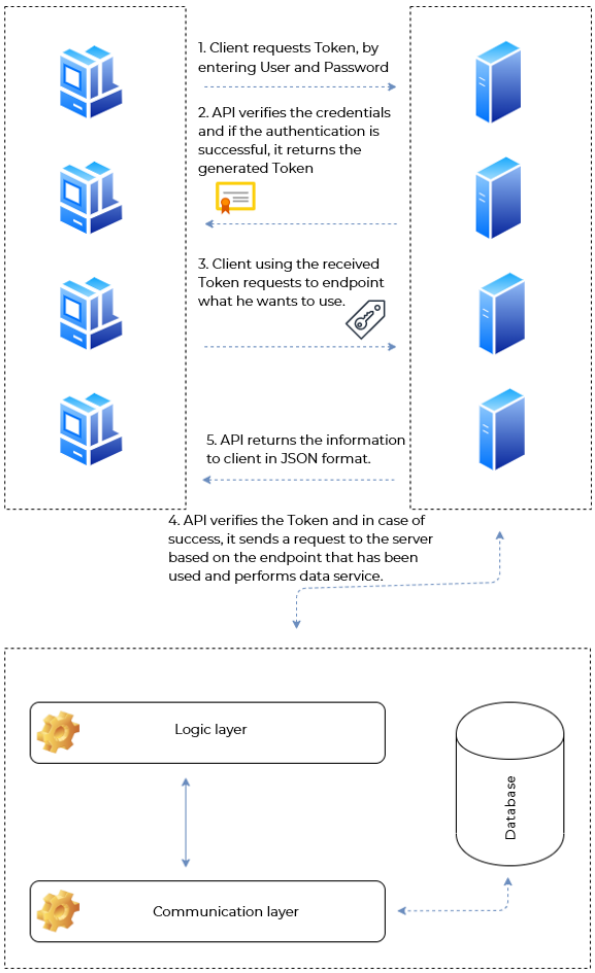

Fig. 5. Token Based Authentication.

The following script describes the API for issuing and verifying documents issued by the government portal.

Access to API methods is realized using the specified link where the input parameters and the expected result in jSon are, as follows:

Input parameters

{

"Language ID": "sample string 1" – Language Parameters: depends how many languages is used

}

Expected Result in jSon

[

{

"Document ID": "sample string 1",

"Description": "sample string 2",

"Issued By": "sample string 3", Depends from which platform the documents are issued True=Yes and False=No

"Verified": "sample string 4" Can be verified through electronic platform

},

```
    {
        "Document ID ": "sample string 1",
        "Description ": "sample string 2",
        "Issued By ": "sample string 3",
        "Verified ": "sample string 4"
    }
]
```


IV. CONCLUSION

In this paper we presented data quality assessment and improvement of personal records dataset and vehicle registration records through analysis of five data quality dimensions. Personal records of datasets are assessed through dimensions such are accuracy, completeness, timeliness, consistency and uniqueness.

After data quality assessment and improvement of the datasets then we implemented specific model for integrating of e-Services in electronic platforms such is government portal of Kosovo. We described physical structure of the integration of issuance and verification services, then we presented a flowchart for integration of e-Services in the electronic platform and also we showed implementation solution through jSon.

Through this paper we highlight the importance of assessment and improvement of data quality as a prerequisite for implementing e specific model for implementing of e-Services through government portal of Kosovo.

Also, all the results retrieved form this paper can guide stakeholders to offer professional and satisfactory e-Services through eliminating problems and obstacles from technical and operational perspective.

REFERENCES

[1] e-Government Portal, South Africa, https://www.eservices.gov.za/tonkana/home.jsf.

[2] J. Sánchez-Torres,and I. Miles, "The role of future-oriented technology analysis in e-government: A systematic review," European Journal of Futures Research, vol. 5, pp. 1–18, 2017.

[3] P. Christen, "Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution and Duplicate Detection," Springer Data Centric Systems and Applications, 2012.

[4] S. Dawes, "Governance in the digital age: A research and action framework for an uncertain future," Government Information Quarterly, vol. 26, pp. 257-264, 2009.

[5] S.K. Aikins, and D. Krane, "Are Public officials obstacles to citizen-centered e-government? An examination of municipal administrators' motivations and actions," State and Local Government Review, vol. 42, pp. 87-103, 2010.

[6] E. Larsen, And R. Lee, "The Rise of the e-citizen: How People Use Government Agencies' Web Sites,"https://www.pewresearch.org/internet/2002/04/03/the-rise-of-the-e-citizen-how-people-use-government-agencies-web-sites, 2002.

[7] Wesam Ibrahaim Mohammad Alabdallat, "Toward a mandatory public e-services," Department of Finance and Administration Sciences, Al-Balqa Applied University Al Salt-Jordan, As-Salt, Jordan.

[8] A. Al-Taher, "E-government: Between theory and practice," Amman: Araya Hall, 2009.

[9] R.Y. Wang andD.M. Strong, "Beyond accuracy: What data quality means to data consumers," Journal of Management Information Systems, vol. 12, pp. 5-33, 1996.

[10] C. Cichy, and S. Rass, "An overview of data quality frameworks," IEEE Access, vol. 5, pp. 24634-24648, 2019.

[11] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," ACM Computing Surveys, vol.41, 2009.

[12] L. Cai, andY. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," Data Science Journal, vol. 14, 2015.

[13] K. Gosand W. Zabierowski,"The Comparison of Microservice and Monolithic Architecture," IEEE XVIth International Conference on the Perspective Technologies and Methods in MEMS Design, pp. 150-153, 2010.

[14] Google Trends,https://trends.google.com/trends/explore?date=now%201-d&q=microservices, 2022.

[15] J. Stubbs, W. Moreira and R. Dooley, "Distributed systems of microservices using docker and serfnode," In Proceedings of the IEEE 2015 7th International Workshop on Science Gateways, pp. 34–39, 2017.