Interreg
Alpine Space

SMARTVillages
EUROPEAN REGIONAL DEVELOPMENT FUND

EUROPEAN UNION

Database and final database report

Work package: WP T2
Activity A.T2.4
Deliverable: D.T2.4.1

Database and final database report

Work package: WP T2
Activity A.T2.4
Deliverable: D.T2.4.1

Version: 1.0

Date: January, 2021

Authors:

Jorge Martinez-Gil[a], Mario Pichler[a]

----

[a] Software Competence Center Hagenberg GmbH

This is a report of the database and the backend that support the functionality of the DEP.

List of content

List of figures

1. Introduction

In the last times, the notion of smart cities and smart villages has gained a lot of popularity. There are already several methodologies and tools to help cities implement solutions in a wide range of domains, including traffic light optimization, improvement of urban solid waste treatment, and communication between vehicles to avoid traffic jams or to make more efficient use of parking space, energy saving in public and private buildings. However, due to their unique characteristics, rural territories are not suitable for implementing such solutions. This is mainly because the amount of data is not so voluminous, diverse or generated at such speed. Therefore, the nature of the information generated is very different and requires different solutions.

For this and other reasons, we have developed an online platform (Digital Exchange Platform, DEP) that can help rural territories to pilot a successful transition to smarter and more sustainable rural models. Some of the functionality that we have developed in this project is devoted to the deployment of such platform.

To do that, our platform is built around a set of tools that allow for analysis and monitoring of smartness indicators from the villages of the Alpine Space. This platform consists of a front-end and a back-end, as well as a set of intermediate tools that allow the analysis of the collected data.

The back-end is the part in charge of the storage of all the information that is necessary to operate, including the data of the smartness self-assessment forms, the good practices, the functionality to make possible a matchmaking operation with the Toolbox (a repository of tools available in a different server), etc. While the front-end is the part that is visible to the people and involves everything that users can interact with. Finally, the intermediate tools are designed to be lightweight. They can be connected to the data services of the platform, to analyze them and offer the result of that analysis to the users who require it.

This technical report focuses on the part related to the back-end. More specifically, on the database that supports the back-end. To implement such database, we use the relational database management system MySQL that contains all the information related to the users of the DEP. Therefore, the contributions of this technical report are as follows:

- We offer a detailed overview of the relational database that supports our online platform
- We offer a detailed overview of the tools we have developed to make the most of the data contained in the database
- We offer our point of view about possible future works and improvements that could be carried out.

The rest of this document is structured as follows: in Section 2, we present the technical details of the database implementing the back-end that supports our online platform. In Section 3, we present the technical details about the lightweight tools we have developed to make the most of the data contained in the database. Finally, in Section 4, we draw conclusions from this work and possible additional improvements that could be made as part of future work.

2. The Back-end

For supporting our back-end, we have decided to use a relational database which is a digital database based on the relational model. Some of the advantages of this kind of databases are the data accuracy, the easy access to data, integrity, flexibility, high security, and so on. One of the most important properties of our database is that users for further processing or analysis can export all the information contained. This information can be accessed through various micro services that we have enabled for this purpose.

The rest of this section is structured in the following way: we introduce the need of using a database of this kind, we explain the architecture that we have designed to act as a back-end, we describe the process of exporting data and we conclude with the legal aspects of the data we are handling.

2.1 Introduction

One of the main ideas behind our platform is that many stakeholders from different backgrounds can discuss and share their strategies to implement more sustainable models of digitization that can help them solve some of their most pressing specific problems. In this sense, it should be noted that although our platform is focused on the Alpine space, it is possible to register even if you do not belong to any of these countries.

We must store information about the processes that can be carried out from the front-end. For example, consider for a moment the process of smartness self-assessment. We need to store both some metadata (name of the village, country where it is located, kind of municipality, population as well as the age and profession of the person filling the questionnaire) and information related to the result in each of the smart categories (People, Governance, Living, Environment, Economy and Mobility) as can be seen in the following semi-structured code.

But we must also store information related to best practices, etc. All this is done through tables of the relational model. This way of proceeding also facilitates the consultation and access to data through the SQL language that is the de facto standard in the field of information systems.

{"metadata": {
"name": "string",
"country": "string",
"kind": "string",
"number_of_inhabitants": 0,
"assessor_age": "string",
"assessor_type": "string"
},
"smart_assessment": {
"smart_people": 0,
"smart_governance": 0,
"smart_living": 0,
"smart_environment": 0,
"smart_economy": 0,
"smart_mobility": 0 }}

However, all this information is processed in background, so users do not have to worry about. In addition, the collection of such information strictly complies with the requirements of data protection laws which state that information should not be collected that uniquely identifies an individual.

2.2 Architecture

The architecture we have decided on for our platform is the classic client-server architecture, where the server processes the requests of a user given through a client. It has a front-end, which is the visible side of the server, and a back-end which is where the data is stored for further processing. It is precisely in the back-end where the database of this project resides.

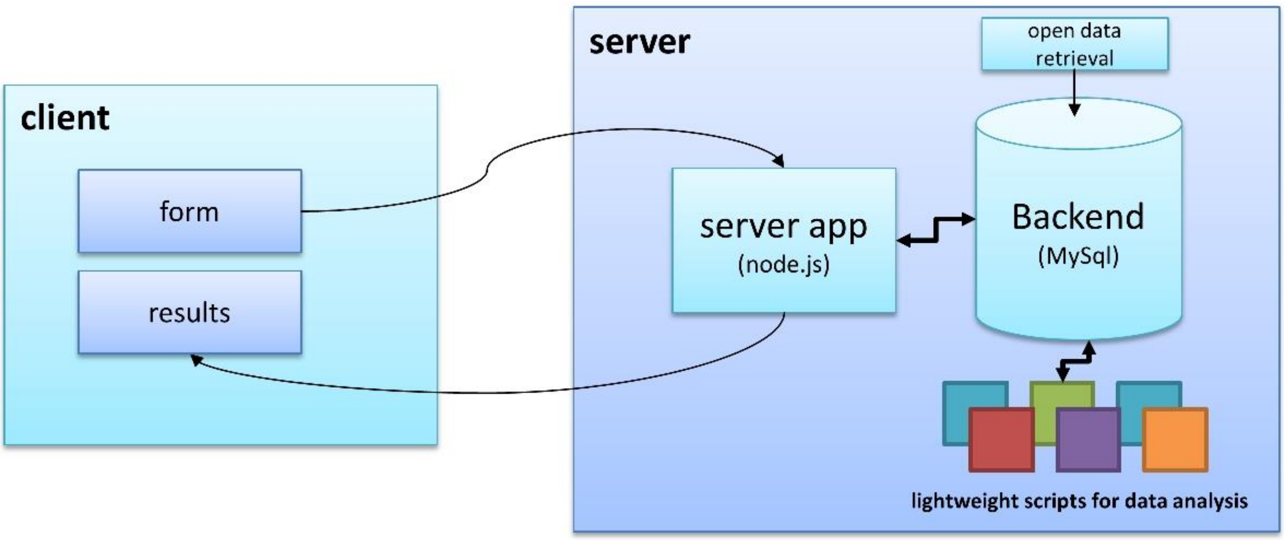

Figure 1. Architecture of our online platform

2.3 Exporting data

The data in the database are exported in raw form, so that they can be processed or analyzed by any person who wishes to do so. We have also built in some functionality that allows us to analyze such raw data. For example, we have functionality that allows us to calculate the rankings of the villages that have completed the smartness assessment. These rankings can be calculated in an absolute way or in a relative way to the desired dimension. In addition, we have developed functionality to calculate the similarity of the villages present in the database but from the smartness assessment results point of view.

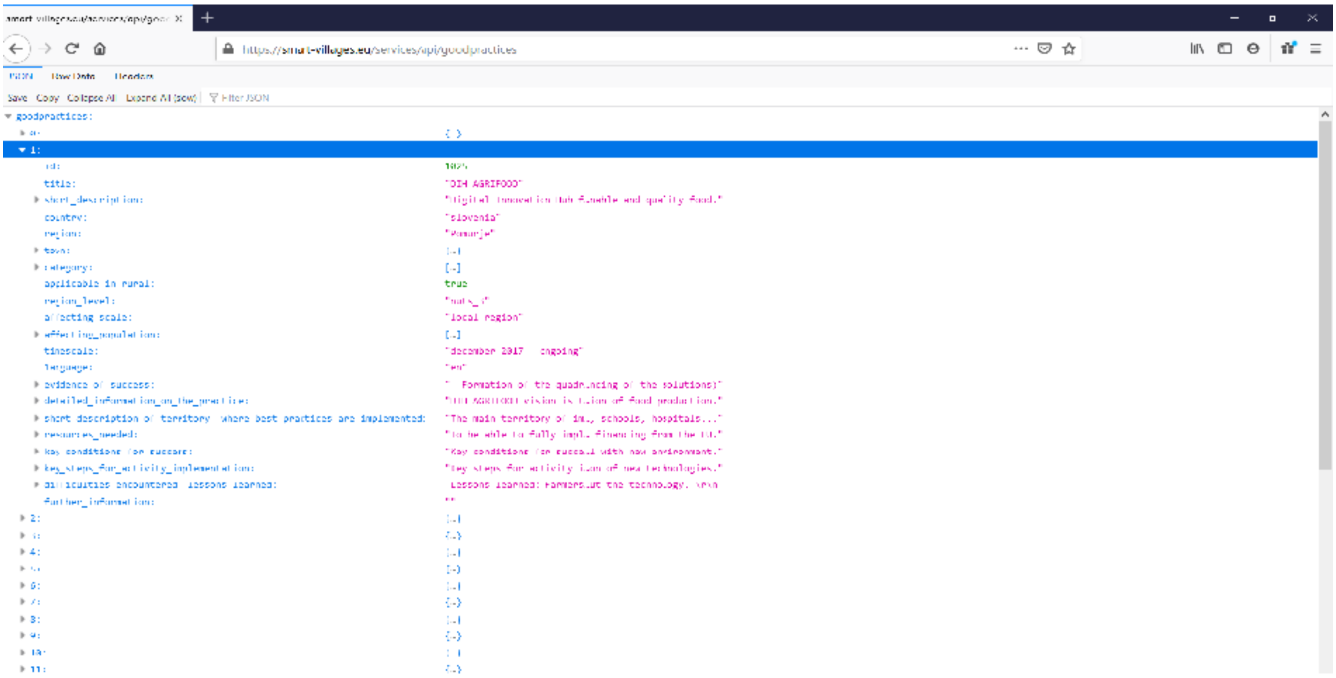

Figure 2. Screenshoot related to the access of the best practices from the web browser

The API points available from our DEP can be accessed through the following URL:
https://smart-villages.eu/language/en/export-data/

In fact, we offer three specific entries for consuming the data through micro services.

- One for data related to smartness self-assessment:
  https://smart-villages.eu/services/api/assessment

- Another one related to the content data from the best practices:
  https://smart-villages.eu/services/api/goodpractices

- And the last one that is related to the metadata from the best practices:
  https://smart-villages.eu/services/api/goodpractices/1367

2.3 Legal Issues

In accordance with data protection laws, we do not require special treatment of such information since we do not store data that can uniquely identify the person who generates it. For example, when a user proceeds to complete a smartness self-assessment process, the information asked is completely generic: name of the village, number of inhabitants, age group and type of organization for which it works. With that information, it is impossible to reconstruct the profile of that person. In addition to all this, this way of proceeding also allows us to offer the data openly to anyone who wants it for further study or analysis, since there is no risk to the privacy of individuals.

3. Additional backend tools

In addition to the database itself, we have developed a collection of auxiliary tools that allows getting the most out of the data stored. For example, we have a machine learning (ML) tool that allows us to discern whether a database entry should be considered legitimate or not. As one can imagine, there could be many entries in which users have not entered information truthfully. However, if we do not apply any mechanism to filter this information, we run the risk that the results of our analysis are not entirely realistic. For that reason, we need tools like this one that allow us to find out the legitimacy of a given entry.

3.1 Introduction

We have developed scripts for the analysis of the information contained in the database. These scripts include functionality to detect things such as the most similar people from a global point of view or looking only at a given dimension.

3.2 Determining the legitimate of entries

One of the most severe problems we have to face when working with data from smart villages is being able to automatically discern the veracity of the information to be analyzed. For example, in our framework, we capture a lot of data and information through questionnaires specifically designed to determine the degree of smartness of a given village. These forms are open to the public, and anyone can fill out the information truthfully or can do so without much thought, in a hurry, or without knowing for sure if the information they are entering is completely true. For us, it is extremely important to have the most reliable information possible; otherwise, the conclusions of our analysis run the risk of not being accurate. For this reason, we have been working on an automatic mechanism capable of verifying the plausibility of the data inserted into the system using different machine learning techniques.

Our solution is based on the notion of automatic classification which is a process intended to predict the outcome from a dataset that has been previously labeled by a human expert. In this case, we face a binary classification problem since there are only two classes: the filled form is valid or is not valid.

In order to do that, the method needs to use some training samples to understand how the given input variables relate to the output values. If the automatic classifier is trained properly, it can be used to detect the validity of new data inputs. In this context, there are many automatic classification methods publicly available, but it depends on the application and nature of the dataset to conclude which one is the best. . For this reason, we have tried several methods here with their standard configuration.

Figure 3 shows us the results obtained for the four classifiers for a sample of 200 entries corresponding the data from the smartness self-assessment. These classifiers are Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random Forest (RF) and Multilayer Perceptron (MLP).

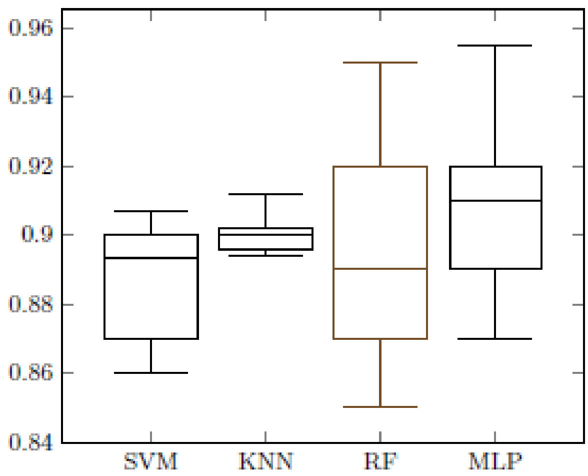

Figure 3. Summary of results achieved with different ML methods

3.3 Clustering

One of the most interesting capabilities that our framework for data analysis can offer is to automatically calculate the clusters or logical aggregation either of villages that share a similar degree of smartness in general or in specific thematic areas. In our specific case, clustering is a data analysis technique whereby given a set of villages (represented by values stating the answers to the questionnaire), we can classify each village into a specific group. In this way, villages that are in the same group should have similar properties, while villages in different groups should have highly dissimilar features.

3.3 Similarity calculation

Another feature we have implemented in our framework is the automatic calculation of the similarity between villages. Of course, this is not a physical similarity, but a type of similarity that measures the degree of maturity in relation to each of the six dimensions of study that we address in the framework of this work. The calculation of similarity is very useful because it allows the local authorities of the rural world to determine which places present some characteristics similar to those of the place in question, so it is possible to look at them as a third-person viewer and analyze what actions they are currently developing.

3.4 Ranking

The ranking functionality is related to the creation of an ordered list of villages to facilitate the understanding of specific factors. By reducing detailed features to a sequence of ordinal numbers, the ranking functionality makes it possible to assess interesting information according to some specific criteria. However, since the results only show the perception that the local authorities have about their village, these results must be taken with caution.

4. Conclusions

In this deliverable, we have presented in a summarized way the details regarding the database that implements our back-end and contains all the necessary information for the operation of the DEP. Our report has covered the technical aspects related to the database as well as a collection of lightweight tools to make use of the data stored.

Many of the problems of the rural world are the depopulation because of rural exodus, where many people leave the place where they have lived during many years in search of new opportunities in the urban world. This leads to an aging of the population, related to the previous point, as young people consider that other more populated places can be more attractive and offer more professional opportunities. On the other hand, the disappearance of public services, because it becomes very expensive to offer a service that will not have a large number of users. Through the DEP, local authorities from many villages can get in touch to share experiences about their progress in digitization and in their fight against the problems, we have mentioned.