The Framework Programme for Research & Innovation
Innovation actions (IA)

Project Title:

SMart mobILity at the European land borders

SMILE

Grant Agreement No: 740931

[H2020-DS-2016-2017] SEC-14-BES-2016 Towards reducing the cost of technologies in land border security applications

Deliverable

D4.3. Statistical Data Analysis and Reporting

Deliverable No.		**D4.3**	
Work package No.	**WP4**	Work package Title and task type	**Border Monitoring Data Collection, Analysis and Alerts Information Management**
Task No.	**T4.3**	Task Title	**Statistical Data Analysis and Reporting - Shared Biometric Matching Service**
Lead beneficiary		**FRAUNHOFER**	
Dissemination level		**Public**	
Nature of Deliverable		**Report**	
Delivery date		**03 May 2019**	
Status		**Final**	
File Name:		**[SMILE] D4.3_v1.0.pdf**	
Project start date, duration		**01 June 2017, 36 Months**	

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°740931

Authors List

Leading Author (Editor)				
Name		*Initials*	*Beneficiary Name*	*Contact email*
Alexandru Todor		AT	FRAUNHOFER	-
Co-authors (in alphabetic order)				
#	*Name*	*Initials*	*Beneficiary Name*	*Contact email*
1	Georgios Stavropoulos	GS	CERTH	
2	Dimitra Triantafyllou	DT	CERTH	
3	Kin Tsun Chiu	KC	eGovCD	
4	Evangelos Markakis	EM	TEIC	

Reviewers List

List of Reviewers (in alphabetic order)				
#	*Name*	*Initials*	*Beneficiary Name*	*Contact email*
1	*John Avramidis*	*JA*	*EULAMBIA*	
2	*Velina T. Dobreva*	*VD*	*CDBP*	

Document history			
Version	Date	Status	Modifications made by
0.1	03/04/2019	Document Creation - ToC	FRAUNHOFER
0.2	24/04/2019	First Draft	FRAUNHOFER
0.3	26/04/2019	Second Draft	FRAUNHOFER
0.4	03/05/2019	Third Draft – Int. review comments	FRAUNHOFER
1.0	03/05/2019	Final version	CERTH

List of definitions & abbreviations

Abbreviation	Definition
BCP	Border Crossing Point
EMSR07	European Migration Statistics Regulation of 2007
EU	European Union
GDPR	General Data Protection Regulation
TCN	Third Country National
EASO	The European Asylum Support Office

Executive Summary

This document aims to describe border monitoring data analysis and management. It particularly refers to statistical information regarding the situation at borders and monitored crossings in terms of traffic, immigration and criminal activities. In addition, the document addresses the topic of biometric analysis and matching for more robust and accurate identification results.

Existing BCP statistical procedures are initially analysed to define the categories of data addressed by SMILE. Then the data requirements are investigated, and the SMILE reporting architecture is presented. Moreover, the SMILE statistics and reporting module with its functionalities are described.

The last section of the deliverable includes a short description of hard and soft bio-metrics and their role in the identification procedure. To obtain better results and overcome any potential restrictions as a result of analysing each biometric feature separately, a shared biometrics matching service is proposed.

Table of Contents

List of figures

1. Introduction

Border monitoring data analysis has a key role in SMILE, since useful information on various aspects of the project is extracted. Topics involving border crossing points (BCPs) data analysis refer to border crossing issues, such as traffic, immigration and criminal activities as well as biometrics processing and identification procedures. Therefore, the goals of deliverable D4.3 "Statistical Data Analysis and Reporting" is to address and analyse these aspects, describe their requirements and restrictions and present the SMILE analytics module and shared biometrics matching service.

The aim of the SMILE statistics module is to provide useful information for BCPs in an organized and comprehensive way with compliance to sensitive data legislation (Voigt, 2017). In this context, existing BCP statistical procedures are analysed to determine the categories of statistical reports. Moreover, the data requirements determine the SMILE statistical architecture, which meets the specifications of flexibility, anonymity, scalability and customization. Furthermore, the SMILE statistics and reporting module comprises a variety of visual statistic components while specific admin personnel have the ability to modify and handle dashboards and components.

In addition to border crossing statistics, border monitoring data analysis includes a biometric matching service. Hard and soft biometrics are handled for identification purposes, while combined information is utilized to obtain more accurate results.

The rest of this deliverable is structured as follows. Section 2 provides an analysis of existing BCP statistical procedures. Section 3 presents the SMILE statistics and reporting architecture. Section 4 includes a detailed description of the SMILE statistics and reporting module with various examples and screenshots. In section 5 the shared biometrics matching service is analysed and section 6 concludes the deliverable.

2. Analysis of existing BCP statistic procedures

Most BCPs publish a series of statistical reports on account of the European Migration Statistics Regulation of 2007 (ESMR07) (COUNCIL, 2007). The scope of this regulation is to provide EU policy and decision makers with up to date statistical information in order to better assess the situation regarding migration and criminality at European borders.

In the following section, the statistical reports published by all our use-case partners are perused, the type of statistical information and how it is displayed is analysed as well as the way the EU uses the aggregated data.

2.1 Bulgarian BCP Statistical Reports

The Bulgarian Main Directorate "Border Police"[1] publishes statistics related to border control. These statistics present information on general police activities regarding the combating and

[1] https://www.mvr.bg/gdgp

prevention of crime as well as migration statistics published monthly in accordance with EMSR07 regulation[23].

The yearly statistical reports provide for example information concerning passenger traffic growth, third-country nationals (TCNs) detained at entry, those detained for illegally staying in of the country, number of illegal immigrants detained at entry or exit, etc.

- Passenger traffic growth
  - Passenger traffic growth at the Bulgarian land borders
  - Passenger traffic growth at the Bulgarian borders
- Third-country nationals detained
  - Total
  - Third-country nationals detained – total
  - Detained at entry
  - Detained at exit (without AFIS registration)
  - Illegal stay in the country
  - Detainee nationality
- Nationality of third-country nationals detained at entry (nationality, number, % from total)
  - Specific country
  - Other
  - Total
- Nationality of third-country nationals detained at exit (border, total, green border, border control point (BCP))
  - Specific country
  - Air Border
  - Maritime Border
  - Total
- Nationality of third-country nationals detained at exit (nationality, number, % from total)
  - Specific country
  - Other
  - Total
- Detained illegal immigrants on entry (border, total, green border, border control point)
  - Specific country
  - Air Border
  - Maritime Border
  - Total

[2] https://www.mvr.bg/министерството/programni-dokumenti-otcheti-analizi/статистика/миграционна-статистика
[3] https://www.mvr.bg/министерството/programni-dokumenti-otcheti-analizi/статистика/годишен-бюлетин-полицейска-статистика

2.2 Hungarian BCP Statistical Reports

The Hungarian National Police has published extensive statistical data pertaining to crime and immigration as well as risk assessment reports derived from this information.[4] The report from 2015 is hereby analysed since it is the only one published in English; however other reports from the Hungarian BCP website, seem to be identical in structure.

The General Situation Report on Border Management published by the National Headquarters of the Hungarian Police in December 2015[5] includes statistics on several categories of registered measures in border management.

The measures include:

- Unlawful acts related to illegal migration, including:
  - Human smuggling
  - Document falsification, including passport, ID card, residence permit, visa and stamp
  - Illegal border crossings
  - Overstays
  - Contraventions of alien policing rules
- Refusals of entry, this includes:
  - Lack of travel document/ valid visa
  - Possession of false/counterfeit/forged travel document/ visa
  - Lack of supporting documents
  - Overstays
  - Lack of financial cover
  - Alert in the national register/ SIS
  - Considered to be a threat to the country
- Warrant of captions includes national and international (SIS alerts)

In addition to the total numbers of each measure, the numbers are also broken down by:

- Border sections
- Counties (Except “Refusals of entry”)
- (Main) nationalities
- Hour of the days (“Unlawful acts related to illegal migration” only)
- Days of the week (“Unlawful acts related to illegal migration” only)
- Gender (“Unlawful acts related to illegal migration” only)

Besides the registered measures in border management, other statistics are included in the report, including:

- Handovers, takeovers according to International readmission agreements
- Transfers
- Ordered ban at entry and stay
- Handovers, takeovers according to Dublin Procedure
- Asylum applications
- Detention in the detention centre of National Police/ ordered by Police
- Persons ordered to leave
- Persons apprehended by border policing authorities
- Returned persons

[4] http://www.police.hu/a-rendorsegrol/statisztikak/hatarrendeszet?page=0

[5] http://www.police.hu/sites/default/files/hatarrendeszet_hk_2015.12_angol_1.pdf

- Passenger flow
- Vehicle traffic
- Composition of passenger flow

The report presents the statistics by comparing numbers for December 2014 and 2015, January to December 2014 and 2015, and the total percentage for 2015. The change in numbers is also presented in the actual numbers and dynamics in percentage.

2.3 Romania BCP Statistical Reports

Similar to the Bulgarian and Hungarian Border Police, the Romanian General Inspectorate of Border Police publishes a series of statistics about migration at the Romanian Border. These statistics are published as part of the Romanian government data portal[6].

The annual statistical reports contain the following information:

- People crossing the border
  - EU citizens
  - Non-EU citizens
- Vehicles crossing the border
- Illegal activities detected includes:
  - Contraventions detected at the border with
    - Hungary
    - Ukraine
    - Rep. Moldova
    - Bulgaria
    - Serbia
    - Airports
    - Ports
  - Criminal offences detected at the border with
    - Hungary
    - Ukraine
    - Rep. Moldova
    - Bulgaria
    - Serbia
    - Airports
    - Ports
- People illegally crossing the border or attempting an illegal border crossing
  - Number of people being prevented to illegally cross the border
  - Foreigners organized in groups includes:
    - Migrant groups
    - Human traffickers
- Number of citizens who were denied entrance at the border
- Criminal offences constituting contraband or customs fraud

[6] http://data.gov.ro/ro/organization/inspectoratul-general-al-politiei-de-frontiera

  - People detained illegally crossing or attempting to illegally cross the border upon which objects were found that constitute contraband
  - People identified to be committing contraband with cigarettes
    - Groups identified committing contraband with cigarettes
- Items retained in order to be confiscated
  - Total value
    - Value of the cigarette packages retained in order to be confiscated
  - Cigarette packages (based on numbers)
  - Vehicles suspected to be stolen
  - Weapons
  - Handcuffs
  - Bullet cartridges
  - Other narcotics (kg)
  - Cocaine (kg)
  - Cannabis (kg)
  - Heroin (kg)
  - Other forbidden substances (gr)
  - Boats
  - Boat motors
  - Fishing nets, monofilaments (meters)
  - Fish (kg)
  - Jewellery (kg)
  - Alcohol (litres)
  - Fake brand clothing (items)
  - Fake brand cosmetics and perfumes (items)
  - Fake brand shoes (pairs)

2.4 Frontex Aggregated Statistical Reports

As stated above, the scope of the ESMR07 regulation that determines the publication of immigration statistics by all BCPs is a better overview of migration trends. The overview is meant to provide regulators with the data needed to make informed policy decisions.

One of the best examples of such aggregated data use is the risk analysis reports published by Frontex (the European border and coast guard agency). FRONTEX publishes a yearly report with risk analysis results (Frontex 2018). The statistical data in these reports is derived from the reports complied by all EU member states on the numbers of people crossing European borders legally or illegally. Each yearly report conveys the aggregated data for the previous year, e.g., the 2018 report contains data for 2017.

When processed and elaborated against geopolitical phenomena, these numbers re-veal some rather interesting trends as follows. The lowest numbers on illegal border crossing events since 2013 were recorded in 2017. This decline is the consequence of decreased detections on the Eastern Mediterranean and Western Balkan routes. Additionally, the reduction in the number of detections on the central Mediterranean route was a great achievement. However, the tremendous rise in illegal detections on the Western Mediterranean route outshines this reduction.

The European Asylum Support Office (EASO) reported that 701997 applications for international protection were lodged in all EU member states, including Norway and Switzerland. This is half the number reported in 2016. Bearing that in mind, there were substantially fewer illegal border crossing detections reported, then applications submitted for international protection. This leads to the conclusion that some of the applicants had crossed EU borders illegally and undetected under a visa-free scheme.

Furthermore, 183548 refusals of entry at European borders were documented in 2017, which is 15% lower than in 2016. In addition, 6700 entry attempts with fake or invalid documents were reported at the external borders, representing a new lowest record since 2012. In contrast, border crossing attempts with fake or invalid documents in-creased by almost 9% within the Schengen area.

Furthermore, 435786 illegal stay detections were recorded by all member states, which is 11% less than in 2016. Since 2015 when the European migration peaked, 2017 was the second year in which the number of illegal stay detections dropped.

Finally, the overall number of illegal human trafficking facilitators recorded decreased by nearly 19%. A probable reason for this is the ever-changing and evolving smuggling networks' means of operation and sophistication levels constantly being adopt. As such, border authorities find it more and more difficult to detect and apprehend them.

The following Figure (Figure 1) illustrates the aggregated report data.

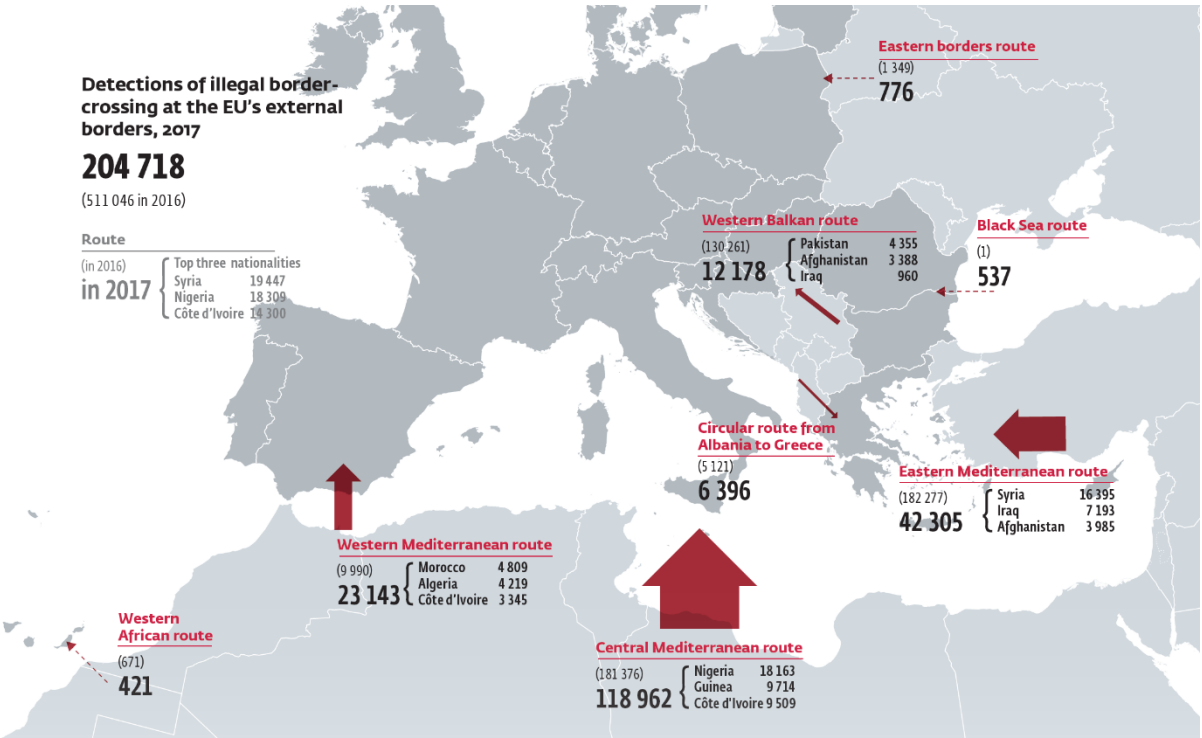


Figure 1 Detection of illegal border crossing within the EU external borders (Frontex, 2018)

From the above Figure, it can be safely deuced that 204718 illegal border crossing attempts were detected in 2017, which is 60% lower than in 2016 (511046). Additionally, it is also observed that the top three nations detected in illegal flows were Syria (19.447 persons), Nigeria (18.309 persons), and Côte d'Ivoire (14.300 per-sons) constituting 9%, 9% and 6% respectively of the overall illegal flows detected.

It can be concluded that the enforcement of checks within European borders denotes the main safeguarding method for the Schengen area and a continuous contribution to the long-term safety of its member states.

3. SMILE's statistics and reporting architecture

In this section, the requirements are derived from the BCP statistical procedures, after which a system architecture designed to meet those requirements is presented.

3.1 Requirements Analysis

This section entails an analysis of the statistical reports published by the border police of Bulgaria, Hungary and Romania, as well as the aggregated reports by Frontex and corresponding EU regulations. Our data sources are based on the in-formation provided to us by the respective BCPs, as well as an analysis of the BCP websites in each language, official EU law and the Frontex website. Subsequent to the analysis, the types of statistical reports are classified in multiple categories. Then the data required for each category is analysed and the kind of access SMILE has to this data is further investigated

3.1.1 Categories of statistical reports

- Category 1: Statistics about traffic at the border
  - This information includes yearly reports of how many people or vehicles crossed the border and real-time information as to the expected waiting time at the border
- Category 2: Statistics about immigration
  - This category includes information about attempts of illegal border crossings, refusals at entry, overstays, etc.
- Category 3: Statistics about criminal activities at the border
  - Criminal activities such as smuggling, human trafficking, stolen vehicles, forged documents etc.

3.1.2 Data requirements by category

Each presented category corresponds to different requirements for type of access to information required to generate these kind of statistical reports.

- *Category 1*: this category requires information from the designated SMILE lane as well as non-SMILE lanes (lanes without pre-registration and the second lane) in order to produce a reasonable overview of border traffic. Non-SMILE traffic lanes are outside the project scope, as stated in D 2.4 and D 7.4.

- *Category 2*: the majority of illegal immigration-related offences are recorded at the green border, which is beyond the SMILE project scope, and data from national and international police databases is required. Information regarding refusals on entry can only be partially recorded in SMILE. The reason is that, when issues with ID documents or visas are detected by SMILE, the persons are transferred to the second lane, (also outside of the SMILE scope).
- *Category 3*: as with category 2, many of the criminal activities are detected at the green border and fall outside the scope of the SMILE project. For instance, either on account of risk analysis or the direct intervention of a border guard, in case an issue is detected with a person trying to cross the border in the SMILE lane, they are sent to the second lane.

3.1.3 Addressing the data requirements and related issues

From our analysis in Section 2, we come to the conclusion that, in order to provide statistical reports similar to those published by the BCPs, the SMILE reporting architecture needs data from multiple databases such as national and international criminal police databases, customs databases, BCP databases external to SMILE. Since the integration with such databases is only partially foreseen in the scope of the project, we have made following architectural decisions in order to address these issues:

Data Integration

We have established a series of requirements to the architecture of the deliverable in order to be able to easily integrate all the necessary databases for producing relevant statistical reports, may they be for a single crossing point, or aggregated reports over multiple countries.

Data Simulation

Since we have no access to the required databases, we addressed the issue by simulating all databases required as well as the information that we expect to be present in the fully developed SMILE platform. Based on the analysis of statistical reports we derived certain distributions of different characteristics of the travellers and generated the necessary data in the system to reproduce these reports. A more detailed description of the data simulation module, as well as the integration with external databases, will be published in D 4.4.

3.1.4 Technical requirements

The data requirement analysis poses a series of technical requirements addressed to our architecture design:

- **Flexibility**: the statistics and reporting architecture needs to be designed so as to adapt to the rapidly changing legal framework coordinating BCP activities. This means the need to enable adding or removing data sources outside the scope of SMILE.
- **Customization**: the changing legal frameworks and further digitalization in the different border control organizations, require end users to be able to add new types of visual statistical components whenever needed and with minimal effort.
- **Scalability**: The system needs to cope with millions of users and travellers a day, as well as be able to visualize and analyse all travellers. It is also necessary to be able to expand the

architecture in case the project scope is extended (second lane, customs, criminal cases etc.)

- **Anonymization**: since all current BCP statistical procedures do not require any kind of personal identifiable data, the system should be designed in a way that enables disregarding all personally identifiable data before it reaches the statistical components.
- **Separation**: the SMILE project is meant to incorporate multiple BCPs on one platform. However, in some cases current regulations do not foresee the real-time sharing of information between BCPs. This means that national BCPs need to have their own copy of their own data and have ownership of the statistical components.

3.2 System architecture specifications

Based on the defined architectural requirements, the system architecture is as depicted in Figure 2.

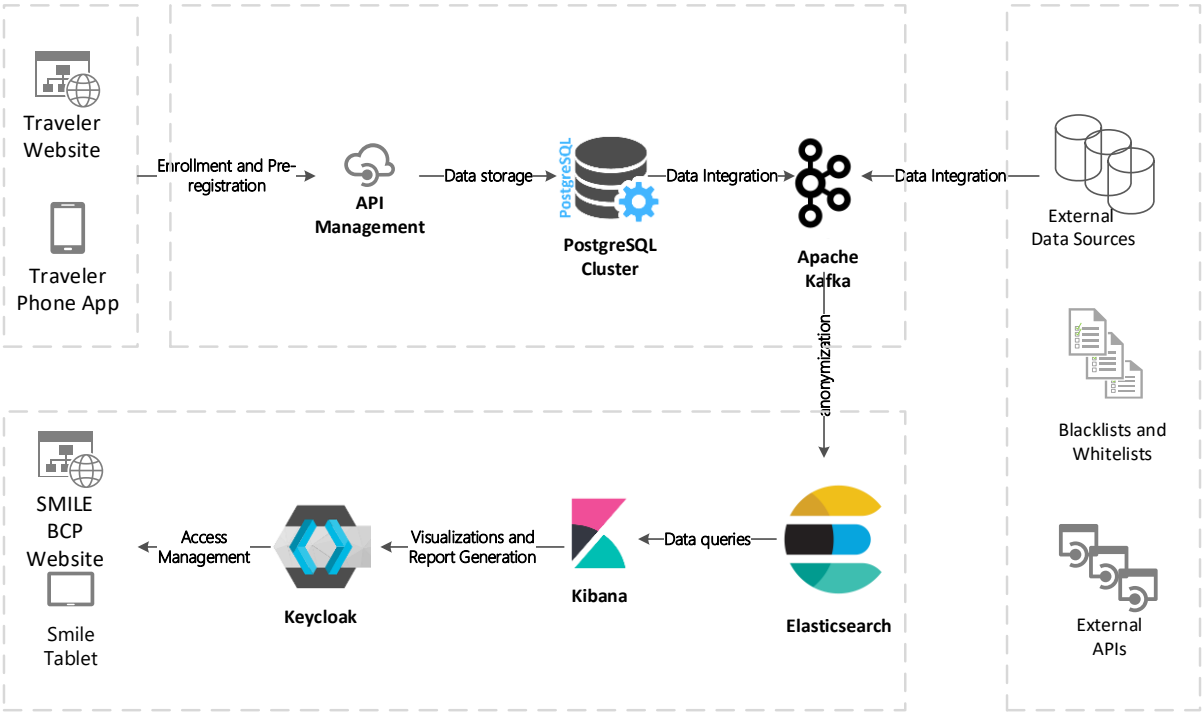


Figure 2: D4.3 Software Architecture

In Figure 2 we show the different components of the smile architecture as well as different zones that are defined for the purposes of data isolation and to enhance security.

Zone 1 contains the data producing applications that travellers use. Zone 2 is an abstraction of the general SMILE architecture presented in D 2.4; it contains all SMILE components outside of the statistical and reporting components. Zone 3 contains all the statistical components that are internal to a BCP. Zone 4 is the external zone and contains external information from the EU-Lisa databases[7] as well as national police databases and blacklists. This architecture will be discussed in more detail in D 4.4.

[7] https://www.eulisa.europa.eu/

Components and their role in the architecture:

Postgres[8]

Postgres is an open source relational database management system. It comes with out-of-the-box shading and cluster support. It can scale up to millions of queries per second and provides first-class tool support. The role in our architecture is as the database management system for SMILE, where all the services store the data.

Apache Kafka[9]

Apache Kafka is a stream processing software platform designed to handle real-time data input with low latency and high throughput. Processing data streams is depending on scalable message queue designed as a distributed transaction log. It provides a multitude of connectors to various database systems and allows us to integrate external data sources in our architecture. Furthermore, by selecting only specific columns from Postgres, we can ensure that no identifiable data is passed to the statistical components.

Elasticsearch[10]

Elasticsearch is a RESTful distributed search and analytics engine for all kinds of documents. It provides scalable near real-time search functionality, as well as a plugin architecture that allows it to be extended with statistical analysis components, as well as alerting.

Kibana[11]

Kibana is a data visualization plugin for Elasticsearch that enables the creation of statistics and visualizations based on the data indexed in Elasticsearch. Some of the visualization options it supports include bar, line and scatter plots and different kinds of charts and maps. The visualizations can be updated in real-time since they are based on live data from the index.

Keycloack[12]

Keycloack is an open source software component that offers integration facilities with multiple identity management systems such as ActiveDirectory/LDAP and Kerberos, as well as 2-Factor authentication. Its role in the software architecture is to offer added security and integration with the SMILE system as well as BCP login systems.

[8] https://www.postgresql.org
[9] https://kafka.apache.org
[10] https://www.elastic.co/products/elasticsearch
[11] https://www.elastic.co/products/kibana
[12] https://www.keycloak.org/

Interactions between components:

Figure 2 also shows the interactions between the different components; we will proceed to explain this in more detail:

When a user pre-registers for a travel the information is passed From Zone 1 to Zone 2. The API management in Zone 2 stores the corresponding data in a PostgreSQL Database. The Kafka connector registers each new travel pre-registration and passes only statistically relevant information to Zone 3. Apache Kafka retrieves other statistically relevant information required for the reports from Zone 4 and passes this to Zone 3. In Zone 3 the data is stored in Elasticsearch. The Kibana visualization component in Zone3 contains a series of predefined dashboards, which can be accessed by authorized personnel from the BCP interface. The dashboards can be exported as PDF reports, or as spreadsheets. BCP admins have the ability to define new dashboards in Kibana and to define access rights to the dashboards.

3.2.1 Requirement Fulfilment

Flexibility: the architectural design should be able to easily integrate other data sources through specialized components in Apache Kafka, as well as by loosely coupling the statistical interface formed by Elasticsearch and Kibana, allowing those components to be used independently of the rest of the system.

Customization: this requirement is addressed by including the capability for BCP personnel to easily add or change the displayed statistics, with minimal training. Furthermore, the data model can be augmented by new data sources without affecting any other system components.

Scalability: the chosen components are among the de facto standards for big data architectures, and are proven to scale to data sizes much greater than required for the project[13][14]

Anonymization: the data requirements analysis shows that the statistics require no data to identify persons. The system is designed so that no information such as names or IDs is sent to the statistical component.

Separation: the data flow from the SMILE system to the statistics module can be specifically con-trolled by the Kafka component. This means that a national BCP will only have access to the information that refers to the component, and BCPs will be allowed to run this system independently and if they so require.

[13] https://benchmarks.elastic.co/index.html
[14] https://kafka.apache.org/performance

4. Smile statistics and reporting module

The SMILE statistics and reporting module has two main components: the BCP Dashboard, and the BCP Statistics Admin Panel. The BCP Dashboard presents a series of predefined visual statistic components that have been created with a specific scope by the BCP admin personnel. The BCP Admin Panel is meant for specialized personnel that can create new dashboards or modify existing dashboards or visual components.

4.1 BCP Statistics and Reporting Dashboard

In the SMILE interface for BCP authorized personnel, we show an example dashboard that is meant to exemplify the analytical capability of the platform. This interface is only visible to the employees that have been granted access to it by the BCP admin by clicking on the statistics menu item in the BCP management panel.

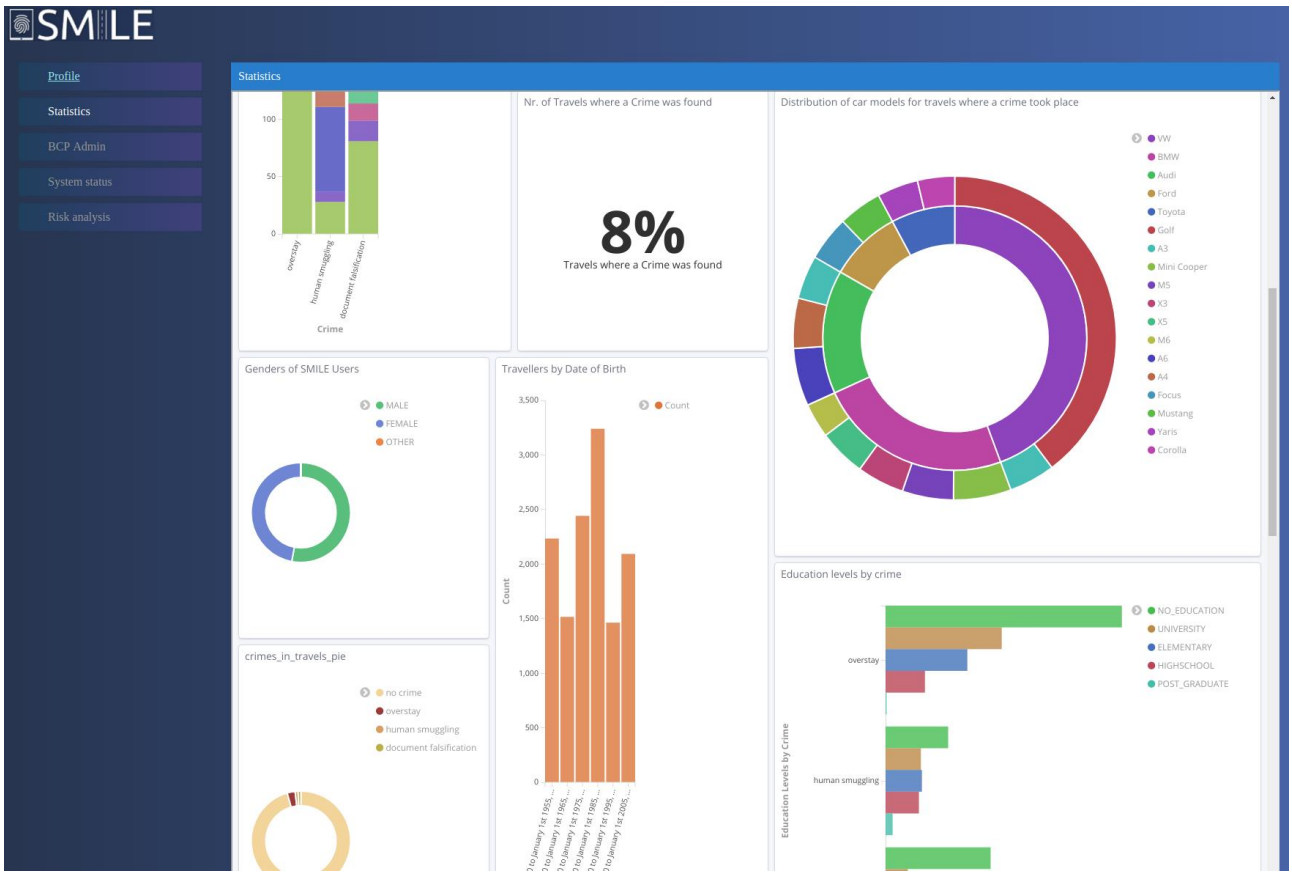
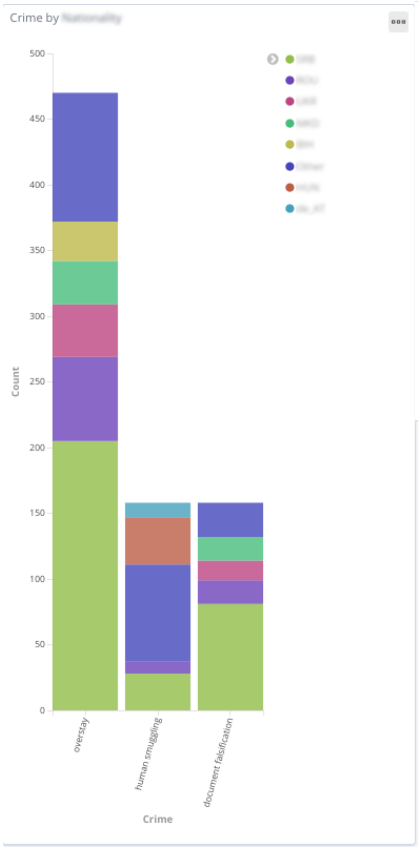
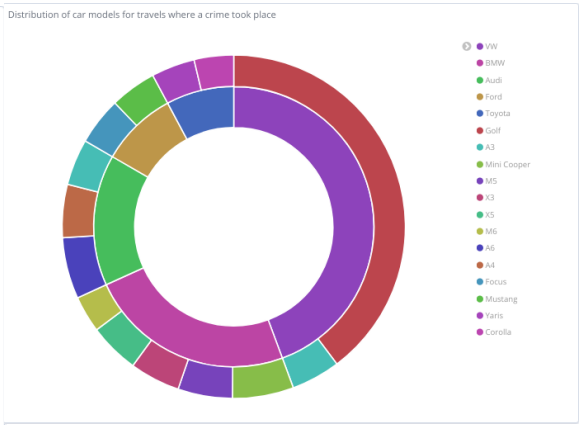


Figure 3 Screenshot of the SMILE Dashboard

This interface shows a series of statistical components, based on the simulated data. We will proceed to explain each component in part, what type of statistic it is, and for which use cases it can be used.

Crime histogram by Age, type of entrance, duration, date of birth, education level, etc.

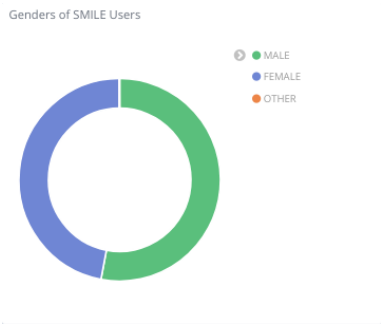




This diagram shows us which car models are used for smuggling.

Statistical value that displays the percentage of travels where a criminal act was detected

User gender counting



The education level by crime can be counted as shown in the following example

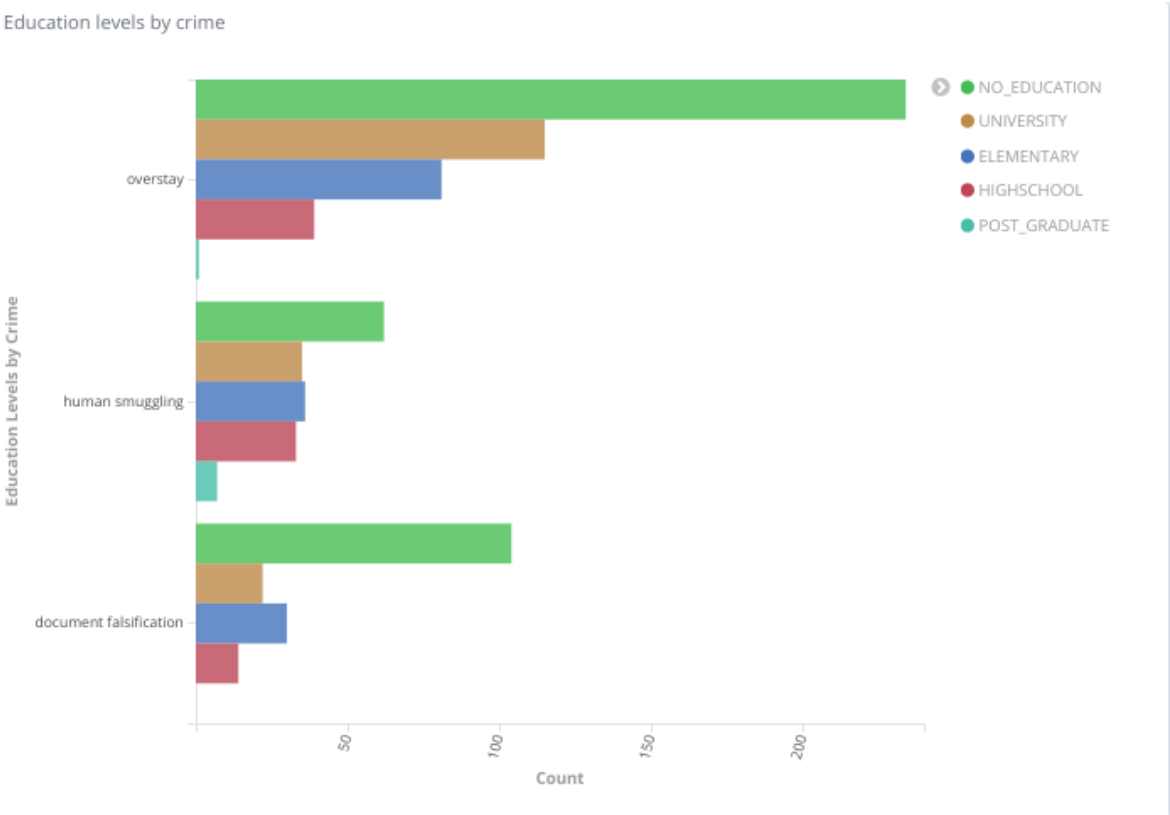


Figure 4 Example Barchart Visualization from the SMILE Dashboard

A very useful feature is to inspect the data of a diagram; this will show us a list of this data.

Distribution of car models for travels where a crime took place

View: Data

Download CSV

vehicle_make.keyword: Desce...	Count	vehicle_model.keyword: Desc...	Count
VW	346	Golf	310
VW	346	A3	36
BMW	186	Mini Cooper	45
BMW	186	M5	40
BMW	186	X3	37
BMW	186	X5	37
BMW	186	M6	27
Audi	118	A6	45
Audi	118	A4	39
Audi	118	A3	34
Ford	69	Focus	35
Ford	69	Mustang	34
Toyota	61	Yaris	32
Toyota	61	Corolla	29

Rows per page: 20

As we see the data are downloadable as a CSV file format that can be loaded or applied to another system or software.

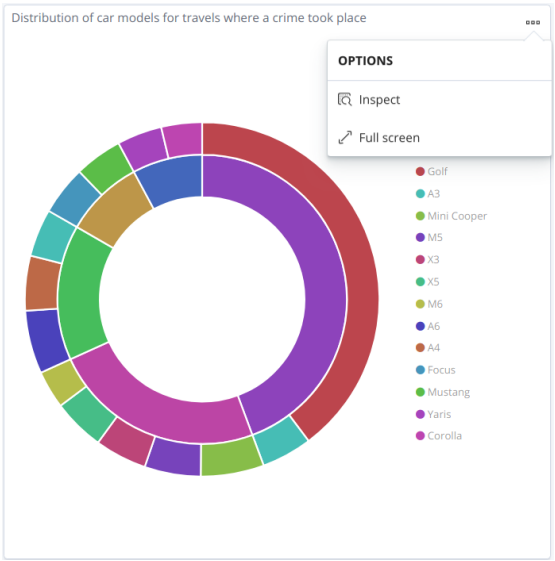
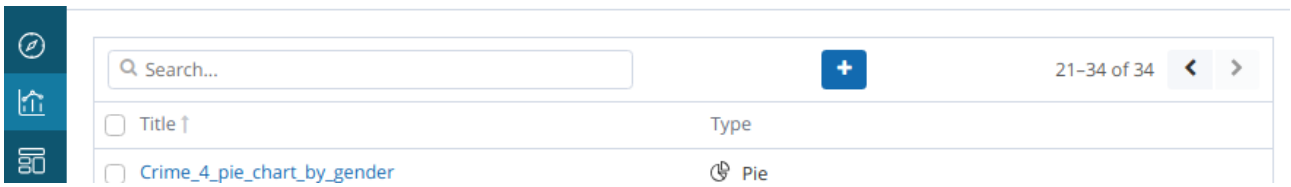


Figure 5 Example Piechart Visualization from the SMILE Dashboard

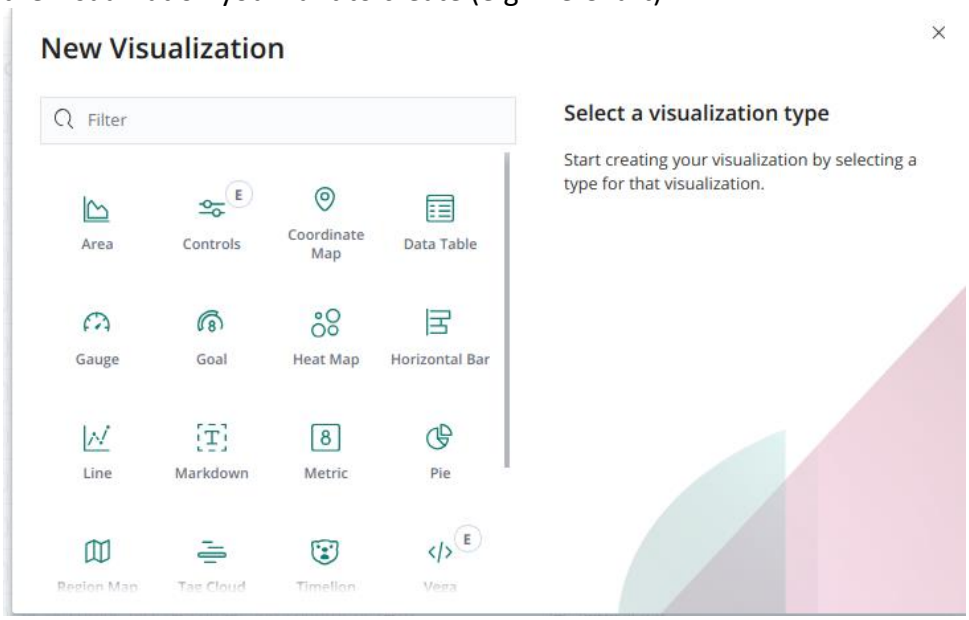
4.2 The BCP Statistics and Reporting Admin Panel

The BCP admin panel, accessible by BCP admin personnel, enables the creation of new dashboards and visualization components. In the following section, we show how to create new visualization components or how to adjust existing ones. The Admin panel can be accessed from the SMILE BCP interface/BCP Amin.

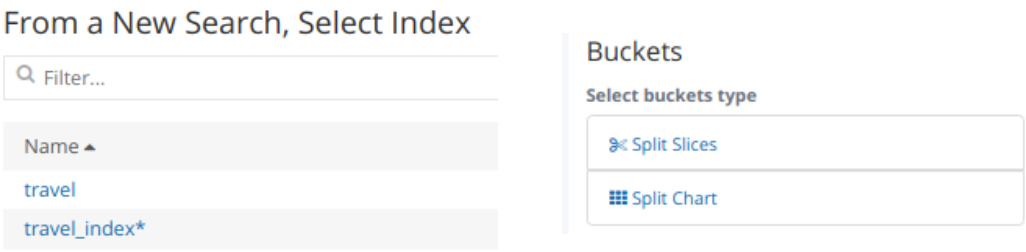
1) In *visualize* menu click the *plus* symbol



2) Select the visualization you want to create (e.g. Pie Chart)

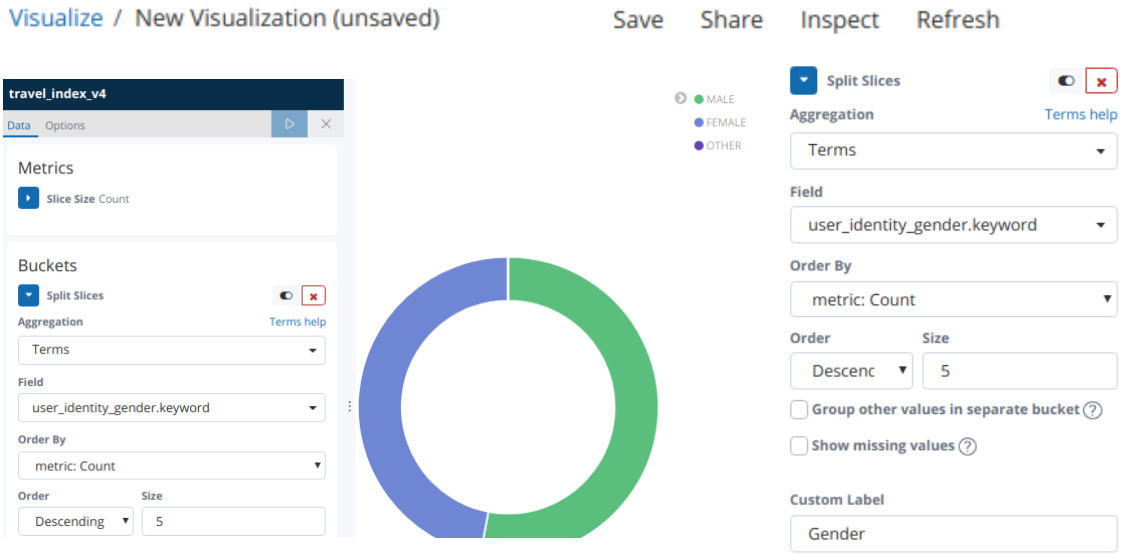


3) From the list of indices select the Elasticsearch index you want to use



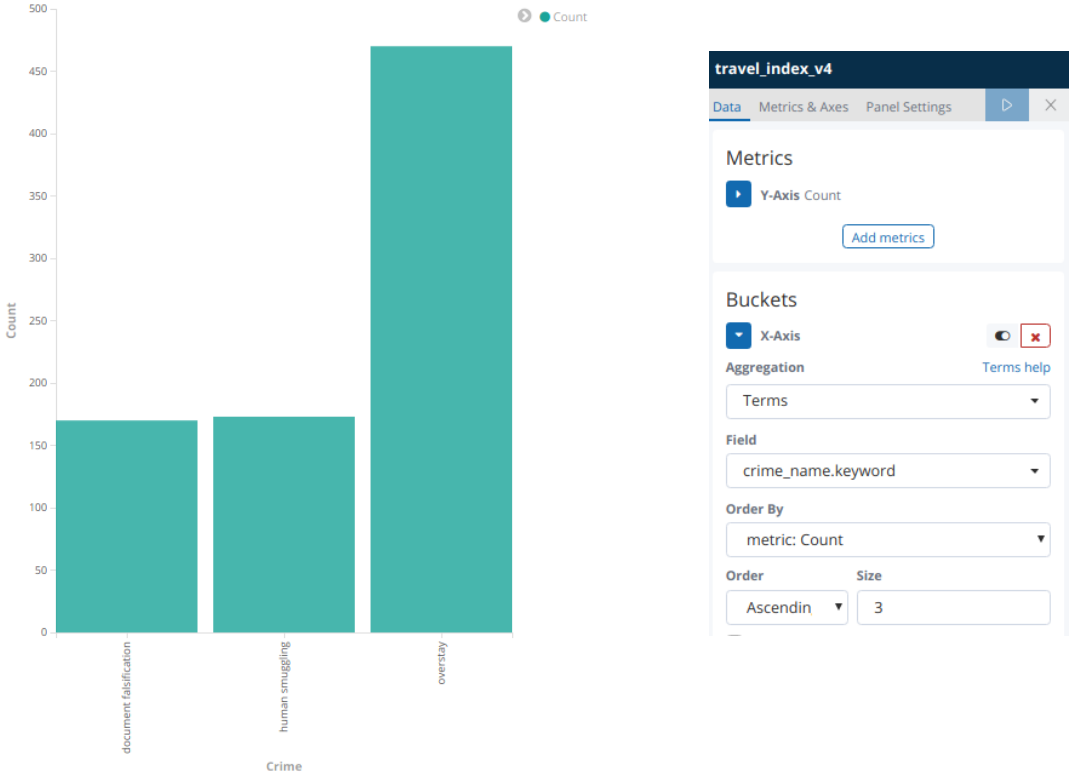
4) Click *Split slices* to get a menu for selecting an aggregation
5) Select *Terms* as aggregation and a Field of interest (gender) and an aggregation type (default: count) and define a Custom Label (*Gender*)
6) Click the *play* button to preview the visualization

7) Should look like this:
8) Save the visualization by clicking the *save* button in the menu bar on top



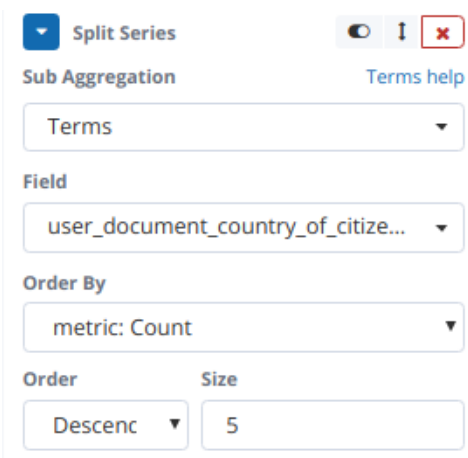
4.2.1 Creating a histogram for crime by “criteria”

1) Select a *vertical bar* from the *create new visualization* menu
2) Select your index of interest
3) In “Buckets” select *X-axis*
4) As an aggregation select *Terms*
5) Select the field of interest (*crime_name*)
6) Click the *play* button to preview the plot. So far it should look like this:



7) Click *Add sub-buckets* at the bottom and select *Split Series*

8) As an aggregation select *Terms* and select the field of interest (*i.e. country of citizenship*)
9) Click the *play* button to preview the plot; it should look like this



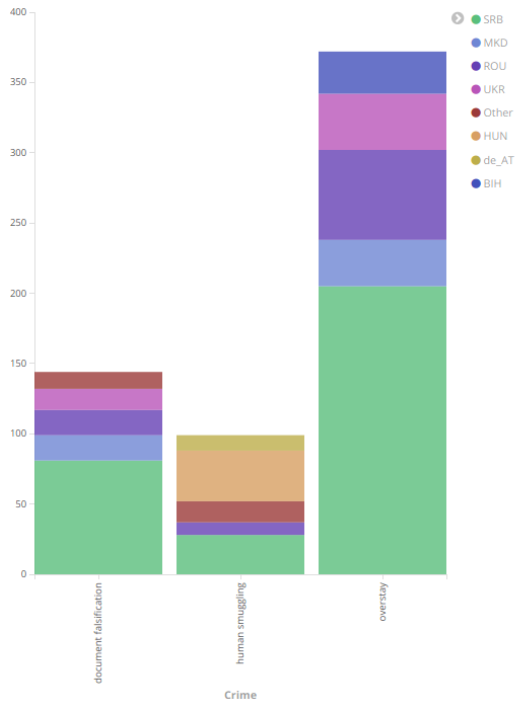


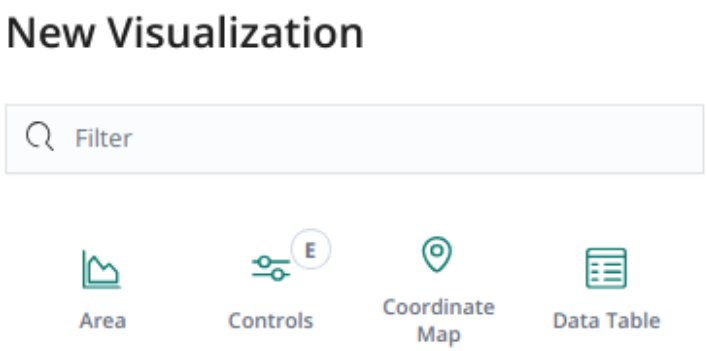
Figure 6 Example Histogram Visualization from the SMILE Dashboard

In the dashboard you can specify the time range:

4.2.2 Creating a Table with counts of border crossings

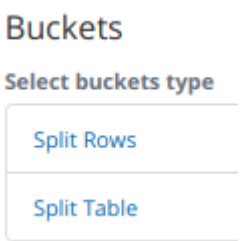
In order to create a table with counts of crossings of citizens from countries that are non-Schengen-members follow these steps:

1) Go to the visualization menu, click the *plus* button and choose *Data Table* as visualization type



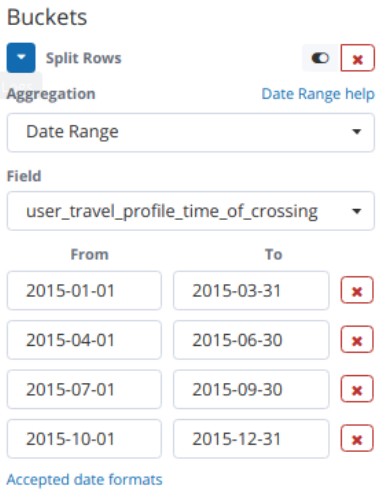
2) Choose your index and proceed

3) Select the appropriate bucket type: *Split Rows*

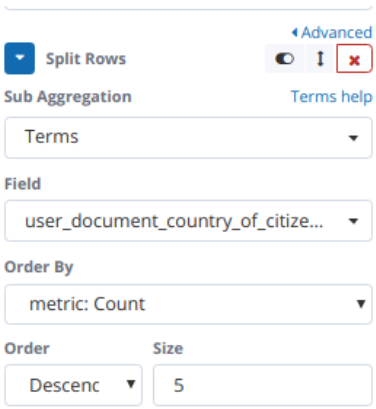


4) Choose the aggregation type: since we want to list the nr. of border crossings by time range we select *Date Range*
5) Choose the field of interest: *time of crossing*

6) Add as many time ranges as needed by pressing the *Add Range* button and specify the date ranges (in the example it was done for the quarters of 2015). After steps 4, 5, and 6 it should look like this:



7) Click the *Add sub-buckets* button to specify more columns and select the *Split Rows* bucket type.
8) As aggregation type select *Terms* and as Field *country of citizenship.* This part should look like this:



9) After specifying Custom Labels and clicking the *play* button to preview the table, it should look like this:

Time of Crossing (Date Range)	Nationality	Count
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	SRB	197
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	it_IT	159
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	ro_RO	151
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	bg_BG	141
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	hu_HU	141
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	SRB	199
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	it_IT	181
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	ro_RO	150
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	KOSOVO	135
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	hu_HU	135

10) Since we are only interested in users from non-Schengen countries, we can apply a filter, matching only the entries of those users. To do that use the search bar on top and specify a query like this (for space reasons only selection shown in the screenshot):

```
>_ user_document_country_of_citizen_ship.keyword :"ro_RO" or user_document_country_of_citizen_ship.keyword : "bg_BG" or user_document_country_of_citizen_ship.keyword : "uk_UA"
```

11) After pressing the *Refresh* button on the top right, the table should be filtered to contain only users from non-Schengen countries. It will look as follows:

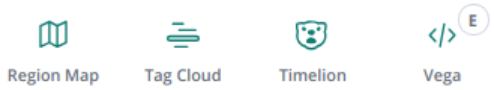
Time of Crossing (Date Range)	Nationality	Count
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	ro_RO	151
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	bg_BG	141
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	KOSOVO	132
January 1st 2015, 01:00:00.000 to March 31st 2015, 02:00:00.000	uk_UA	111
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	ro_RO	150
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	uk_UA	139
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	KOSOVO	135
April 1st 2015, 02:00:00.000 to June 30th 2015, 02:00:00.000	bg_BG	128
July 1st 2015, 02:00:00.000 to September 30th 2015, 02:00:00.000	ro_RO	145
July 1st 2015, 02:00:00.000 to September 30th 2015, 02:00:00.000	uk_UA	141

Figure 7 Example Table Visualization from the SMILE Dashboard

12) To save the table use the save button of the menu bar at the very top

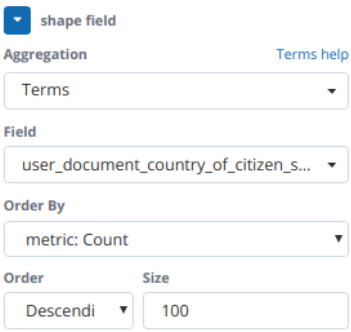
4.2.3 Creating a Region Map visualization of nationalities of SMILE users

1) Go to the visualization menu, click the *plus* button and choose *Region Map* as visualization type

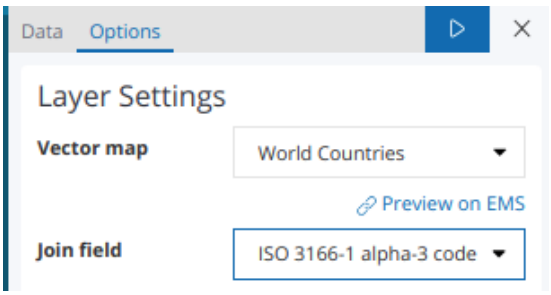


2) Choose your index and proceed

3) As bucket type select *shape field*
4) As aggregation select *Terms* and select *Country of Citizenship* as Field



5) Depending on how the nationalities are represented in the data select the correct type for the *Join Field* option in the *Options* menu, like this:



6) After using the *plus/minus* buttons in the top left corner of the map in order to zoom in on the countries of interest and pressing the *play* button to preview it should look like this

Figure 8 Example Map Visualization from the SMILE Dashboard

4.3 Statistical Report Generation

After a dashboard has been created and a specific time frame for the data has been selected, it can be exported as a pdf report. The reports can be further improved by activating the “Optimized for printing” option.

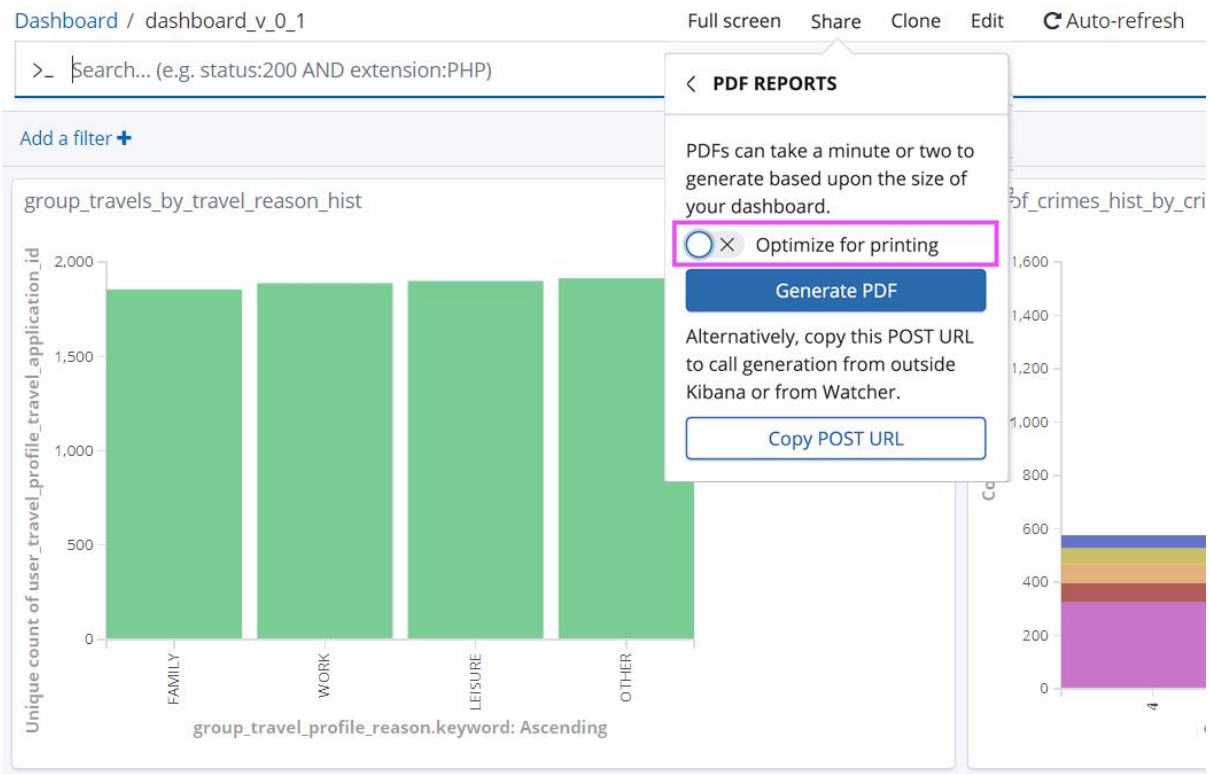


Figure 9 Generating a PDF Report from the SMILE Dashboard

5. Shared biometric matching service

The SMILE system employs several biometric traits in order to verify the identity of its users during the trip pre-registration and border crossing stages. These traits are individually extracted from each registered SMILE user, through live photos and passport data, using the various SMILE devices of the SMILE framework (users' mobile devices, BCP officers' devices etc.).

The specific biometric traits employed within the SMILE system can be classified into two categories:

- Primary biometrics (also known as "Hard" biometrics), which remain constant regardless of the capture conditions, thus presenting the necessary discriminative capacity to allow accurate person identifications
- Soft biometrics, which are non-unique personal attributes that can provide indications about the identity of a person, working complementary to primary biometric traits in order to enhance person identification accuracy in unconstrained settings

While each biometric trait, on its own, can present different advantages and limitations, the combined utilization of all these features allows the SMILE system to identify, in a robust and accurate fashion, all the registered users regardless of the individual SMILE device used to capture their appearance.

5.1 Hard Biometrics used in SMILE

The SMILE system utilizes three hard biometric features towards person identification[15]:

1. Iris
2. Face
3. Fingerprint

5.1.1 Iris Recognition

Iris recognition exploits the complex patterns of the human iris, which are unique for each individual and remain unaltered during his lifetime, in order to identify a person. Close-up images of one or both irises are captured, usually with the use of NIR cameras, with pattern-recognition algorithms employed to match the detected iris patterns to a database of identified subjects.

5.1.2 Face Recognition

Face recognition relies on the extraction of features from the overall human face, using various imaging sources (RGB, depth etc.). While some facial features can be altered, such as facial hair, or eyewear, recent advances in Deep Learning, have led to the generation of automated face recognition architectures which can provide high accuracy results even under challenging conditions.

5.1.3 Fingerprint Recognition

Similar to Iris recognition, fingerprint-based authentication takes advantage of the complex fingertip patterns in order to verify the identity of a user. As they are nearly unique and relatively difficult to alter, even though not as difficult as the Iris, fingerprints are suitable as long-term markers of human identity. Moreover, as fingerprint recognition has been the primary identification method for almost a century, in contrast to face and iris which have only recently started to be broadly adopted, there are extremely large fingerprint databases available, which can help with the identification of potentially high-risk individuals.

Modern fingerprint acquisition techniques usually employ touch scan devices, grouped into two sub categories: optical readers and solid-state readers. Meanwhile, in recent years touchless optical scanners have also been introduced.

5.2 Soft Biometrics used in SMILE

Complementary to the primary biometric traits, the SMILE system also employs seven soft biometric features, which are extracted from the faces of the users. For the acquisition of these features, a mobile-optimised deep learning framework is employed[16], using RGB images captured by the users' mobile devices.

[15] A more detailed description in given in Deliverable D3.2 "Individual-based hard biometric authentication modules"

[16] The soft biometrics module is described in detail in Deliverable D3.3 "Soft Biometrics Module"

Table 1: The facial soft biometric traits extracted by the soft biometrics module

Gender	male / female
Age Group	infant / child / young adult / adult / senior
Race	Caucasian / African / Asian / Hindi / mixed
Eyes State	open / partially open / closed
Eyewear	glasses / no glasses
Smile	smiling / not smiling
Beard	beard / no beard
Moustache	moustache / no moustache

5.3 Shared matching biometrics service.

Individual biometric systems may have to contend with a variety of limitations such as noisy data, intra-class variations, restricted degrees of freedom, non-universality, spoof attacks, and unacceptable error rates. On the other hand, multimodal systems combine any number of independent biometrics, overcoming some of the limitations arising when using just one biometric as the verification tool.

Within this scope, a multimodal biometric matching service is employed in SMILE, utilizing all the available biometric traits, towards the authentication of users. More specifically, the biometric authentication of travellers is three-fold: live biometrics captured at the border crossing, by the SMILE infrastructure, are compared with biometrics extracted from the user's e-passport data, as well as the biometric features stored in the user's profile (matching between e-passport biometrics and the user profile is performed independently during the capture of the corresponding data). Moreover, comparisons against various third-party national and international databases are performed, using each time only the available biometric features, in order to identify high-risk persons.

The hard-biometric data comparison requires absolute matches between all three modalities, while small deviations are allowed for the soft biometric features, as they can be temporally variable (i.e. changing facial hair) and are used complementary to enhance the verification accuracy of the hard biometrics[17]. Towards this end, each biometric extraction module, including both hard and soft biometrics, has been individually evaluated during the development process, described in detail in Deliverables D3.2 and D3.3, in order to ensure low error rates which, carry over to the multimodal verification process.

The whole biometrics verification process at the BCPs is controlled through the developed SMILE biometric matching service, which runs in the background on the BCP officers' mobile devices, and updates the corresponding entries continuously on the SMILE database, based on the captured biometric features and the biometric verification results.

In order to ensure the smooth and uninterrupted operation of the matching service, special care has been given to make sure that all underlying algorithms are optimized for low power

[17] The detailed biometric matching algorithms are described in Deliverable D3.4 "Privacy Preserving multifactor & multimodal travellers verification at land BCPs"

mobile devices and can run in real time on such systems, with the soft biometrics module, for example, requiring less than 130ms to run on a mid-range mobile device (see D3.2 & D3.3).

Besides the technical specifications, the multimodal biometric matching service also employs privacy preserving and enhancing techniques, described in Deliverables D8.2 and D8.3, while also following the ethics guidelines of the SMILE project, as defined in Deliverable D8.6

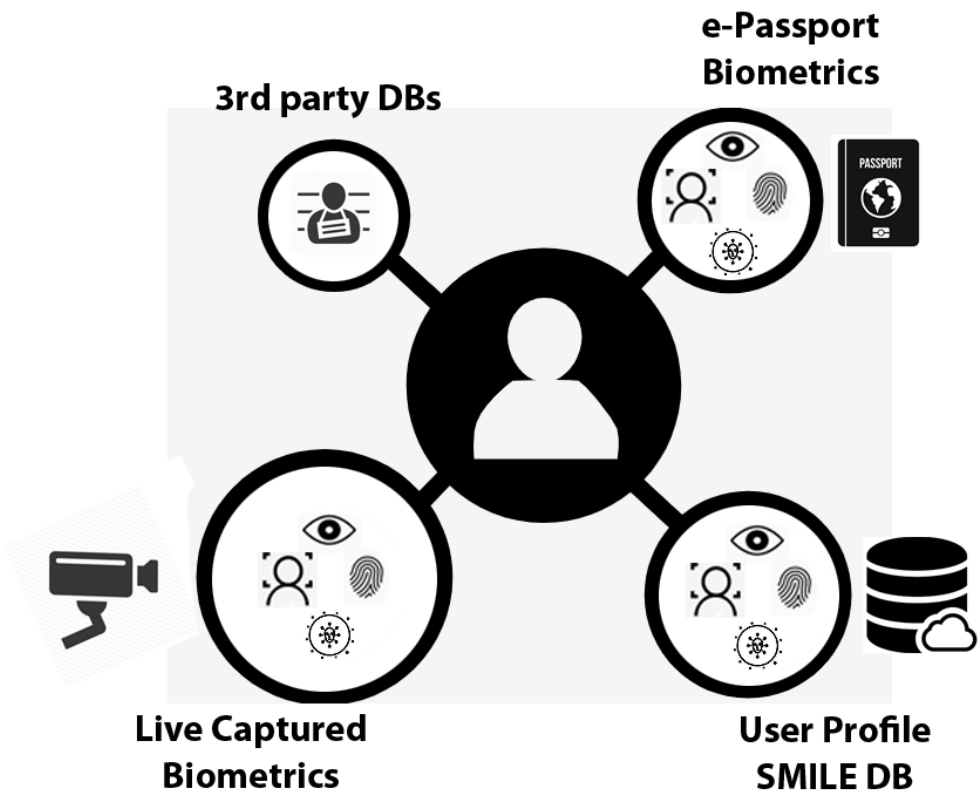


Figure 10 The SMILE matching biometrics service. The hard and soft biometrics captured live at the BCPs are compared with the corresponding biometrics extracted from the user's e-passport and stored user profile. Comparison against 3rd party databases is

6. Conclusion

The types of statistics that BCPs currently publish is regulated by the European Council (COUNCIL, 2007). Currently these statistical reports are produced with a lot of human intervention, and the precise content varies between different BCPs. The SMILE project, as an integrated system that can provide functionality for multiple BCPs at once, offers statistical components that can be used to generate different kinds of statistical reports automatically, offering flexibility and new visual data analysis insights to the users of the platform. Data integration challenges have been identified, and the proposed architectural solution aims to solve them by offering a flexible data integration solution, which allows the statistical reporting components of SMILE to be integrated and used together with the internal software systems of the BCPs, or other EU projects that provide functionality that is outside of the scope of the SMILE Project. The developed architecture also addresses the privacy by design goals of the project, by removing personally identifiable data from the statistical components.

Furthermore, in order to overcome any performance issues that result from weather conditions and other influencing factors, a shared matching biometrics service was proposed, which combines multiple biometric approaches.

References

COUNCIL, T. E. (2007). Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76. *Official Journal of the European Union*, OJ L 199, 31.7.2007, p. 23–29.

Frontex. (2018). *Frontex Risk Analysis for 2018.* https://data.europa.eu/euodp/data/dataset/ra-2018.

Voigt, P. a. (2017). The EU General Data Protection Regulation (GDPR). *Springer.*, Vol. 18.