D6.5 Meta Fusion and Risk Analysis Tool 1

Document Due Date: 28/02/2021 [M18]
Document Submission Date: 28/02/2021 [M18]

Work Package 6: Alternative Technologies to Identifying People

Document Dissemination Level:
Public

D4FLY - **D**etecting **D**ocument frau**D** and i**D**entity on the fly
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833704

Abstract

Part of Work package 6, Task 6.5, aims to develop a risk analysis framework based on a machine learning based metadata fusion approach to assist traveller identification at the border crossing point. The input of the risk analysis framework includes the metadata sources from other tasks within D4FLY including document check and fraud detection, biometric verification and presentation attack detection to anomaly and travel pattern detection. External data sources such as API and EES information can also be fused into the system. The output from the risk analysis framework could be presented in a universal/standard warning system to show a flag to the border guards to simply indicate the risk level and help them quickly identify any suspicious patterns or threats at the frontline. The developed framework should be adaptive to be deployed in different border scenarios and to suit different EU Member States' border checkpoints.

The work involves: firstly, to define a model for processing metadata, and secondly to propose a risk analysis framework based on a metadata fusion approach. The aim is to assist border authorities to plan and to execute tasks related to more efficient identification, and to reduce the risk of false negative assessments.

This document is the first deliverable of this task, focussed on the first phase of the development of Task 6.5 and describes the activities and development progress within the task. In summary, the main activities and progress accomplished during the first period (M10-M18) to be reported in this document include:

- Undertook literature review and background study on metadata fusion and risk analysis frameworks
- Conducted interviews with the D4FLY end users to help better understand the potential use of the technology to be developed and identify the requirements/needs from the end users when developing the system. Detailed results and analysis on the responses are provided
- Investigated different data fusion methods and selected a suitable fusion model for processing metadata
- Defined the architecture of the risk analysis framework
- Proposed the initial UI design for presenting the risk analysis results to the border guards
- Considered and discussed ethics, privacy and security issues

Project Information

Project Name	Detecting Document frauD and iDentity on the fly
Project Acronym	D4FLY
Project Coordinator	Veridos GmbH
Project Funded by	European Commission
Under the Programme	Horizon 2020 Secure Societies
Call	H2020-SU-SEC-2018
Topic	SU-BES02-2018-2019-2020 Technologies to enhance border and external security
Funding Instrument	Research and Innovation Action
Grant Agreement No.	833704

Document Information

Document reference	**D6.5**
Document Title	**Report**
Work Package reference	WP6
Delivery due date	28/02/2021
Actual submission date	28/02/2021
Dissemination Level	Public
Lead Partner	**UoR**
Author(s)	**UoR: Lulu Chen** **VTT: Sirra Toivonen, Niko Lehto, Anni Karinsalo**
Reviewer(s)	**First Review: James Ferryman, Laura Salmela** **Second Review: Martin George** **Third Review (security review): Dimitris Kyriazanos**

Document Version History

Version	Date created	Beneficiary	Comments
0.1	31.11.2020	UoR	Initial draft
0.2	15.12.2020	UoR	Updated structure
0.3	21.01.2021	UoR	Added content in various sections
0.4	27.01.2021	UoR	Added content in various sections, integrated input from VTT
0.5	03.02.2021	UoR	Added content in various sections
0.6	09.02.2021	UoR	Added input from VTT, made updates in various sections
0.7	13.02.2021	UoR	Added updates from VTT, made some minor updates in various sections, finalised text for first internal reviews.
0.8	16.02.2021	UoR	Updates based on comments from first reviews
0.9	21.02.2021	UoR	Further updates from first reviews; ready for second review

0.91	25.02.2021	UoR	Updates based on comments from second review; ready for security review
1.0	27.02.2021	UoR	Minor updates based on security review; final edits

List of Acronyms and Abbreviations

ACRONYM	EXPLANATION
API	Advanced Passenger Information
BMS	Biometric matching service
CIRAM	Common Integrated Risk Analysis Model
D4FLY	Detecting Document frauD and iDentity on the fly
EC	European Commission
ECRIS-TCN	European Criminal Records Information System
EES	Entry/Exit System
ETIAS	The European Travel Information and Authorisation System
EU	European Union
EURODAC	Information on European Asylum Applications
FastPass	A harmonized, modular reference system for all European automated border crossing points (EU FP7 research project)
GDPR	General Data Protection Regulation
IBM	Integrated Border Management
ISO	International Organization for Standardization
KPI	Key Performance Indicator
NISO	National Information Standards Organization
OSINT	Open Source Intelligence
PAD	Presentation Attack Detection
PNR	Passenger Name Record
PROTECT	Pervasive and UseR Focused BiomeTrics BordEr ProjeCT (EU H2020 research project)
SBC	Schengen Borders Code
SISII	Second generation Schengen Information System
SLTD	Interpol Stolen and Lost Travel Documents database
SOCMINT	Social Media Intelligence
TRESSPASS	robusT Risk basEd Screening and alert System for PASSengers and luggage (H2020 research project)
VIS	Visa Information System
VSATIS	State Border Guard Service Information System

Table of Contents

1 Introduction

The main objective of the tasks in D4FLY Work Package 6 – Alternative technologies to identifying people, is to explore alternative solutions for people identification. This deliverable reports activities within Task 6.5 – Meta fusion and risk analysis, which focuses on creating a risk analysis framework that supports determination of risk for an identifiable traveller during the border crossing. It also considers ethics, privacy and security issues.

Smartphones, tablets and other mobile devices are in everyday possession of many people and carry valuable data in and of themselves. In addition to the analysis of contacts, call lists, SMS, social networks and messages from messaging services, it is also possible to extract movement data, emails and passwords from most systems. For crimes and criminal acts, e.g. corruption, data theft, or cyber-crime, digital evidence plays an important role in the investigation. Travellers and digital documents produce a significant amount of metadata that can be analysed and checked for plausibility. This data could be especially valuable to fight impostor fraud at manual border posts without automated biometric authentication capabilities. Above all, Task 6.5 investigates and researches the possibilities of using metadata in the traveller risk analysis to assist border authorities to plan and execute identification tasks effectively and reduce the risk of false negative assessments.

The main objective of this task is to develop and test a universal system to automatically build a risk profile and determine risks related to identifying travellers by gathering and analysing information from a variety of sources, which includes document and travel metadata, as well as metadata derived from fused biometrics (Task 5.6), alternative technologies for identifying people (including output from Tasks 6.1-6.4), tactical and travel pattern anomalies (Tasks 8.4 and 8.5) and interoperable databases and other data sources available to the border authorities. Such a risk analysis system would be beneficial for enhancing the border security and speeding up the border check process. The output from the risk assessment based on metadata fusion could be presented in a universal warning system to show a flag to the border guards to simply indicate the risk level and help them quickly identify any suspicious patterns.

Risk assessment can heavily depend on the country and border types (e.g. air, sea and land border). An interview study with end users will be conducted during the work to better understand the needs when using automatic risk analysis based on information fusion to help with border check. The task aims to deliver a more generic risk analysis framework that is configurable by the authorities to meet the special requirements of different border checking situations/scenarios by each individual country and be able to select available/suitable input metadata sources. An alert system indicating the traveller risk level via a Graphical User Interface will also be developed based on the needs from the end users.

1.1 Background

Task 6.5 starts in M10 (July 2020) and ends in M30 (February 2022), and the contributors to the task are UoR and VTT. There are two deliverables from the task:

TABLE 1 DELIVERABLES OF TASK 6.5

Deliverable number	Deliverable title	Type	Dissemination level	Due date

D6.5	Meta fusion and risk analysis tool 1	Report	Public	M18 – February 2021
D6.10	Meta fusion and risk analysis tool 2	Demonstrator	Public	M30 – February 2022

The task will be developed in two phases and there are two main outcomes accordingly:

1) During the first phase, the state-of-the-art will be reviewed, and different strategies will be investigated. User requirements will be established. A suitable data fusion model will be selected and defined for processing and combining a variety of metadata. The architecture of the risk analysis framework will be defined. The GUI for the developed risk analysis tool will be designed.
2) During the second phase of the task, a risk analysis framework will be implemented that helps border authorities to plan and to execute tasks related to person identification, by automatically assessing the risk level of a traveller crossing the border more efficiently and hence reducing the risk of false negative assessments. The framework will be evaluated on both synthetic and real data. The GUI will be prototyped.

1.2 Aim of this document

This document introduces the first phase of the development for Task 6.5 and describes the activities and development progress within the task. In summary, the main activities and progress accomplished during the first period (M10-M18) to be reported in this document include:

- Undertook literature review and background study on metadata fusion and risk analysis frameworks (Section 2)
- Conducted interviews with the D4FLY end users to help better understand the potential use of the technology to be developed and identity the requirements/needs from the end users when developing the system (Section 4)
- Investigated different data fusion methods and selected a suitable fusion model for processing metadata (Section 5)
- Defined the architecture of the risk analysis framework (Section 5)
- Proposed the initial UI design for presenting the risk analysis results to the border guards (Section 5)
- Considered and discussed ethics, privacy and security issues (Section 3.5 and 4)

As this is a research task and also the dissemination level of the deliverable is public, several limitations have been identified during the study and presentation within the task and report. These limitations include:

- Data protection and ethical issues: obtaining access to certain types or sources of data, or associating data from different sources, may be restricted under GDPR or other regulations
- Security limitations: due to the dissemination level, some information cannot be reported directly in this document. This includes topics such as identification of specific risks, prioritisation of the risks as established in the European Commission

guidelines for the H2020 Programme on the classification of information in research projects[1].

- Data availability: there is a lack of real-life data that can be used in training, validation and evaluation of the system. The plan is to use synthetic data for initial testing.

1.3 Input and output

This task is closely linked to other work packages and tasks in the project. Tasks that provide direct input or are relevant to Task 6.5 include:

WP5:

- Task 5.6 – Biometric fusion

WP6:

- Task 6.1 – Using smartphone sensors for identifying people
- Task 6.3 – Smartphones based enhanced traveller verification
- Task 6.4 – Blockchain & Distributed Ledger

WP7:

- Task 7.1 – Detecting morphed faces
- Task 7.2 – Counter spoofing & presentation attacks detection
- Task 7.3 – Counter spoofing with additional sensors

WP8:

- Task 8.2 – Improving automated forgery detection in travel and identity document reading devices
- Task 8.3 – Document fraud detection for breeder documents
- Task 8.4 – Tactical anomaly detection in documents
- Task 8.5 – Travel patterns from passports

The output of Task 6.5 will be fed back into the overall D4FLY system:

WP4:

- Task 4.1 – Platform user interface design
- Task 4.2 – System Architecture

[1] https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/secur/h2020-hi-guide-classif_en.pdf

2 Related work on risk analysis and metadata fusion

This section focusses on the work that has been carried out in investigating state-of-the-art on risk analysis related to border control and exploring the potentials of using data science, especially metadata fusion, for developing a risk analysis approach to assist border authorities managing different border crossing types.

This section leads with a short background of the risk assessment and analysis principles via an introduction to the ISO 31000 Risk management standard principles. The general travellers' risk assessment and checking at border control is then introduced. The traveller risk assessment options are guided by the regulation and depend on information available through the means of travel. Frontex has contributed to the harmonised risk assessment principles by providing the CIRAM model. Information of the outcomes from the related previous projects (FastPass and TRESSPASS) that have developed automated risk analysis frameworks and tools is summarised in this section. As some specific deliverables from those projects concerning the metadata risk analysis or the risk analysis in general are not public deliverables, the description on those topics is generic. In this section, the state-of-the-art on data fusion has also been reviewed and summarised.

2.1 Risk analysis for border control

Standard ISO 31000 provides the ground on which risk management and assessment development [9] is based. It emphasises that risk management aims to create and protect value, improve performance and encourage innovation. The general aim of the risk analysis is to be involved in developing an understanding of the risk. It contributes to risk assessment[2] and decisions on whether risks need to be addressed/treated, as well as on the most appropriate handling strategies and methods. The methods used may be qualitative, semi-quantitative or quantitative. In general, risk management process consists of phases to identify the risks, to determine the consequences and probabilities of the identified risks taking into account the presence (or not) of existing controls, and their effectiveness and complexity and to develop mitigation measures. ISO 31000:2018 [9] guides the risk analysis to consider following factors: the likelihood of events and consequences, the nature and magnitude of consequences, complexity and connectivity, time-related factors and volatility, the effectiveness of existing controls and sensitivity and confidence levels. Risk analysis provides an input into decision-making where options involve different types and levels of risk [9]. In other words, the objective is to provide evidence-based information and analysis to make informed decisions on how to treat particular risks and how to select between options. The level of information required depends on the application, the availability of reliable information, and the decision-making needs.

The risk analysis in the context of border control (especially in D4FLY) focuses on the risk assessment made on the traveller at the time the border check takes place. Currently, the border control measures at the external borders constitute one tier in the European integrated border management (IBM) [41] whereas the other tiers (measures in third countries, measures with neighbouring third countries, risk analysis and measures within the Schengen area and return) support holistic border management of Europe. A holistic risk

[2] In the ISO 31000 standard, the risk assessment process includes three phases: risk identification, risk analysis and risk evaluation. Risk assessment is being followed by risk treatment. Before a risk assessment is initiated, the context where it is focused shall be established.

analysis could also consider the whole border check systems and processes; however, these are not in the scope of this deliverable.

In border control, the accurate identification of individuals and access to authentic and verifiable information are critical. Border authorities utilise intelligence and available information on travellers who are aiming to enter or exit the territory, in order to determine effectively those who are qualified to enter and those who must be stopped at the borders. The information provided by the travellers at the border come in the form of passport or ID-card and other needed documentation, their biometrics and oral or behavioural feedback form the basis of the risk assessment. Border guards also check the relevant databases for possible hits. Depending on the traveller's origin and the border type, there will also be various amounts of advanced passenger information available. Automated checks, in general, are used for low-risk travellers. In the manual lines, travellers are assessed by the first line border guards who can assess travellers' risks based not only through the border check (e.g. inspecting presented documents), but also on their behaviour observation and interview responses. Manual line inspection may also be provided with specific profiling information to steer the traveller assessment at certain points in the border control process.

To better identify individuals who present a higher risk the EU has established new border control related regulations during recent years. As an outcome of a longer preparatory actions two Regulations, Regulation (EU) 2019/817 and Regulation (EU) 2019/818, were established to address the interoperability of different European wide databases. These introduce enhanced schemes to identify persons who use false or multiple identities at the border or within the EU territory. Regulation (EU) 2019/817 defines the framework for EU interoperability between information systems in the field of borders and visa information systems; and Regulation (EU) 2019/818 in the field of police and judicial cooperation, asylum, and migration. The aim of the new regulation is to remove the problems caused by the current EU information systems concerning identity fraud, or other security threats, arising from the fact that the current databases do not communicate with each other, and to establish a new framework that allows for improved identification of a person that has already been recorded in one of the databases (e.g. VIS[3], EURODAC[4], SIS II[5], EES[6], ETIAS[7] and ECRIS-TCN[8]). This also includes biometric data. The common BMS (shared biometric matching service) will be built and it facilitates the identification of a person that has been registered in several databases by using a single technical component to compare a person's biometric data (fingerprints and facial images). The new systems will make targeted and intelligent use of the information in the databases. The framework includes development of several technical components that enable interoperability. The proposals have been built on a two-step approach which means that the initial searches are made on hit/no-hit basis and only if a flag is raised the law enforcement authorities can request further information [38].

To promote integrated border management and to enhance consistent information exchange in a structured way with different strategic partners and develop high-standard control of the borders across Member States and as part of Frontex strategic analysis at the borders, a

[3] Visa Information System

[4] Information on European Asylum Applications

[5] Second generation Schengen Information System

[6] Entry/Exit System

[7] The European Travel Information and Authorisation System

[8] The European Criminal Records Information System

Common Integrated Risk Analysis Model (CIRAM) has been developed [34]. It aims to serve as a strategic tool to harmonise risk assessments at European level and to develop a conceptual framework in preparation of risk analyses. CIRAM also supports the four-tier integrated risk management model. The risk in the border management context is defined as: *"a magnitude and likelihood of a threat occurring at the external borders, given the measures in place at the borders and within the EU, which will impact on the EU internal security, on the security of the external borders, on the optimal flow of regular passengers or which will have humanitarian consequences."* [12][35]

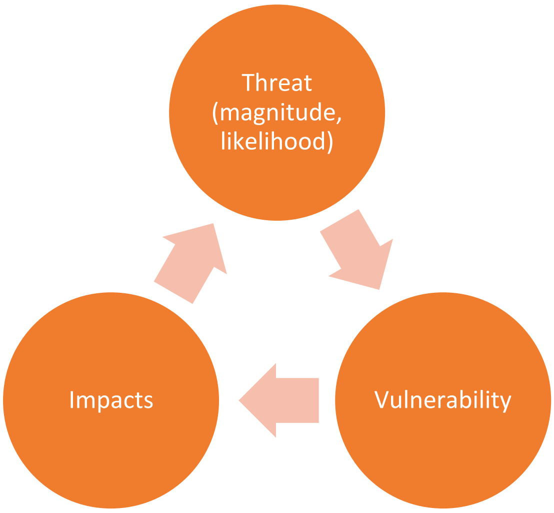


FIGURE 1 RISK IN THE BORDER MANAGEMENT CONTEXT [12]

In the CIRAM model, threat is defined as a force or pressure acting on the external borders, whereas vulnerability is defined by the capacity of the system to mitigate a threat. In the assessment of the vulnerability, the policies, geographical, technological or operational aspects may have an effect. Impacts may be analysed e.g. in terms of security effects on the societies, flow of passengers, humanitarian situations or disruptions of services. Intelligence that consists of collecting, analysing and distributing information is at the heart of the CIRAM risk analysis principles. The analysis of the available data or information enables forming interpreted outcomes of the pieces of information [12].

In addition, CIRAM risk analysis can be used to analyse situation at border crossing points, for example, by the European Commission to decide on operational priorities and the distribution of Community funding in the border control domain, but also to plan unannounced visits to inspect Member States' compliance with the Schengen Borders Code [13]. Frontex risk analysis considers border risks jointly including the most vital migration control and other border-crossing issues such as smuggling, terrorist crimes or trafficking [14].

2.2 Related projects

There are several EU projects that have developed technologies related to border risk analysis. Two representative examples which relate closely to D4FLY are described below.

FastPass project

EU FP7 project FastPass [2] established and demonstrated a harmonised, modular approach for Automated Border Control (ABC) gates. FastPass assessed possible risks in ABC systems and developed a framework for future harmonized security assessments. In addition, the project designed an alarming module based on fusing different input sources which includes results from biometric verification and fusion, biometric Presentation Attack Detection (PAD), video surveillance events (e.g. loitering, left luggage, tailgating) and document reading (passport reader). A rule-based fusion model was developed and evaluated using simulated data. UoR and VTT were the partners who collaborated on the FastPass task, so the model to be developed in D4FLY will be based on the results from FastPass. FastPass mainly focussed on combining output from biometric verification and video surveillance-based event detection. In D4FLY, a much wider range of data sources will be explored.

TRESSPASS project

EU H2020 project TRESSPASS [3] proposed a novel approach that links existing risk-based approaches into a multi-threat, multi-modality and four tier risk-based border management system-of-systems. The main objective of the project was to develop accurate risk indicators from available data and background information, to calculate a risk for each traveller and based on the risk, adjust security checks required for each traveller. The KPI of the security system performance development included efficiency, traveller satisfaction and operational cost reduction indicators. An analytic framework was proposed for modelling risk and a systematic approach was developed for quantifying risk, based on a set of indicators that can accurately be measured across all four tiers of the Integrated Border Management. Data fusion is applied in the tiered system for risk assessment.

TRESSPASS optimised various risk indicator estimations aiming to increase their accuracy with higher levels of confidence [36]. The performance indicators of the automated risk analysis included: effectiveness, flow rate, efficiency and level of ethical compliance. As part of the output, the project proposed an automated real-time risk assessment for airport passengers using deep learning. Data fusion was based on the output of the sensing devices, especially on the surveillance cameras. TRESSPASS also listed some identified challenges or restrictions in developing the automated risk assessment system, which includes for instance, use of data that is GDPR compliant, definition of risk/anomalous behaviour, applicable sensors, cost-benefit consideration of the additional risk assessment [37], etc.

Task 6.5 of the D4FLY project has a clearly different scope and focus from above-described projects. This task firstly will deliver a multimodal metadata fusion framework that combines a variety of metadata sources using a selection of fusion schemes to assess a traveller's risk level. Secondly, the metadata input source covers a wider topic and security issues for border crossing including metadata derived from document check, biometric verification and travel patterns, etc. External input sources such as (Entry/Exit System) EES and video surveillance events will be considered, however, the main types of data sources are from the other tasks developed in D4FLY. Finally, there is a focus on the explainability of the risk scores generated, specifically which risk factors / data inputs to the process contributed to the final risk assessment.

2.3 Metadata fusion

According to the United States National Information Standards Organization (NISO), metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information [16]. The term metadata is used differently in different scenarios. There are several types of metadata and the three main types of metadata defined in the literature include [17]:

- **Descriptive metadata** describes a resource for purposes such as discovery and identification
- **Structural metadata** indicates how compound objects are put together
- **Administrative metadata** provides information to help manage a resource

In the context of D4FLY, the term metadata refers to the structured/processed data rather than the raw data obtained directly from e.g. the sensor capture. A fusion process combines a series of metadata from multiple sources into a single estimate (deterministic or various clusters or groups of manageable estimates) expected to be more accurate and informative than multiple sources [1].

2.3.1 Metadata fusion for border control

The aim of the metadata fusion in D4FLY is to develop tools and intelligence to enhance the interpretation of the results provided by different D4FLY technologies and tools during the border check situation. An additional aim is to support the border guards in the interpretation of results in hectic checking situations and to enhance the introduction of additional sensors and technologies in border control. The final aim is to introduce intelligence to the interpretation of results in order to be able to interpret also those situations where the sensor output is not self-evident but when combined with other information gathered at the border check situation may show that the traveller is of low risk or that further analysis should be made.

The D4FLY metadata fusion analysis is not intended to replace border guards but rather to support their interpretation of the available additional pieces of information efficiently. Additionally, by applying machine learning or data science techniques, an automated risk analysis process is presumed to be able to increase accuracy and reduce false negatives. Risk analysis of a traveller currently mostly relies on manual check, thus, using an automatic system that combines a range of factors would potentially help increase the speed and efficiency of the process.

2.3.2 Metadata fusion schemes

Data fusion has been applied in multimodal biometric systems for border control previously [18]. The fusion of evidence from various biometric modalities can be performed at sensor level, feature level, score level and decision level [27][28][29]:

- Sensor level fusion: combines biometric traits from multiple sensors prior to feature extraction. Sensor level fusion is useful in multi-sample systems, when a sensor can capture two or more samples of the same trait and create a more accurate description for that trait
- Feature level fusion: combines information from multiple features sets extracted from different samples. Feature level fusion normally needs to deal with large dimensionality and variance of feature sets, especially when different traits are combined
- Score level fusion: combines multiple likelihood in the form of matching scores using different fusion schemes. Previous work claims that score level fusion is more effective and produces better matching then other levels of fusion [27]. In score level fusion, selection of a normalisation scheme is an important step for obtaining good performance

– Decision level fusion: combines either the decisions of separate algorithms, or decisions made separately on different evidence. Decision level fusion has the least complexity but less effective than score level fusion due to the limited amount of information it can provide at the fusion stage

There is not a universal fusion method that can always provide better results than the other methods. Various factors, such as the biometric traits being fused and data quality, can affect the result. Score-level fusion has been more widely applied in all biometric fusion applications and is generally preferred because it offers the best trade-off in terms of the information content and the ease in fusion [30]. Marasco and Sansone [31], in this context, conducted experimental comparisons on different fusion methods and claimed that adding biometric traits to the fusion does not necessarily increase the performance.

Where biometric metadata is combined with other types of metadata, there is a lack of literature or standardised methods to combine these sources. For example, there exist no joint standards for the joint fusion of results from spoofing tests, different forms of quality information, results from document verification process related to anomaly recognition, and no universal standard on the entry/exist stamps. It is therefore important to address in this work how to quantify different types of data in a standard approach that can be fed into the fusion system, e.g. can all types of data be normalised into a score as biometric verification.

Parallel and serial fusion

There are currently two types of fusion modes: parallel fusion mode and serial (or sequential) fusion mode. The former fuses the information of all traits in the system simultaneously, while the latter uses traits in the system one by one in sequence. Most previous work in the literature has focussed on parallel fusion at feature-level, score-level, decision-level or sensor level. Parallel fusion requires that all data sources need to be always available for each user at the border crossing point. In contrast, the serial fusion mode usually provides more flexibility such as how the input data sources can be arranged and ordered and adopt different parameters in the chain.

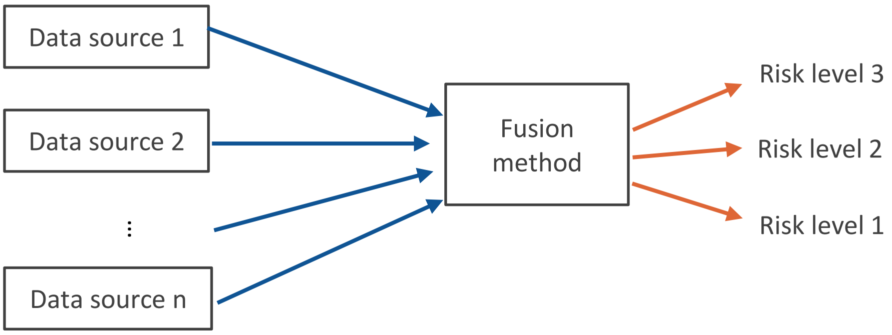


FIGURE 2 A TYPICAL PARALLEL FUSION PROCESS CHAIN. ALL BIOMETRIC TRAITS ARE FUSED SIMULTANEOUSLY INTO THE FUSION METHOD IN ORDER TO OUTPUT A DECISION.

In the serial fusion mode in the context of biometric verification, the users go through the authentication process stage by stage. At each stage, a certain type of trait is sampled and matched against a template library. Once the user passes the authentication at a certain stage, all the remaining stages can be by passed, i.e. most users do not have to go through the whole chain of stages for the authentication [32]. Thus, serial fusion mode usually can provide more flexibility in ordering the traits and parameterising the corresponding matchers in the chain, hence, more user convenient. Serial fusion of multiple matchers represents a good trade-off between the widely adopted parallel fusion and the use of a mono-modal verification system [27]. Marcialis et al. claimed that the one advantage of serial fusion over parallel fusion is that the majority of genuine users should be accepted by using only one biometric, i.e. the first one in the processing chain (this can be particularly true if some partitioning of users is possible [33]). Figure 3 illustrates the process chain of a typical serial fusion system.

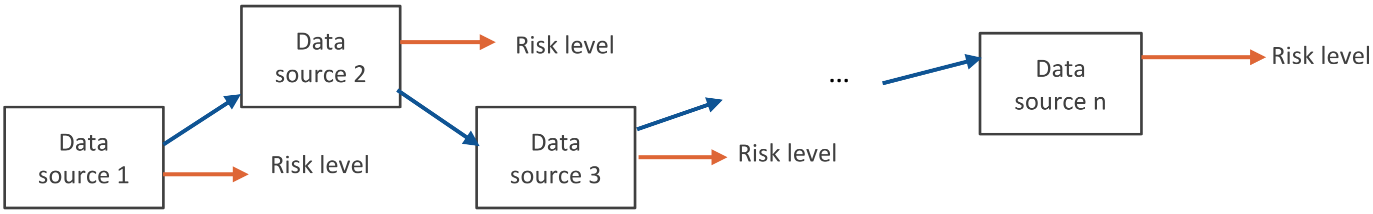


FIGURE 3 A TYPICAL SERIAL FUSION PROCESS CHAIN

Rule-based and learning based fusion

1) Rule-based fusion

A rule-based system is a system where input is processed with defined rules to produce an output. An example of a simple rule is: if a fake travel document is presented or presentation attack detected at biometric verification process, a high-risk level should be determined, and an indication of the risk level should be presented via the border guard system. As an output, the metadata fusion module will initiate an alert message if a rule is matched against the given data/information. For instance, when everything is normal/legitimate, the biometric information indicates that the traveller is 98.9% the person as he/she claims to be, and the video surveillance indicates normal traveller behaviour. Applying the risk analysis procedures (according to the desired threshold limits), the decision can be made automatically, and the traveller may cross the border. Alternatively, if some of the previous checks would have failed, the fusion based risk assessment provides a message indicating the issue to be flagged as an alarm (e.g. biometric match does not exceed the given threshold).

2) Machine learning based fusion

An Artificial Neural Network (Neural Network, NN) is a computing system of artificial neurons, inspired by the brain's biologic neural networks. Neurons are structured into layers, in which the first layer is called the input layer and the last layer is called the output layer. The depth of a neural network influences its ability to detect complex features.

The most common approach to train neural networks is called supervised learning. It means the system learns by processing examples, in which each entry contains a known input and a corresponding result. The process is iterative and after training the neural network is able to generalize and classify new inputs to pre-known result categories. Applications of the neural network-based fusion is limited into use cases where a collection of pre-defined input-result entries are available. For example, neural networks are used in different tasks related to statistical pattern recognition, including speech recognition, image recognition, in which supervised learning material is possible to be gathered beforehand.

In the metadata fusion module, a neural network based fusion algorithm could be trained for a specific task in specific control point which takes multiple data sources as inputs and provide one output for the rule-based system. In other words, neural network based fusion algorithm can operate as a non-rule-based fusion that is incorporated into the rule based fusion algorithm.

2.4 Summary

Efficient and effective border control is a key request to safeguard the security and mobility of the citizens in the EU. Border guards must fulfil their responsibilities to protect the EU's borders from the threats posed by the travellers while at the same time supporting the societies, business communities and travellers' rights by facilitating fair, efficient and speedy border processes. The preceding sections have reviewed previous work and possible approaches to develop traveller risk analysis so that automated assessment could be achieved to support the border guards. This section has especially considered potential approaches to develop metadata analysis using data fusion schemes to support the interpretation of the alternative technologies.

Also, combining all the available data sources using a machine learning based fusion method would potentially help increase the accuracy and reduce the risk of false negative assessments. Such a system would also potentially increase universality across different border types across different member states, and lead to contribution to standards on certain types of metadata.

To generate such a risk analysis model for border control, a series of challenges need to be addressed:

1) Firstly, what kind of metadata sources are available to use for different border crossings and which type of metadata matter most to each border scenarios?
2) As there is no universal standard for all types of data, how to use the data that are not normalised or standardised?
3) Although this is a research task to investigate the possibilities of using all sorts of metadata to assess a traveller's risk, under current rules, obtaining access to certain types of data is probably restricted under GDPR or other regulations. This aspect should be taken into account while developing the system
4) To identify the types of risks or threats that the border authorities consider as important
5) How to present the risks to the border authorities in an efficient and clear method?
6) What kind of data fusion model fits best with the border control scenarios?
7) How to test, validate and evaluate the developed system? What kind of data would be available for this purpose?

These questions and challenges are addressed in this deliverable in the following sections.

3 METADATA SOURCES IN D4FLY

The main point of data fusion is to combine all available data sources together to automatically detect and assess a traveller's risk. D4FLY project covers topics from document check to biometric verification. This section introduces the metadata types that are relevant to the D4FLY project and can be potentially used in the risk analysis framework and includes feedback from the end users (Section 4). There are also relevant external data sources that have also been identified and described in this section.

TABLE 2 RISK ASSESSMENT EVENTS RELEVANT TO D4FLY AND THEIR RELATED BORDER SCENARIOS

No.	Type	Metadata source	Metadata as input
1	Document check	Automatic document forgery detection	Forgery detection score/result (i.e. pass/fail/invalid and a confidence score)
2	Document check	Breeder document fraud detection	Document matching score/result (i.e. pass/fail)
3	Document check	Face morphing detection	Morphing detection score/result
4	Biometric verification	Biometric fusion process	Biometric fusion score/result
5	Biometric verification	Biometric PAD	PAD score/result
6	Biometric verification	Biometric verification based on smartphones	Verification score/result
7	Anomaly detection	Tactical anomaly detection	Anomaly detection result (i.e. result on a predefined rule: correct, ignored, incorrect)
8	Travel pattern	Traveller travel pattern by automatic international stamps extraction	Travel pattern (i.e. country recognition, in/out recognition, data recognition)
9	External data sources	To be considered: for instance, interoperable databases, video surveillance events, etc.	External (e.g. a database result can be a match/no match)

3.1 Document and travel metadata

The first type of input comes from the travel document check. Three types of metadata can be obtained from the travel document inspection process in D4FLY.

Automatic forgery detection (1)

Task 8.2 – Improving automated forgery detection in travel and identity document reading devices focusses on developing tools to enhance automatic document inspection, specifically automatically checking Kinegrams® and other optically variable features embedded within different document types. The outcome from the automated forgery detection process is firstly a recognition of genuine or fake travel documents, and secondly distinguishing original documents from printed copies. The metadata obtained from this module to be fed into the risk analysis framework will be the forgery detection results (e.g. detection score and result).

Breeder document fraud detection (2)

Task 8.3 – Document fraud detection for breeder documents develops solutions to detect fraud in breeder documents (e.g. birth certificate and marriage certificates) which are used as proof of identity in document issuance or immigration processes. Manual analysis of breeder documents is time-consuming and requires specific competencies from authorities, as there is no single standard for breeder documents in general and persons can provide them in their original language. This task seeks to use machine learning techniques to automatically recognise documents and retrieve similar details and the appropriate reference documents to detect document fraud.

Face morphing detection (3)

Recent studies have shown that an intermediate frame in a morphing (transforming and blending) between two face images of different people can deceive commercial biometric verification systems to match both faces with a single morphed image [10], and even trained humans can be fooled by such morphed images [11]. Task 7.1 – Detecting morphed faces develops different approaches/algorithms to automatically detect morphed face images during passport checks. The results (e.g. detection scores) from face morphing detection can used for risk assessment.

Detecting forgery or fraud in travel documents can indicate a high-level risk. If these kinds of attempts would be detected during first line border check, the traveller would be sent to the second line for further examination.

3.2 Biometric verification and fusion

Another main focus in D4FLY is biometric verification in border control. The data from biometric verification process can be fed into the risk analysis framework.

Biometric fusion results (4, 6)

Task 5.6 – Biometric fusion combines all the individual biometric verification results and produces a verification result. The biometric fusion module produces a single biometric verification result by combining verification scores from multiple biometric modalities (Task 5.1 – 5.4). The output from biometric fusion is a result from direct live identity check at the border.

This also includes the outcome from WP6 tasks on using smartphone technologies for identifying people, which will be combined into the biometric fusion process.

Presentation attack detection (PAD) (5)

Task 7.2 and 7.3 focus on developing solutions on automatically detecting biometric presentation attacks. The output from each biometric PAD module will be PAD score indicating the possibility of an attack. The PAD score and its corresponding biometric module name will be the input for the risk analysis framework. If a presentation attack is detected, a high-level risk score should be assigned. The detailed information on the potential attack detected can be shown on the border guards' screen. The traveller should be sent to second line check.

3.3 Tactical and travel pattern anomalies

<u>Tactical anomaly detection</u> <u>(7)</u>

Task 8.4 – Tactical anomaly detection in documents focuses on recognizing anomalies at a tactical level. For example, an otherwise normal document contains contradictive information in relation to stamps, dates or names present in the passport. A combination of bottom-up learning and top-down rules is used to recognize tactical anomalies. The anomaly detection results (e.g. a detection score or result) along with detailed information on the detected event will form the input to the risk analysis framework.

<u>Travel pattern</u> <u>(8)</u>

Passports include visa pages with stamps with entry and exit information of travellers. This travel information is commonly used as one of the indicators of risk, which may lead to more thorough inspection of the document. There have been attempts to recognize stamps [4] but these approaches assume that the stamp is a plain colour object on a monotone background. Task 8.5 – Travel patterns from passports focusses on automatically analysing international stamps on visa pages to extract travel information and developing a tool to extract information from stamps in the passport. The extracted travel information can indicate the traveller's travel pattern/history which can be very useful for detecting potential risks.

3.4 Other types of input source

The metadata types described above have been addressed in the D4FLY project and technologies are being developed in other D4FLY tasks as described above. In the metadata fusion task, the focus will be to use these types of input sources. However, the design of the data fusion system will also take into account other/external types of metadata (metadata group 9) that can be available to the border authorities and be included in the metadata fusion process to enhance the overall risk assessment of travellers, for instance:

- Results from video surveillance: in the FastPass project [2], video surveillance events (loitering, left object, tailgating within eGate, etc.) were one of the main data sources used in the developed alarming module
- Interoperable databases such as, Entry/Exit System (EES) information, API (Advance Passenger Information) information, Interpol Stolen and Lost Travel Documents database (SLTD), Car registration number check (e.g. for stolen cars) for land border crossing check, and PNR (Passenger Name Record), etc.

These external data sources will not be the focus in D4FLY, as using these data may be subject to legal challenges or restricted by GDPR or other data protection rules, which is not in scope of this task.

3.5 Data protection and ethical issues

Developing automated risk analysis tools for border control may raise new ethical concerns and therefore ethical issues must be carefully considered throughout the development process. In border control, ethical issues are varied and as the technologies advance, the ethical considerations must follow. In data fusion, ethical concerns may be related to data sources or what kind of fusion logic will be implemented. Depending on data sources, the data reliability should be carefully considered (e.g. open source data). Open Source Intelligence (OSINT), where information and knowledge are gathered from publicly available sources, is a growing field in the security domain as well as SOCMINT (Social Media Intelligence) where information available on social media networks is analytically exploited.

When new technological tools are considered for border checks, they should be always evaluated from the perspectives of necessity and proportionality. To avoid risks, careful attention of the compliance with the Schengen Borders Code (SBC), fundamental rights and data protection regulations should be made [39].

In the D4FLY project, the metadata used for the fusion will only include the data sources that are available from other tasks in the D4FLY which have been identified and summarised in the text (Section 3.1, 3.2 and 3.3) above. These data types represent the information that can be collected at the border check including passport control and the border check corridor solutions. Other types of data such as interoperable databases will be taken into account when designing the approach and algorithms and only synthetic data of these external data types will be used for evaluation purposes if necessary. As this task progresses in the project, an ethical impact assessment will continue to identify any relevant privacy and ethical concerns. These concerns will be discussed with partners and mitigations will be developed. The reader is referred to Deliverable D3.3 [42] for an initial assessment and plan for further steps.

4 USER REQUIREMENTS STUDY

In the context of Task 6.5, interviews with the D4FLY project end-users were organised. The sessions were held in January and February 2021. The aim of the interviews was to collect viewpoints and feedback from end-users both at a general level with regards to metadata fusion based traveller risk analysis, and specifically focusing on individual research objectives within Task 6.5, such as prioritisation of the risks. This section introduces the methodology adopted for the end-user interviews and provides a detailed summary of the outcome from the interviews.

4.1 Methodology

The interview questions and interview structure were designed together by the contributors to this deliverable (i.e. researchers from VTT Technical Research Centre of Finland and the University of Reading). The interview questions were organised according to the following thematic structure:

- **State of the art** encompassing current trends and challenges in metadata fusion based risk analysis
- **General guidelines and viewpoints** focusing on applicability to different border check scenarios and specification of requirements for the development of a risk analysis module
- **Additional points** identifying relevant metadata and input sources and specification of the border guard graphical user interface

During the interviews, the public dissemination level of the deliverable along with interview results were emphasized. Therefore, no specific risk level or detailed risk prioritisation was discussed during the sessions.

The results presented in the following sections are a synthesis of viewpoints and comments expressed by each end-user in the interview sessions. In the data analysis phase, the interview themes were redefined to better fit the responses from the interviews. In both the data analysis and reporting phases, the responses of individual end-users were anonymised with no real names or links to identity being revealed in the deliverable. Together with the formal security assessment process of the D4FLY project, the end-users reviewed and accepted the edited contribution prior to inclusion in this deliverable.

4.2 Results of end-user interviews

4.2.1 Configuration of risk analysis model according to different border environments

Border type plays a significant role in distinguishing what kind of metadata-based risk analysis model can be used for analysing traveller risk at first line border checks (i.e. suitability and feasibility of different models). Also, larger traveller groups in certain travel modes can be considered in general to be low risk (e.g. passengers on cruise vessels). Border types (air, land, sea) and the location of the border crossing point heavily affect which kind of data is or could be available and how the data can be accessed prior and during first line border checks. The availability of different data from a wide range of systems determines the scope and method for risk analysis that is pragmatic and feasible for each border crossing point. In a more controlled and static environment[9] (e.g. an airport terminal), the spectrum of available systems accessible for risk analysis purposes is much higher than in a less controlled environment. A rough schematic for a modular risk analysis approach is provided in Figure 4.

In different border types, also the presence and availability of data from other law enforcement authorities (e.g. customs) differ. In some countries, one authority may perform both border guarding and customs functions at a border crossing point, while in others, there may a clear organisational distinction between 'a border management authority' and 'the customs' [19]. Also, overlapping responsibilities are possible.

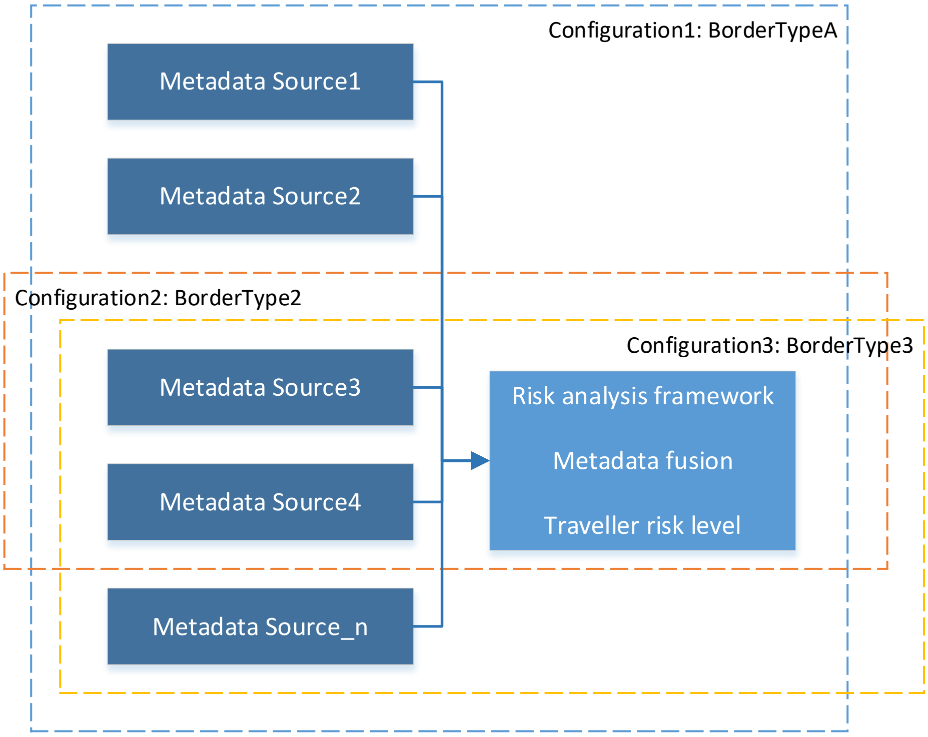


FIGURE 4 MODULAR APPROACH TO METADATA FUSION IN DIFFERENT BORDER ENVIRONMENTS

4.2.2 Data availability and current data fusion practices

Database queries to national or EU wide databases play a significant role in current risk analysis processes (e.g. Second generation Schengen Information System (SISII), Visa Information System (VIS), national databases, Entry/Exit System (EES) of the State Border Guard Service Information System (VSATIS), Interpol Stolen and Lost Travel Documents database (SLTD) [20] etc.). At the first line, the database query process starts when the document(s) of a traveller is scanned. Database checks provide **a match/no match** response to the workstation/graphical user interface of a first line border guard, and depending on which database produces the alert, next steps are decided (e.g. necessity to send a person to second line). According to document number and issuing country, document reference system may provide certain document samples as pop-ups to a border guard's user interface.

[9] From a border control perspective, a controlled environment can be understood as an enclosed area whose border check relevant parameters can be largely regulated or monitored (e.g. environmental conditions, passenger flow).

Depending on border type, border guards may receive analysed Advance Passenger Information (API) [21] or Passenger Name Record (PNR) [22] information. PNR data may contain, for example, the following information on the traveller:

- Dates of travel and travel itinerary,
- Ticket information,
- Contact details like address and phone number,
- Travel agent,
- Payment information,
- Seat number and baggage information

Recent developments in border checks of cruise passengers include new systems that enable digital information communication of the passenger lists and maritime vessel routes. Booking information collected by stakeholders other than air carriers, such as shipping companies, could also be considered useful for traveller risk analysis. Booking information for a cruise may include similar data as PNR, such as contact details, flight details (e.g. if arriving to a cruise from another country), payment details and information on travel companions. This information is available before a traveller embarks a vessel. In the current setting, this information is not used for risk assessment at the first line. The current legal framework does not permit the use of this data.

In addition to these, technologies used for manual and automated border checks may provide information for traveller risk analysis (e.g. travel document scanners, biometric sensors, video surveillance). There is also a range of sensors available to assess the behaviour of the traveller in different ways (e.g. body temperature, breathing, heartbeat, voice)[10] . However, there are different viewpoints in determining which kind of traveller behaviour can be considered as abnormal in different border environments and modes of travel.

At the moment, there are no fully automated data fusion processes involved in processing various information at the first line border check; in some environments, there may be partial automation included in the risk analysis, such as the processing of freight information. As an example, an automated data fusion process could mean the combination of travel document information with information about travel destination(s) or previous journeys. Also, flight routing information[11] could be considered. In a sea border environment, important risk indicators may also be visited ports, home port and other voyage related information. The situation is the same for alerts that originate from sensors or devices/equipment used for border checks, such as passport scanners and related background systems. Each result is examined individually. Together with different potential alerts originating from databases or sensors (e.g. video surveillance), the behaviour of the traveller is a key input on how the information is interpreted. Overall, certain data sources are considered as a challenge from risk analysis perspective. Although very important, PNR is viewed as a complex dataset, and there are a number of limitations for achieving higher automation in in the analysis of PNR data.

[10] The development or integration of these type of sensors in not within the scope of the D4FLY project.

[11] Information about a traveller’s departure point and possible waypoints/stop-overs on a route to the final destination.

4.2.3 Categorisation of risks

To categorise risks, end-users support the use of simple and commonly applied models, such as those based on 'traffic lights'. In a traffic light model, risks are labelled according to three colours, which indicate the following in the context of border checks at the first line:

- Green: Low or no risk
- Yellow/Amber: Necessity to perform additional checks or seek further information/a thorough examination in second line control
- Red: Necessity to check a person on the second line and/or detainment

In developing new ways to perform traveller risk analysis, it should be kept in mind that automation should not decrease necessary human vigilance nor support overreliance on automated solutions, especially in the low or no risk cases. With regards to risk category *yellow*, an assessment is being made whether additional checks can be performed at the manual booth at the first line or the person is escorted to the second line for further investigation. In addition to plain colour categorisations, a confidence score could be attached to each category. The confidence score would indicate for example how many percentages above green an analysis result is. A key issue is to ensure that the result display is understandable and sufficiently explanatory. The output could be presented to the border guards as a pop-up window with the alert and links to additional information.

4.2.4 Challenges for further implementation of automated risk analysis

If the level of automation is to be increased in risk analysis, the main limitations originate from legislation, namely the Schengen Borders Code (SBC) [23] and the General Data Protection Regulation (GDPR)[24]. Both regulate the collection, use, saving and sharing of personal data and other traveller related information prior to and during border checks, and influence the sharing of information among key associated organisations, such as immigration authorities, customs, facility operators and commercial actors within a border crossing point. Also, the use of data stored in different databases is clearly regulated and can only be done according to a predefined purpose. For example, fingerprints stored in the Entry-Exit System (EES) can be only used to confirm identity in the context of a border check. They cannot be used for other purposes (e.g. database cross-referencing). An important aspect to consider is also proportionality – non-severe misdemeanours or crimes visible on authorities' databases are not a cause to direct a person to second line investigation or prohibit a person from crossing EU's external borders. With respect to metadata, it is often not considered as personal information and thus not being regulated by the GDPR. However, there may be certain exceptions to this[12]. Overall, the sharing of metadata between different stakeholders might be possible and support better and effective interagency cooperation at border crossings.

It needs to be noted that applicable legislation or regulatory basis changes during a border check if a person is directed from first line border check to second line investigation. At the second line, GDPR no longer applies. Instead, activities conducted at the second line are

[12] "Under the GDPR, an identifiable person is someone who can be identified — either directly or indirectly —by their name, an identification number, or their geolocation data. This information could potentially be embedded within the metadata of any file shared online. While it might not be immediately visible to the human eye, these metadata could be extracted or read." [25]

regulated by law enforcement directives[13], also including particular administrative procedures and duties that the authorities need to perform.

Technically, the transcription of traveller biographic data across different languages and alphabet systems is seen to cause challenges for risk analysis, as information may be stored differently to databases or documents. For example, different lettering may be used to spell a person's name in different official documents (e.g. name/surname in a driving licence vs in a passport). The challenge is that currently there is no unique identifier for a person that would ensure correct identification. At the EU level, there is an initiative to develop a Common Identity Repository (CIR) [23] to facilitate the situation.

Before a combination of biometric identifiers (like face and fingerprints in EES) is commonly used as a unique identifier, a metadata risk analysis system might produce an unacceptable number of false positives. Therefore, a key priority is to investigate that proper risk indicators are used as source information and that the analysis tool relies on verified data. Depending on data source, all background information stored in different systems may not be accurate or correct (e.g. information provided when booking a travel). One possible solution is to use a confidence score that would indicate the reliability of the information provided by different data sources. Data sources could also be prioritised or ranked, and a feature importance score might be used for that purpose. Depending on context, different pieces of information on a traveller's booking details may trigger an alert (e.g. time of booking, means of payment, services purchased on board).

4.2.5 Future outlook

Today, risk assessments or risk analyses are supported by traveller risk profiles which are compiled and updated on a regular basis depending on identified changes in certain risk factors or parameters (e.g. travel trends, changes in the origin of used travel documents). The risk profiles are distributed to on-duty border guards in daily briefings or in other intervals. Current risk profiles are thus available and applicable for a certain period of time; there are no 'real time' risk profiles. In the future, the metadata of the document could be used to assign automatically the traveller or document to the risk profile and to inform the border guard who would take the necessary decision.

Risk assessments often follow a shared model, such as CIRAM 2.0 developed by Frontex. Also, the legal framework is an important driver. The end-users see a potential in fusing data used for the creation of risk profiles in the current setting. Also, the utilisation of a broader range of data sources would benefit traveller risk analysis at the first line. The availability of data from different sources needs to be considered. From example with regards to PNR, air carriers are obliged to "transfer PNR data… (a) 24 to 48 hours before the scheduled flight departure time; and (b) immediately after flight closure, that is once the passengers have boarded the aircraft in preparation for departure and it is no longer possible for passengers to board or leave" [24]. Often, air carriers prefer the latter 'wheels-offs' option that may in short flight distances[14] limit risk assessment possibilities to certain extent. The most important thing is information collection time from data input to recommendation output. Automation may also help detecting trends that are not yet identified in manual investigations.

[13] There may be national differences between EU Member States on which directives or regulations are applicable.

[14] Flight time being for example 30 minutes.

A preferred solution would be so-called 'up stream' risk analysis which would allow the denial of travel before a person embarks on a journey and arrives at a border crossing point. This applies, however, mostly to air travel and potentially to other cases in which travellers are transported across borders by carriers (e.g. train connections). The implementation of the European Travel Information and Authorisation System (ETIAS) [25] system may facilitate the situation at land borders and other border environments where typically no or little prior information of border crossers is received in advance. On the whole, automated risk analysis could be applied to all border checks processes at all types of BCPs. In building automated risk analysis, one also needs to take into account the dynamic nature of risk trends. In other words, effort should be made in creating a living system which can accommodate changes in risk trends and patterns over time, both immediate and long-term. Due to sudden changes affecting border environments and border crossing points, risk trends may change even within a 24-hour period.

With regards to travel documents, pages including country stamps and other information are considered important for risk analysis. Fully automating the extraction of stamp information is challenging, and depending on the presence and the quality of the stamp not always seen as feasible. Work package 8 (Document Verification) of the D4FLY project is particularly focusing on these aspects. Data from the RFID chip of the travel document, holder picture and fingerprints are also considered important sources of metadata.

Even though metadata fusion allows the handling of a significantly wider amount of data sources, human capabilities continue to bear relevance as only a little nuance in the behaviour of the traveller during discussion, detected by border guard, may lead to additional checks. Additionally, there always needs to be a human element checking the result before an intervention is made. This is also in accordance with EU ethics rules which highlight that the final decision must be made by human (border guard): travel risk analysis system (and metadata) could be provided for border guard only as recommendation. Therefore, human judgement should always be used to evaluate the result(s) of any automated risk analysis. The result may be used for example as guidance towards the questioning of a traveller and support the focusing of border guard attention to certain aspects in the person's travel details (current and/or prior history). All procedures should be very clearly described, and it should, in any case, be clear to the border guard what to do in each situation.

With automation, the assessment of intent is considered very difficult. In other words, what a traveller might do after entering or exiting a country. For first time travellers, this is particularly challenging, as they do not appear in any consulted systems. In the long run, any automated solution should not lead to the deskilling of border guards of their key competencies and knowledge in detecting persons of interest.

4.3 Summary of requirements

The set of requirements extracted from the end-user interviews are summarised in Table 3.

TABLE 3 SUMMARY OF END-USER REQUIREMENTS FOR METADATA FUSION BASED RISK ANALYSIS.

Requirement theme	Requirement description
Configuration of risk analysis model according to different border environments	The risk analysis module should be configurable to different border types (primarily air border BCPs, land border BCPs, sea border BCPs). The module should enable the inclusion or exclusion of metadata sources in the data fusion process depending on the availability of specific data at a border

	crossing point or a place where border checks are being performed (e.g. harbour for pleasure boats). The inclusion or exclusion of metadata sources should not affect the outcome of the risk analysis process. The rules that trigger an alert in the data fusion process should be modifiable according to the risks to be detected in a specific implementation environment.
Data availability and current data fusion practices	The risk analysis module should be able to integrate and/or interface with multiple types of metadata sources, e.g. databases of national law enforcement authorities, European law enforcement databases, global databases (e.g. Interpol), data transfers from and/or databases of commercial carriers or other commercial actors (e.g. PNR), sensor-based data (e.g. document/biometric verification processes, video surveillance). The risk analysis module should identify and accommodate internal complexities of different metadata sources or datasets being used in the risk analysis process.
Categorisation of risks	The risk analysis module should apply a standard model for the categorisation of risk (e.g. in a scale of three with colouring similar to traffic lights). A confidence score could be attached to each risk category to indicate confidence in the result. The result display should be made understandable and sufficiently explanatory. The output could be presented to the border guards as a pop-up window with the alert and links to additional information.
Challenges for further implementation of automated risk analysis	The legal compliance of the metadata fusion based risk analysis has to be carefully assessed as regulations may limit the use of particular metadata sources for the purposes of first line border checks, particularly those that are external to a border management authority. The risk analysis module needs to accommodate challenges related to the accuracy and reliability of data stored in different databases and systems. A confidence score could be used to indicate confidence in each metadata source. A feature importance score could be used to prioritise or rank different data as all metadata sources are not of the same importance depending on context.
Future outlook	The risk analysis module needs to accommodate variance in data availability from different metadata sources (e.g. the timeframe for data transfer and access may differ across sources, and it may not be possible to influence to this for the benefit of the risk analysis).

	The risk analysis module should be a living system which can accommodate changes in risk trends and patterns over time. The risk analysis module is a tool to support, not to displace, the work being done by border guards. There always needs to be human oversight to the risk analysis result before an intervention is made. All follow-up procedures proposed by the risk analysis module should be very clearly described, and it should be clear to a border guard what to do in each situation.

5 RISK ANALYSIS FRAMEWORK DESIGN

In this section, the detailed design of the proposed risk analysis framework will be described, including the framework architecture, risk indicators, graphical user interface (GUI), and metadata fusion schemes.

5.1 Risk analysis framework architecture

As summarised at the end of Section 2, there are several stages to be considered when designing the system. Figure 5 illustrates the stages of a risk analysis cycle [26].

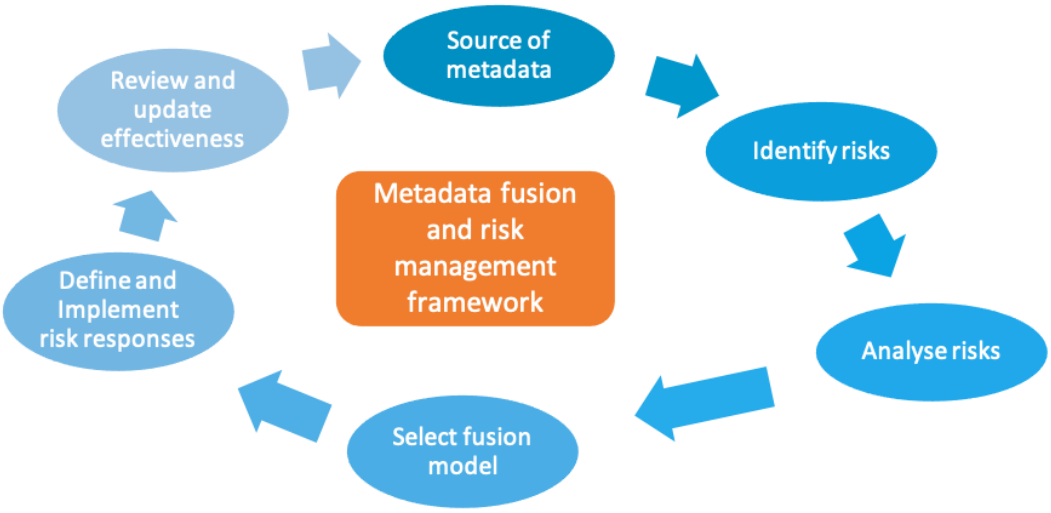

Source of metadata
Identify risks
Analyse risks
Select fusion model
Define and Implement risk responses
Review and update effectiveness
Metadata fusion and risk management framework

FIGURE 5 RISK ANALYSIS FRAMEWORK CYCLE

The first four stages have been addressed including:

- The metadata sources – Section 3
- Identify risks and analyse risks – Section 4, and
- Selection of fusion model – Section 2

These are the main focusses of this deliverable.

5.1.1 Architecture

As introduced in Section 3, five main types of input metadata source are included in the D4FLY risk analysis framework. Figure 6 below illustrates the risk analysis process of the proposed framework.

Input: the input metadata are listed on the left of flow chart which have been introduced in detail in Section 3

Output: the output from the risk analysis framework is a simple and clear risk level such as no risk, low risk, medium risk or high risk

GUI: The output of the risk level will be presented to the border guard on the border guard GUI system using an efficient and clear method

Feedback and update: the border guards can provide feedback based on the results/output and configure the system based on the border types, specific scenarios, etc.

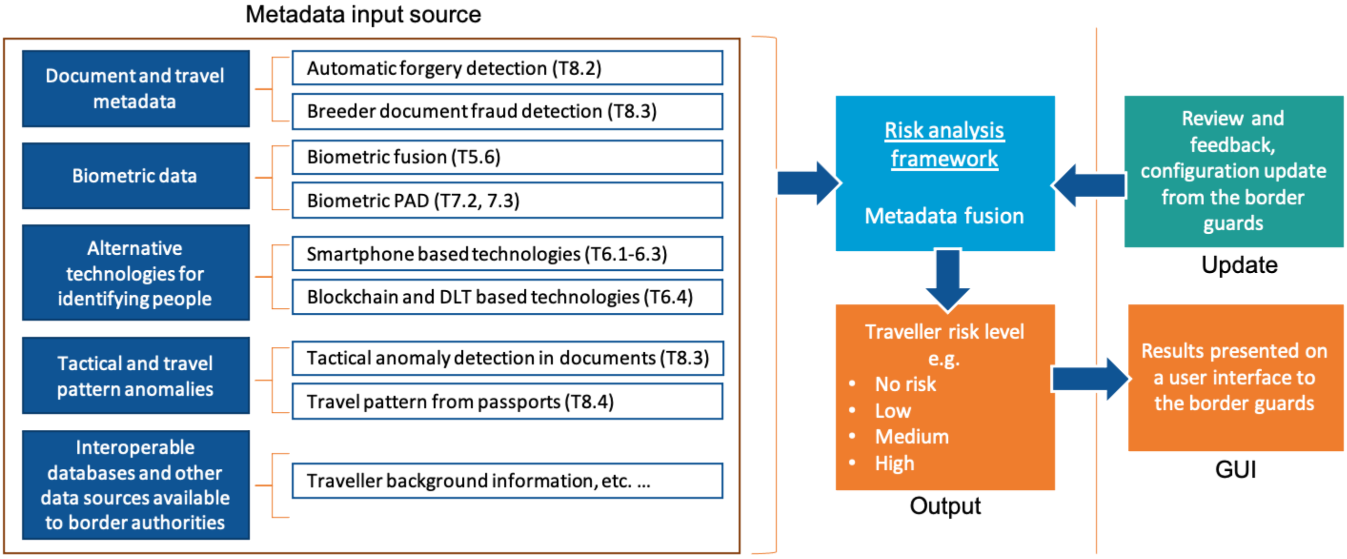

Metadata input source
Document and travel metadata
Automatic forgery detection (T8.2)
Breeder document fraud detection (T8.3)
Biometric data
Biometric fusion (T5.6)
Biometric PAD (T7.2, 7.3)
Alternative technologies for identifying people
Smartphone based technologies (T6.1-6.3)
Blockchain and DLT based technologies (T6.4)
Tactical and travel pattern anomalies
Tactical anomaly detection in documents (T8.3)
Travel pattern from passports (T8.4)
Interoperable databases and other data sources available to border authorities
Traveller background information, etc. ...
Risk analysis framework
Metadata fusion
Review and feedback, configuration update from the border guards
Update
Traveller risk level e.g.
No risk
Low
Medium
High
Output
Results presented on a user interface to the border guards
GUI

FIGURE 6 D4FLY RISK ANALYSIS PROCESS FLOW CHART

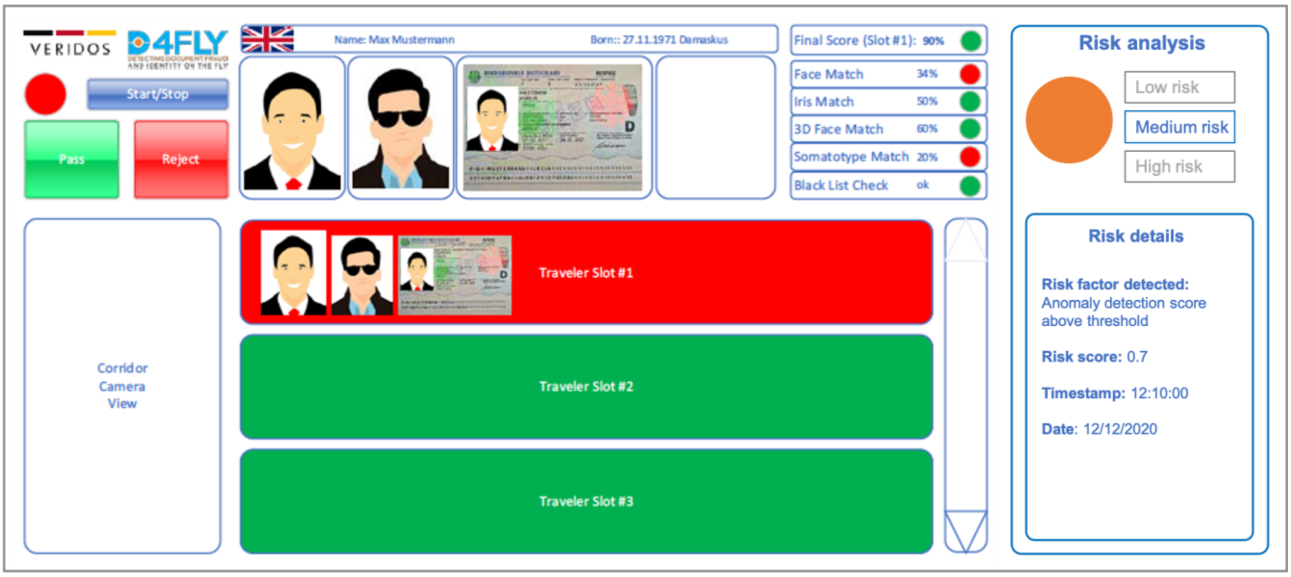
5.1.2 Traveller risk levels

The output of the metadata risk analysis is the level of risk of the traveller. The output result is based on the intelligent fusion of the metadata input and aims to provide the border guard with new information to support the decision at the border check situations. As the automated tools and technologies aim to develop the facilitation and speed of the border check situation, the presentation of the metadata fusion results aims to support this objective. The representation of the analysis in the border guard user interface is proposed to be categorised according to the model in the Common Integrated Risk Analysis Mode (CIRAM) methodology to low, medium and high levels of risks (see Table 4). When needed, border guards are also able to find background information of the fusion raw data and analysis results.

TABLE 4 THE CIRAM PROPOSED RISK LEVEL DESCRIPTIONS [12]

Level of risk	Description
Low	Acceptable risk. The impact can be dealt with, and the vulnerability is acceptable, but the threats must be monitored to discover changes that could increase the risk level.
Medium	Tolerable risk, but the impact is not easily dealt with given current capacity in place. A small increase of the magnitude of the threat could jeopardise the effectiveness of the response. The development of the threat must be monitored on a regular basis, with consideration to whether necessary measures have to be implemented.
High	Unacceptable risk. The impacts cannot be dealt with adequately with the given capacities and before risk reducing treatment has been implemented

The reliability of the input data source should be addressed in practice. However, for research purposes, all input metadata sources are assumed to be fully trusted and accurate. As considering the reliability factors from various data sources can be complicated and applying confidence scores, for example, will complicate the model training process. Therefore, evaluation and handling data reliability and accountability is not in scope for the current work, and can be addressed in the future research opportunities.



Name: Max Mustermann
Born:: 27.11.1971 Damaskus

Final Score (Slot #1):	90%	●
Face Match	34%	●
Iris Match	50%	●
3D Face Match	60%	●
Somatotype Match	20%	●
Black List Check	ok	●

Risk factor detected: Anomaly detection score above threshold

Risk score: 0.7

Timestamp: 12:10:00

Date: 12/12/2020

FIGURE 7 FIRST PROPOSED UI DESIGN BASED ON CURRENT BORDER STATION UI DESIGN [43]

As shown in the system architecture design (Figure 6), a Graphical User interface (GUI) will be designed and developed to present the risk level to the border guard in this task. Design of the GUI takes into account the feedback from the end users. The presentation should be simple to follow and universal, for instance, using a simple traffic light indicator to flag the risk level, with relevant risk factors that can be browsed in detail.

Deliverable 4.1 – UI specification 1 [43] introduces the current design of the UI for the border guards' screen. A traffic light style indicator design is adopted to present the risk analysis results to the border guards based on the three-tiered risk levels (low, medium and high). Figure 7 shows the initial GUI design that is based on and integrated into the current border station UI. Once any risk factor is detected to indicate an amber or red light, the detailed risk analysis results will be presented to the border guard so that the border guard can handle the situation accordingly based on the information. The detailed design and development of the GUI and how it can be integrated into the D4FLY system will be presented in the next deliverable D6.10.

5.2 Proposed metadata fusion scheme/algorithm

As introduced in Section 2.3, there are two main types of fusion schemes: rule-based fusion and learning-based fusion using e.g. neural networks. In this section, the proposed metadata fusion scheme to be used in D4FLY is described.

An event that causes a high-impact risk, is defined to be a significant action, deviation or result in the border control situation from the sources defined in Section 3 that may have an influence on border security. Such events that flag an alarm may need further attention or actions from the border guard. In general, there are three groups of events in the context of border security:

- Accidental – an event caused by unfamiliarity with the system which causes errors in the checking process, e.g. lack of dexterity, wandering attention, or leaving objects behind, etc.
- Deliberate – an attempt to avoid being recognised as a person on a watchlist or to evade the control completely,
- Alert – where a passenger can pass through the ABC, but their identity or profile triggers an alert

In the metadata fusion module, events that can be detected or monitored within the D4FLY defined technologies or process modules are focused on: Document and travel checks, biometric recognition and fusion, alternative technologies for identifying people and tactical and travel pattern anomalies. For example, the events can include low fused score of the available biometric recognition, travel pattern analysis, or a person presenting a fake passport. An example of the imaginary high-risk assessment is presented below.

The metadata fusion module combines data from multiple sources, e.g. biometric fusion and travel pattern data, and then filters the analysed information and provides high-quality feedback concerning the travellers to the border guards. The metadata fusion module outputs a decision on whether a traveller is entitled to cross the border or if he or she must be further checked manually by the border guard. If applicable and desired, an automatic entry may be allowed for travellers who pass the automated analysis made by the metadata fusion module.

The scope and goal of the risk analysis is to filter and construct a compact representation of the risk level of each traveller for the border guards. The border guards will have more confidence regarding the plethora of information from a compact indication of each traveller's risk level. The risk analysis system provides situational picture information from the border guard point of view to support efficient and effective border control situation.

5.2.1 Rule-based fusion

The module is based on a rule-based system for processing the situation at hand. The rule-based system compares the identified/noticed events against defined rules to produce an output (action/alarm). An example of a rule is: if there are inconsistences in the traveller's travel patterns, then an indication of a risk can be raised or the risk level of the traveller in question is raised. As an output, the metadata fusion module provides an action flag (e.g. an alarm, notice, or remark) of available and supported actions. From the architecture implementation point of view, the action flags are plain strings.

The D4FLY information system integrates and uses the metadata fusion module by providing an event string and executes a desired behaviour that is returned as an action by the metadata fusion module. The individual rules can react to one or multiple anomalous events. Respectively, the rules can result in one or multiple actions.

The metadata fusion module combines the biometric fusion result and detected abnormalities from other data sources. The metadata fusion module does not have a fixed (hardcoded) set of inputs and outputs because they are defined in the ruleset (event strings and action strings). Therefore, the metadata fusion module implements a simple language for defining the rules for generating an alarm flag. The ruleset can be modified to best suit a specific border check setup, border control point setups or border guard preferences. For example, the air border may map a specific set of anomalous events to actions, while the sea border may use a different set of rules. The ruleset is defined in plain text, and it can be edited by a corresponding organization if needed. The following figure illustrates the high-level idea of flagging an alarm and the assessment of the risk level.

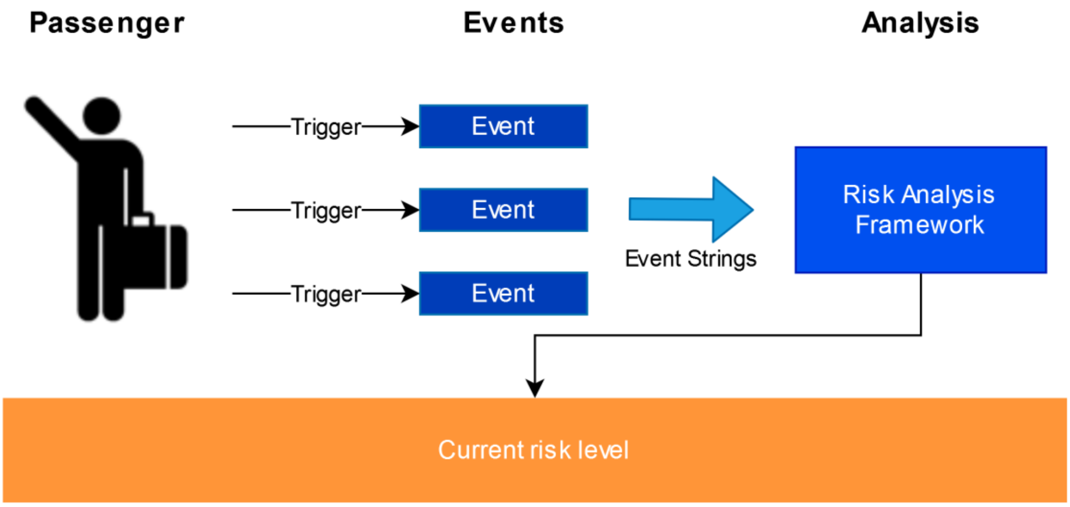


FIGURE 8 AN OVERVIEW OF THE METADATA FUSION MODULE WORKFLOW. IN THIS THE PASSENGER TRIGGERS EVENTS. THE OBSERVED EVENTS, WHICH ARE TRIGGERED INDEPENDENTLY, RAISE THE RISK SCORE OF A PASSENGER. THE RISK LEVELS HAVE SPECIFIC THRESHOLDS, WHICH DETERMINE THE RESULTING RISK LEVEL GIVEN A RISK SCORE. THE OBSERVED EVENTS ARE STRING VALUES, WHICH ARE PUSHED INTO THE METADATA FUSION MODULE BY THE CORRESPONDING EVENT MODULES.

For convenient integration and access, the metadata fusion module is planned to be implemented as a web service, which is called risk analysis service. The risk analysis service implements the required remote methods for:

- Recording event strings for passengers.
- Clearing event strings from passengers.
- Acquiring the current passenger risk level.
- Acquiring the currently recorded passenger event strings.
- Acquiring the immediate passenger risk level via a set of event strings.
- Acquiring a set of event strings, which are supported by the alarm flag service.

The remote procedure calls can be implemented using Simple Object Access Protocol (SOAP), which is widely supported by various programming platforms. SOAP defines the public methods of the alarm flag service using XML, which is a commonly used format for storing information. The following figure illustrates the high-level architecture of the alarm flag module:

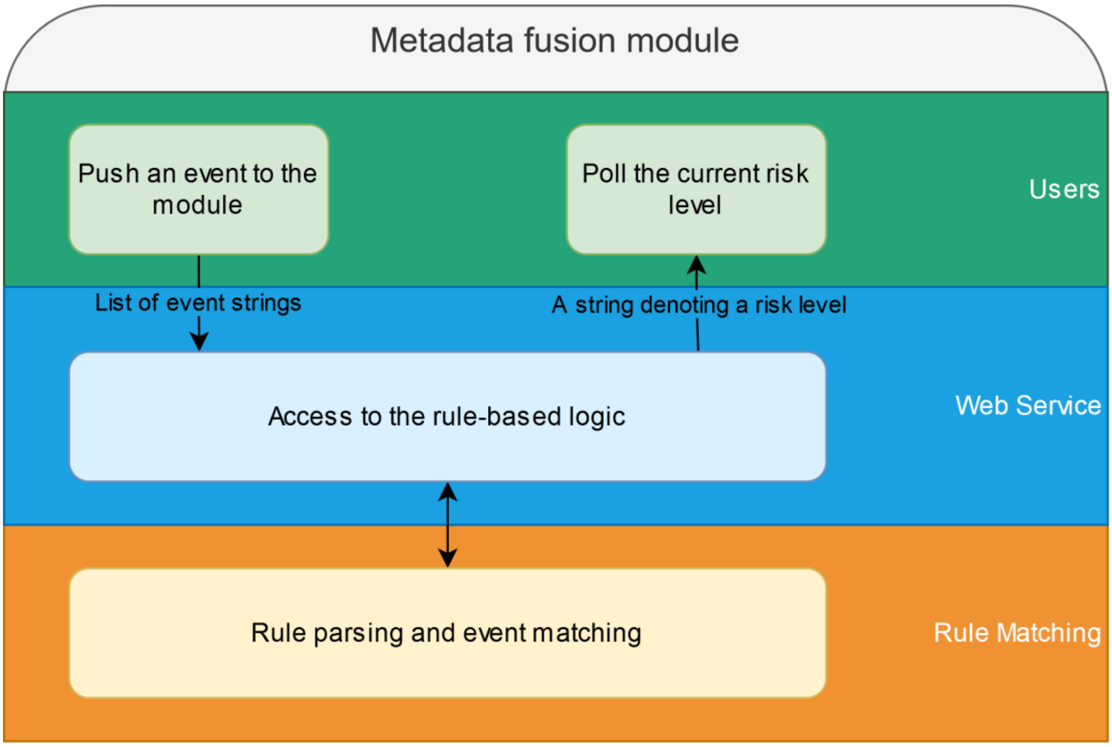

Metadata fusion module
Push an event to the module
Poll the current risk level
Users
List of event strings
A string denoting a risk level
Access to the rule-based logic
Web Service
Rule parsing and event matching
Rule Matching

FIGURE 9 AN OVERVIEW OF THE METADATA FUSION MODULE ARCHITECTURE BASED ON A RULE-BASED FUSION LOGIC. THE METADATA FUSION MODULE IS ACCESSED THROUGH A WEB SERVICE. THE WEB SERVICE (ALARM FLAG SERVICE) IMPLEMENTS METHODS FOR USING THE RULE-BASED MATCHING OF THE EVENTS AND ACTIONS. ADDITIONALLY, THE ALARM FLAG SERVICE IMPLEMENTS METHODS FOR POLLING THE RISK LEVEL OF A TRAVELLER

5.2.2 Neural network-based fusion

A neural network based fusion algorithm could operate as a non-rule-based fusion that is incorporated into the rule based fusion algorithm. A neural network classifier is trained for a specific task in specific control point, so that it takes multiple data sources as inputs and provide one output for the rule-based system. This kind of system can utilize different data sources with different properties as input and provide eligible amount of increased risk levels as output when enough training data is available.

In theory, providing real-time feedback, such as a result of inspection, to this kind of classifier would enable continuous re-training resulting in an increase in accuracy and adaptation to detect new risk profiles from metadata. Additionally, federated learning could enable transferring sensitive metadata risk profiles and aggregating information from different locations.

5.3 Evaluation criteria

The FastPass project, referred to in Section 2.2 above, applied a standard approach for evaluation of the risk analysis framework. In this task, a similar approach will be implemented to evaluate the performance of the developed system. The evaluation will focus on three criteria defined in more detail below:

Accuracy/Security:

The most important quality assessment would be to guarantee a minimum number of errors in classifying travellers into either risk levels or based on binary yes/no entry decisions. The overall method has to bear a low number of false alarms and should not leave out any potential threat. Standard scientific classification error rates should be employed for this evaluation.

As any binary classification-based system (and even in case of a risk-level interface with continuous scores, a binary assessment is feasible by the introduction of thresholds), it is scientifically useful to assess the entire system using an accuracy/precision-based approach.

The accuracy is the proportion of true results (both true positives and true negatives) in the population. In the context of a border control scenario assessment, this would refer to the number of fraud attempts detected as such (true positives) and genuine travellers being permitted entry (true negatives). A problem with this type of scenario evaluation is the collection of test data, as it cannot be assumed that within the short time of data collection it is possible to collect a sufficiently large number of fraud data (and thus might not be able to show statistical significance). Therefore, the scenario assessments are likely to be conducted with artificially recorded fraud attempts to show the feasibility of the method.

While accuracy is a good indicator to judge the degree of closeness of measurements of a quantity to that quantity's actual (true) value, the second indicator critical in the assessment of classification systems is precision, which is the proportion of the true positives against all positives and indicates the degree to which repeated measurements under unchanged conditions show the same results.

Given the fact that the rule-based engine can be interpreted as a classifier, it can be evaluated – given that ground truth data is available (considering the problem related to ground truth acquisition).

Performance:

Another desired criterion for the data fusion module is to return a classification result in real-time. While the combination of techniques is probably not the most critical part, the type of information used for fusion can make a huge difference. For example, it may be more time consuming to retrieve information from the travel patterns than doing biometric verification (e.g. face recognition). An adaptive data source selection process may be applied to provide a better performance with significant higher throughput without degrading the overall accuracy.

In summary, the system has to be real-time-capable, i.e. information has to be available instantly, and combinations with faster throughput at the same high level of accuracy could and should be preferred.

Usability:

Border guards should be able to intuitively use the system – rules should be modular and easy to exchange, and ideally different templates should be available. While this type of evaluation is likely to be the most time-consuming (and also subjective) form of assessment, as it cannot be automated, it is useful to integrate already established results from Work package 2 on User needs and requirements as well as Work package 9 on Field/Operational environment testing (including a task on Evaluation methodology).

While it is difficult to collect ground truth even in a real-world demonstration (it is unlikely that fraud attempts would be captured due to the low probability of such attempts), it is envisioned to have simulated behaviour and annotated scenarios for this type of evaluation using the introduced test scenarios above.

The evaluation task will also the focus of the second phase of this task and results will be reported in the next deliverable D6.10.

5.4 Evaluation using simulated data

There exists a lack of real-life data that can be obtained for training and testing purposes, especially real border crossing data that contains different types of risk factors. Therefore, to evaluate accuracy, performance and usability in a privacy preserving manner the evaluation will be initially based on the usage of synthetic data. The synthetic data represent different metadata source outputs, which are used as an input for meta-fusion tool. Synthetic data is generated such that each data source type has a specific distribution of random values. These values together form a feature vector which is utilized when risk profiles are determined. Synthetic data can be generated using random number generators with different distribution configurations or using generative adversarial networks [40]. Generating synthetic data is a separate process from a tool whose purpose is to provide input values for a meta-data analysis tool.

The synthetic data enables testing of the overall framework. It is possible to test features including accuracy and adaptability of framework by using different kinds of artificial sensor sets, and by leaving part of sensors outside of test feature vectors. Training and classification performance of the classifier will be also identified. Usability testing with the synthetic data can help to identify different delays and what kind of output is most useful for the end users.

6 CONCLUSIONS

Presently, there are almost no automated risk analysis tools at border crossing points to assist border authorities in managing and identifying the risks efficiently. Task 6.5 aims at developing a universal automated risk analysis system based on different fusion schemes that can adopt different metadata types, both being developed within D4FLY and external data sources such as interoperable databases, and can be adapted for deployment at different border types (air, sea and land borders). The risk analysis framework generates a universal risk profile for the travellers by combining a variety of metadata sources based on machine learning/data science techniques to efficiently and accurately detect risks/threats of the travellers. The outcome from the system could be presented in a standard warning system to show a simple flag to the border guards to indicate the risk level, and help them quickly and accurately identify any suspicious patterns. The system should increase throughput and accuracy, enhance security and reduce false negatives.

In D4FLY, this task focusses only on the risk analysis at the border crossing point. Risks can occur at any stage, such as ticket booking, etc. The potential of exploiting the system to be used at other check points can be further discussed in future work.

Throughout the document, the activities and progress carried out and results obtained during the first phase within the task (M10-M18) have been reported in detail. One of the main activities conducted was the interview with D4FLY end-users to help better understand the potential use of the technology to be developed, and to identity the requirements and needs from the end-users when developing the system. Literature review and background study on metadata fusion and risk analysis frameworks was undertaken to investigate different methods and techniques. How to best present the risk assessment output to the border guards has also been investigated. A suitable fusion model for process metadata has been proposed and the architecture of the risk analysis framework has also been defined.

In this document, several challenges and issues have been addressed for developing such a system:

- Data protection and privacy issues under GDPR: in D4FLY the metadata used for the fusion will only include the data that is available during the border check and that is gathered at border check, including passport control and the border check corridor solutions
- As there is no universal standard for all types of data, how to use the data that are not normalised or standardised in the fusion will be further investigated
- Lack of real-world data, especially with a variety of risks presented, could be a challenging issue for learning-based algorithm development and evaluation. One solution is to use simulated data

In the next phase, work will focus on:

- Development and evaluation on the fusion schemes. Testing using synthetic data on end-user defined scenarios has been planned
- Detailed design and implementation of the border guards' GUI system
- Implementation of the risk analysis framework
- Explore the possibility to integrate the framework into the D4FLY system
- Evaluation of the whole system using synthetic data and real data if possible

REFERENCES

[1] Dasarathy BV (1997) Sensor fusion potential exploitation-innovative architectures and illustrative applications. In: Proceedings of the IEEE

[2] EU FP7 FastPass project

[3] EU H2020 TRESSPASS project - robusT Risk basEd Screening and alert System for PASSengers and luggage)

[4] B. Micenkova et al., Stamp verification for automated document authentication, 6th International Workshop on Computational Forensics (IWCF), 2014

[5] Merriam-Webster online dictionary. https://www.merriam-webster.com/dictionary/metadata

[6] Rajamäki, J.; Sarlio-Siintola, S.; Simola, J. (2018) Ethics of Open Source Intelligence Applied by Maritime Law Enforcement Authorities. In Audun Josang (Ed.) Proceedings of the 17th European Conference on Cyber Warfare and Security ECCWS 2018, 28-29 June 2018, Oslo, Norway. Academic Conferences and Publishing International Limited, 424-431.

[7] Common Integrated Risk Analysis Model Summary booklet. version 2.0 Reference number: 17600/ 2013 en. https://frontex.europa.eu/assets/CIRAM/en_CIRAM_brochure_2013.pdf (accessed on 13.1.2021)

[8] Sebastian Klipper. Information Security Risk Management Risikomanagement mit ISO/IEC 27001, 27005 und 31010. Vieweg+Teubner Verlag |Springer Fachmedien Wiesbaden GmbH 2011

[9] SFS-ISO 31000:2018. Risk management. Guidelines. Finnish Standards Association SFS, 2018

[10] M. Ferrara, A. Franco and D. Maltoni: The Magic Passport, IJCB 2014

[11] D. J. Robertson R. S. S. Kramer, A. M. Burton (2017) “Fraudulent ID using face morphs: Experiments on human and automatic recognition”. PLoS ONE 12(3): e0173319

[12] Common Integrated Risk Analysis Model: A Comprehensive Update: Version 2.0, FRONTEX, 2012, Frontex. Ref. No. 16999

[13] Paul, R., Harmonisation by risk analysis? Frontex and the risk-based governance of European border control. Journal of European Integration, 39(6), first online 24 April

[14] Paul, Regine. (2018). Risk Analysis as a Governance Tool in European Border Control (pre-print).

[15] EU H2020 FLYSEC project

[16] National Information Standards Organization (NISO) (2001). Understanding Metadata. NISO Press. p. 1. ISBN 978-1-880124-62-8

[17] Zeng, Marcia (2004). "Metadata Types and Functions". NISO. Archived from the original on 7 October 2016

[18] EU H2020 PROTECT project

[19] Center for the Study of Democracy (2011) Better Management of EU Borders Through Cooperation. Study to Identify Best Practices on the Cooperation between Border Guards and Customs Administrations Working at the External Borders of the EU. Available at https://ec.europa.eu/home-affairs/sites/homeaffairs/files/e-library/docs/pdf/customs_bgs_final_en.pdf

[20] Stolen and Lost Travel Documents database https://www.interpol.int/en/How-we-work/Databases/Stolen-and-Lost-Travel-Documents-database

[21] Advance Passenger Information https://ec.europa.eu/home-affairs/what-is-new/work-in-progress/initiatives/border-law-enforcement-advance-passenger-information-api-revised-rules_en

[22] Passenger Name Record https://ec.europa.eu/home-affairs/what-we-do/policies/law-enforcement-cooperation/information-exchange/pnr_en

[23] REGULATION (EU) 2016/399 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 9 March 2016 on a Union Code on the rules governing the movement of persons across borders (Schengen Borders Code) (codification). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0399&from=EN

[24] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[25] Bamford (2020) *What the GDPR Means for Your Digital Content and Metadata.* Available at https://www.extensis.com/blog/what-the-gdpr-means-for-your-digital-content-and-metadata

[26] Risk Management for Cargo and Passengers – A Knowledge and Capacity Product, Inter-American Development Bank, Technical notes, No. IDB-TN-294

[27] A. A. Ross, K. Nandakumar, and A. K. Jain, Handbook of Multibiometrics. New York: Springer, 2006.

[28] L. M. Dinca and G. P. Hancke, "The Fall of One, the Rise of Many: A Survey on Multi-Biometric Fusion Methods," in IEEE Access, vol. 5, pp. 6247-6289, 2017.

[29] Alessandra Lumini, Loris Nanni, Overview of the combination of biometric matchers, Information Fusion, Volume 33, 2017, Pages 71-85

[30] K. Nandakumar, Y. Chen, S. C. Dass and A. Jain, Likelihood Ratio-Based Biometric Score Fusion, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 342-347, Feb. 2008.

[31] E. Marasco and C. Sansone, An Experimental Comparison of Different Methods for Combining Biometric Identification Systems. In: Image Analysis and Processing – ICIAP 2011. Lecture Notes in Computer Science, vol 6979. Springer, Berlin, Heidelberg

[32] Q. Zhang, Y. L. Yin, D. C. Zhan, and J. L. Peng, "A Novel Serial Multimodal Biometrics Framework Based on Semisupervised Learning Techniques," IEEE Transactions on Information Forensics and Security, vol. 9, pp. 1681---1694, Oct 2014.

[33] A. Ross, A. Rattani and M. Tistarelli, Exploiting the “Doddington zoo” effect in biometric fusion, IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Washington (BTAS), DC, 2009, pp. 1-7.

[34] Strategic Analysis (europa.eu), last accessed on 15 February 2021.

[35] Common Integrated Risk Analysis Model. version 2.0. Summary booklet. Frontex. 2013

[36] Kyriazanos D.M., TRESSPASS: Simulation and Field Tests for Risk-based BCP security and integrated EU border management. EAB-RPC 2020 Virtual Conference. 2020.

[37] Kyriazanos D.M. TRESSPASS: developments in risk based BCP security. EAB Research Projects Conference (EAB-RPC). 2019.

[38] Luyten K. and Voronova S. Interoperability between EU border and security information systems. BRIEFING. European Parliamentary Research Service. 2019. Interoperability between EU border and security information systems (europa.eu). accessed on 10.2.2021.

[39] Diana Dimitrova, Els Kindt. Legal requirements. In Recommendations for Future ABC installations. Ed. Sirra Toivonen. VTT. 2017. Recommendations for future ABC installations. Best practices (vttresearch.com) accessed 10.2.2021.

[40] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, Generative Adversarial Networks, arXiv preprint arXiv:1406.2661

[41] Regulation (EU) 2016/1624 of the European Parliament and of the Council of 14 September 2016 on the European Border and Coast Guard and amending Regulation (EU)

[42] D4FLY deliverable D3.3 – Privacy, data protection, social and ethical impact assessment 2, submitted in Feb 2021

[43] D4FLY deliverable D4.1 – UI specification 1, submitted in May 2020

LIST OF FIGURES

LIST OF TABLES