

D5.3 Recommendations and specifications for background data based risk analysis - Summary

The main goal of D5.3 is to offer recommendations and specifications for data-based risk analysis performed at the borders. The document responds to a need for a technology solution to support border guards in identifying, automatically and in real-time, high-risk travellers and trigger alerts before they are allowed to cross the border. Such a solution should be based on an analysis of the main information systems available to border control operators. Equally important for achieving this goal is an evaluation of the efficiency of different big data analytics techniques in terms of processing relevant information and presenting results so that they are immediately useful to the operator.

In order to meet this objective, D5.3 begins by providing an inventory of information systems and sources on which national border authorities in the EU may rely for the purpose of conducting risk assessment. The inventory comprises information systems that are already in use, alongside systems that are not yet operational but may become so in the future.

An analysis of those information systems reveals a high degree of fragmentation within the EU's current information landscape, with notable differences existing between the various information systems. Such heterogeneity may create obstacles to risk assessment, such as inconsistencies between data stored in different databases, similarities between data belonging to different persons and difficulties in retrieving all relevant data due to its scattering across multiple databases. In addition to these challenges, an important finding emerged from the analysis of information systems is the prevalence of data in alphanumeric text format, both structured and unstructured, complemented by a limited amount of data in image format, consisting primarily of biometric data (fingerprints and facial images). This peculiarity was taken as the starting point for evaluating the potential of different big data analytics techniques as methods to support risk assessment. Having described two basic methodologies for mining large and heterogeneous datasets (SEMMA and CRISP-DB), a variety of algorithms have been evaluated, illustrating the respective advantages and limitations. Two broad families of algorithms have been considered to this end: algorithms applicable to unstructured data and algorithms applicable to structured data.

Building on the result of the evaluation of big data analytics techniques, D5.3 proposes solutions for improving future border control risk assessment in the EU. Such solutions are based on the interoperability framework proposed by the EC in December 2017 and are presented in relation to two hypothetical scenarios. Scenario 1 focuses on the case of an EU citizen who is a victim of human trafficking and detected at a Schengen air border, whereas Scenario 2 involves a third-country national using a fraudulent identity to illegally cross a Schengen land border.

Within the context of Scenario 1, an Automatic Early Warning Risk Identifier System (AEWaRIS) is outlined, capable of analysing PNR and API data jointly with data from other information systems as well as to produce early warnings about travellers posing a risk. With regards to Scenario 2, a Multimodal Biometric De-duplication (MBD) system is proposed, designed to enable searches on all the data related to a certain traveller through a single query.



On top of the risk analysis solutions outlined above, D5.3 ends by offering a set of recommendations specifically directed at ensuring good performance and enhancing the future capabilities of the proposed systems.

Main author(s): Alessandro Bonzio (Z&P), Fabio Ruini (Z&P), Alessandro Zanasi (Z&P)

Contributors: Raul Sevilla (ATOS), Jose Maria Miranda (ATOS)

Month of delivery: March 2018

