D3.2 Privacy and Data Protection Impact Assessment

Document Due Date: 04/2020 (M8)
Document Submission Date: 30/04/2020

Work Package 3:
Ethical, social and legal (including data protection and privacy) impact assessment

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

This deliverable represents the findings from a privacy impact assessment conducted during the first 8 months of the D4FLY project. It includes preliminary recommendations to feed into the design and building of the D4FLY tools. This report highlights the privacy, including data protection, risks apparent at this stage of the project in terms of the extent to which the D4FLY tools have been designed and/or developed. It also assesses the potential impact of these risks so that steps may be taken to minimise any risks presented by the envisaged D4FLY tools, to ensure the project takes a privacy-by-design approach, and to help enhance GDPR compliance and demonstrate accountability.

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	D3.2
Document Title	Privacy and Data Protection Impact Assessment
Work Package reference	WP3
Delivery due date	30.04.2020
Actual submission date	30.04.2020
Dissemination Level	Public
Lead Partner	TRI (Trilateral Research)
Author(s)	Goldberg, Zachary & Muraszkiewicz, Julia (TRI)
Reviewer(s)	Toivonen, Sirra (VTT) Kyriazanos, Dimitris (NCSRD) Reuter, Armin (VD)

Document Version History

Version	Date created	Author	Comments
0.1	19.03.2020	Zachary Goldberg	Draft 1
0.2	06.04.2020	Julia Muraszkiewicz	Review and development: Draft 2
0.3	07.04.2020	Zachary Goldberg	Draft 3
0.4	17.04.2020	Sirra Toivonen	Draft 4
0.5	20.04.2020	Zachary Goldberg	Draft 5
0.6	24.04.2020	Dimitris Kyriazanos	Draft 6
1.0	30.04.2020	Zachary Goldberg	Final Version for Submission

List of Acronyms and Abbreviations

ACRONYM	EXPLANATION
ABC	Automated Border Control
AI	Artificial Intelligence
D4FLY	Detecting Document frauD and iDentity on the fly
Data Management Plan	DMP
EC	European Commission
EES	Entry and Exit System
EU	European Union
FP	Framework Programme
GDPR	General Data Protection Regulation
H	Horizon
HHI	Fraunhofer Heinrich-Hertz-Institute
IATA	International Air Transport Authority
ICAO	International Civil Aviation Organization
IND	Immigration and Naturalization Service Netherlands
M	Month
NTNU	Norwegian University of Science and Technology Norway
PIA+	Privacy, Ethical, Societal Impact Assessment
RNM	Royal Netherlands Marechaussee
SIS	Schengen Information System
TNO	Netherlands Organization for Applied Scientific Research
TRI	Trilateral Research
UoR	University of Reading
WAT	Wojskowa Akademia Techniczna Im. Jaroslawa Dabrowskiego
WP	Work Package

Table of Contents

1 INTRODUCTION

This document is the privacy impact assessment (PIA+) report (D3.2) for the D4FLY project. Since the D4FLY technological tools aim to verify travellers' identities through the use of biometrics and detect fraudulent breeder[1] and travel documents through the use of artificial intelligence, their development and deployment may significantly impact fundamental rights, EU values, and laws related to privacy and data protection. Of course, they may also have societal and ethical impacts, but the societal/ethical impact assessment is the subject of the subsequent deliverable, D3.3. The introduction of relevant concepts related to the concept of privacy such as autonomy was presented in D3.1 Privacy, Data Protection, Social & Ethical Issues Overview (M3). The purpose of this PIA+ is to map the data flows of the project's technological tools both in the project development phase and in the deployment phase, to assess the risks and opportunities that the tools could pose for privacy and data protection and to propose how to mitigate these risks or advance the opportunities with suggested recommendations and solutions. The analysis carried out in the PIA+ will, in turn, inform the technology development. Thus, the PIA+ provides an analysis of how the relevant technology should be developed to ensure that it promotes and protects privacy. In order to ensure transparency and foster public trust, as far as possible, the results of the assessment will be made publicly available via a deliverable published on the D4FLY website.

1.1 Background

A privacy impact analysis, including data protection, helps assess the potential risks and mitigations related to individuals' privacy and personal data emerging from the design, development and deployment of new technologies.

1.2 Aim of this document

The aim of this document is to summarize and report the results of the PIA+ of the D4FLY tools. It is intended that consortium partners read this report, with continued consultation with Trilateral, in order to take on privacy mitigations and opportunities into their design of D4FLY tools. As described in the Grant Agreement and Description of Work, this deliverable will assess the D4FLY tools' foreseeable impact on privacy and data protection, whereas the ethical and societal impact is the focus on the subsequent deliverable, D3.3.

1.3 Input / Output to this document

Input into this deliverable include: Institutional experience at Trilateral conducting PIA+ reports for EU FP7 and H2020 projects dating back to 2004; desk-based research involving a literature review including: academic papers and books, reports from standard bodies and international organizations, blogs, white papers and policy papers regarding the nature of privacy and issues relating to migration and border crossing, EU and national privacy and data protection laws and policies; interviews with consortium partners; and information acquired through an ethics workshop designed and held for consortium partners in January 2020 in Amsterdam. Output is this deliverable.

[1] Breeder documents are those used to apply for lawful travel documents such as birth, death, and wedding certificates.

2 WHAT IS A PIA?

A privacy impact assessment (PIA) is a systematic process for identifying and addressing privacy and data protection issues in the design, development and deployment of new technology while also considering the future consequences and impacts of proposed actions in relation to privacy and data protection. It can be described as an early warning system that can help expose risks regarding the project/technology/policies that are in development.[2] In the context of this deliverable we are focusing on privacy and data protection principles, and therefore the impacts focus principally on mitigating potentially negative consequences to these values and principles, but we will also consider opportunities to protect and promote these values. There are also prudential positive impacts that derive from taking these issues into account. Literature shows that organisations that take privacy seriously can benefit in monetary terms, design terms, and competitive advantage.[3]

A PIA+ is a process best undertaken at the initial stage of a project to have the maximum opportunity to affect the development of the technology. In that sense, a PIA+ aims to mitigate any risks related to privacy and data protection issues that may be present. (Recall that the ethical and social impact assessment is the subject of the subsequent deliverable, D3.3.) For technology development projects, a PIA+ acts as a foundational component for achieving meaningful "privacy-by-design" by providing information to support design decisions.[4]

> *Privacy by design calls for privacy to be taken into account throughout an entire engineering process.*[5]

However, a PIA+ is not a privacy or data protection audit. Instead a PIA+ examines relevant issues and includes engagement from stakeholders. As highlighted by Wright: "Engaging stakeholders, including the public, will help the assessor to discover risks and impacts that he or she might not otherwise have considered. A consultation is a way to gather fresh input on the perceptions of the severity of each risk and on possible measures to mitigate these risks".[6]

An important part of the PIA+ process is the preparation of a report, which documents the process itself. It helps a project consortium to identify the privacy impacts and what must be done to ensure that the project is not a liability. It also helps the project to assure stakeholders that the organisation takes their privacy issues seriously; it seeks the views of those who could be interested in or affected by the project. A PIA+ report is not the termination of an analysis,

[2] Wright, D, 2012.

[3] Cavoukian, A, 2011.

[4] Kroener and Wright, 2014.

[5] Information Commissioner's Office, 2008.

[6] Wright, 2012: 58.

but is rather a single, yet crucial, step in an ongoing process of reflection on privacy matters relevant to the project.

2.1 EU Policies Guiding a PIA

Current guidance from data protection regulators in the EU and outside recommend the use of privacy impact assessments.[7] Additionally, Data Protection Impact Assessments (DPIA) are required under the GDPR[8], and should be conducted on any high-risk data processing activity, before it commences. The current activity does not replace the DPIA that organisations may have to conduct in the future, but outputs from this process can support those other activities. The current process can help the projects to meet their other legal requirements in terms of data protection and privacy. Lastly, The Madrid Resolution adopted by the International Conference of Privacy and Data Protection Commissioners in November 2009 encourages: "The implementation of privacy impact assessments prior to implementing new information systems and/or technologies for the processing of personal data, as well as prior to carrying out any new method of processing personal data or substantial modifications in existing processing".[9]

Privacy is also protected by other laws and policies in Europe including:

- ISO/IEC 29100:2011
- ISO/IEC 27001:2005
- Universal Declaration on Human Rights 1948
- European Convention on Human Rights 1953
- Charter of Fundamental Human Rights of the European Union 2009
- OECD Guidelines on the Protection of Privacy and Trans border Flows of Personal Data 1980
- Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data 1980
- APEC Privacy Framework 2004

While this report is not about ensuring legal compliance with the aforementioned instruments, the number of legislative requirements that protect privacy, including information privacy, are indicative of the importance of assessing privacy risks associated with the D4FLY tools.

---

[7] ICO, 2020. *Conducting privacy impact assessments code of practice.*

[8] GDPR. Art. 35. The GDPR makes Data Protection Impact Assessments (DPIA) mandatory (Article 35) for processors (including technologies) that are likely to result in a high risk to the rights and freedoms of natural persons. The Regulation makes clear that the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. When the risks are identified, the GDPR expects that an organisation formulates measures to address these risks. This assessment should happen prior to the start of processing the personal data and should focus on topics like the systematic description of the processing activity and the necessity and proportionality of the operations.

[9] International Conference of Data Protection and Privacy Commissioners (2009).

2.2 What is Privacy? Literature Review

Privacy, including information privacy or data protection, is recognized by some scholars and policy makers to be a fundamental human right, with various international guidelines, accords and frameworks (listed above in 2.1) providing the basis for national laws, policies and international agreements globally. The United Nations recognised the right to privacy in the Universal Declaration on Human Rights 1948, under Article 12, which stipulates:

> *No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.*

In Europe, rights to privacy, as contained within the European Convention of Human Rights focus on "respect for private and family life, home and correspondence".[10]

Moving on from a descriptive focus on external, especially governmental, interference with individual privacy, participants in contemporary scholarly debates concerning the notion of privacy focus on its normative status, and consider it to be either instrumental to the development and exercise of capacities and intrinsic values such as autonomy and dignity (which principally relate to ethics, and, hence, will be discussed in subsequent deliverables), or as an intrinsic value itself.[11]

Considered as an instrumental value, privacy is seen to be essential for the development and exercise of other important values. For example, it has been argued that although privacy may have diverse interpretations and numerous contexts in which it is relevant, these interpretations and contexts are unified by possessing pertinence for the exercise of autonomy and for human dignity.[12] Bloustein argues that privacy defines one's essence as a human being, which includes individual dignity and integrity, personal autonomy and independence. Respect for these values is what both grounds and consolidates the concept of privacy, whether we define it as control over personal space, over information, over one's image, one's movements and associations, or otherwise. Consequently, violations of privacy are *ipso facto* demeaning to an individual's *person*ality and an offense to human dignity.[13] Placed in a legal context, the common conceptual thread linking diverse privacy cases prohibiting dissemination of personal information or non-consensual surveillance is the value of protection against abuses to individual freedom and human dignity.

Considered as an intrinsic value, privacy is seen to be valuable in itself and not because it provides for the exercise or development of other values or valuable capacities such as autonomy and dignity. To adopt this perspective, one must narrow the scope of privacy and explain its particular value. This task is a challenge considering the diverse contexts in which privacy appears and assumes normative value. On one account, privacy is argued to be valuable because it establishes intimacy amongst individuals.[14] For example, Fried defines

[10] European Convention of Human Rights, Art. 8.

[11] E.g., see Becker, 2019.

[12] Bloustein, 1964.

[13] Kupfer, 1987.

[14] Fried, 1970; Gerety 1977; Gerstein, 1978; Cohen, 2002.

privacy as control over information about oneself. He goes on to argue that privacy is necessarily related to an individual's ability to form intimate relationships involving respect, love, friendship and trust. As a consequence, it has intrinsic value in persons' lives.[15] Arguably, love, friendship and trust are only possible if individuals enjoy privacy, recognize its value for each other individual, and choose when to lower barriers of privacy to establish intimacy with others. Practically speaking, if we consider privacy to be intrinsically valuable, then violations of privacy are wrong because simply they violate privacy and not because they are affronts to human dignity or detrimental to the exercise of a person's autonomy.

Privacy can be interpreted as a *public value* meaning that it has value not just to the individual, but also to the democratic political system. For example, Daniel Solove argues that privacy promotes and encourages the moral autonomy of citizens, an essential requirement of governance in a democracy.[16] Others have argued that privacy is rapidly becoming a *collective value* in that "technology and market forces are making it hard for any one person to have privacy without all persons having a similar minimum level of privacy".[17]

Considering privacy to be both public and collective is especially germane for the D4FLY project. Regarding its collective characteristics, Stephen Kabera Karanja has argued that the sorting nature of border control is prone to discrimination.[18] It focuses too much attention on travellers because of their country of origin, skin colour, ethnicity, race, gender and, sometimes, religion. Discrimination exposes travellers, especially those from marginalized social groups, to undue attention that result in loss of anonymity and privacy. Undue attention can be dehumanizing especially when it elicits unjustified suspicion; it may injure an individual's dignity and integrity. Notions of privacy that are predicated on individual protection offer inadequate protection in border control situations where travellers are categorized and controlled *as groups*. Consequently, paying heed to the notion of group privacy may be necessary especially in the context of border control.

Regarding its public value, there is wide consensus that privacy is an important value for European citizens and residents proving worthy of protection. Significantly, there is also wide consensus among Europeans that they do not possess the amount of control over their privacy as they wish or that they deem necessary. Concerning personal data, and according to Directorate-General for Communication Special Eurobarometer 431 Report on Data Protection, only a minority (15%) feel they have complete control over the information they provide online, and 31% think they have no control over it at all. Two-thirds of respondents (67%) are concerned about not having complete control over the information they provide online.[19] Project partners ought to take seriously what the public perception of the project's

[15] Admittedly, it remains unclear why privacy should be considered under this view as intrinsically valuable rather than instrumentally valuable for intimate relationships, which are intrinsically valuable. However, the goal of this report is not to discuss the philosophical merits of the distinct views, but to present them as possible perspectives in the scholarly landscape.

[16] Solove, 2008.

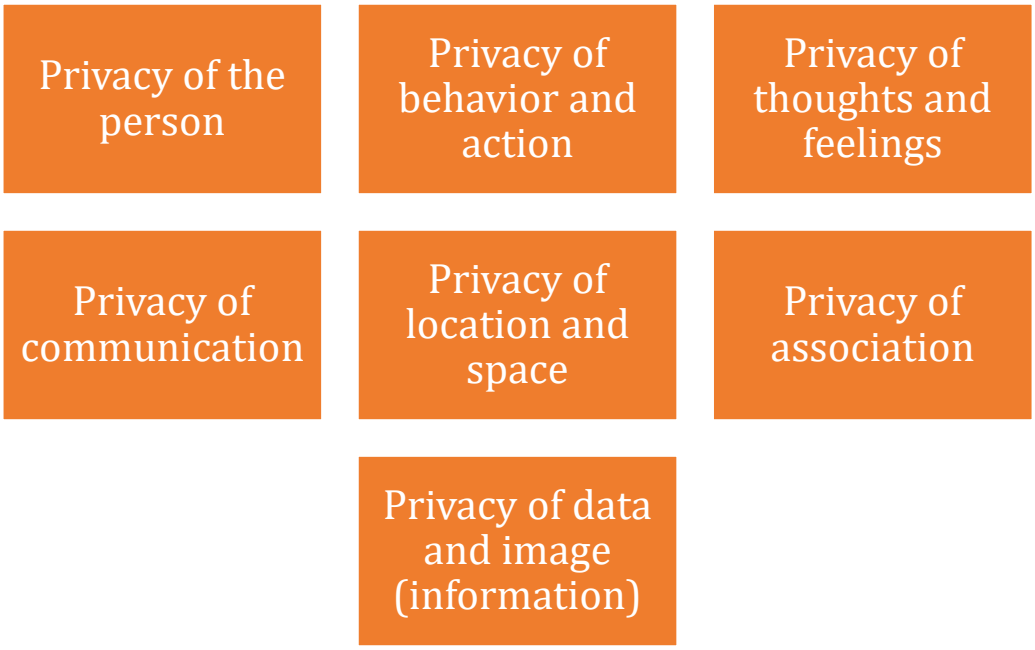
[17] Regan, 1995: 213

[18] Karanja, 2008.

[19] Directorate-General Special Eurobarometer 431 "Data protection", 2015.

tools might be and consider how to reassure travellers that their privacy is valued and their personal information will be protected.

To ensure that the D4FLY project identifies all contexts in which privacy might be relevant for its tools, it is important to note that privacy can take on different meanings in different contexts,[20] and that we can distinguish the following general categories where privacy is important: [21]



Especially regarding context, it is pivotal to note how important context is for the D4FLY project. To cross a border, travellers already voluntarily give consent to be identified by border guards. This is the known and accepted context within which the D4FLY tools are being developed. None of the D4FLY tools will be used outside of this context in which travellers freely give consent to be identified. This freely given informed consent established a sharp contrast with contexts in which biometric data is taken or surveillance carried out surreptitiously.

These insights regarding the complexity of the concept of privacy and the its wide-ranging contexts, as well as the above-mentioned legal policies and regulations, inform the current analysis of privacy and data protection within the D4FLY project.

[20] Nissenbaum, 2009.

[21] Finn, et al. 2013.

3 METHOD

3.1 Engaging with Partners

The consortium is continuously working on developing and implementing privacy-by-design. TRI is leading these efforts and has to date set up the following infrastructure and accomplished the following tasks in order to achieve this end:

1. Liaised with partners from the outset of the project to discuss data flows and privacy concerns. Sent partners data protection questionnaire to complete. Completed phone or in-person interviews with each technical partner.
2. Participated in every WP TelCo.
3. Asked partners to send monthly updates regarding planned activities with human subjects.
4. Advised on planned activities with human subjects.
5. Produced templates for the project's informed consent sheets.
6. Advised the communications partner on website related to privacy notices and privacy requirements, for example for cookies.
7. Worked closely with partners to write all of the EC ethics requirements deliverables (WP 11) due to date.
8. Liaised with ELAG members and sent all ethics-related deliverables to them.
9. Suggested a quarterly meeting with ELAG members and PC. The first meeting took place on 16 April 2020, and the next one is planned for the end of May 2020.
10. Taken trips (prior to the corona virus travel restrictions) to meet partners and observe their tools. Travelled to Raytrix in Kiel, Germany, Veridos in Munich, Germany, and UoR in Reading, UK. (Skype calls with all other partners.)
11. Taken trips to visit End Users to understand their need and requirements and to see first-hand how the tools could potentially be deployed. Destinations: Lithuanian/Belarus land border; Dutch Immigration and Naturalization Document Processing Centre in Zwolle, NL.
12. Reviewed DPIA deliverable from partner TNO.
13. Provide advice for data sharing agreement between partners TNO, IND, and RNM.

3.2 Report on a day-long interactive workshop

On 22 January 2020 the D4FLY project held a day-long "Ethical, social and legal (including data protection and privacy) impact assessment Workshop" to which all consortium partners were invited and was organised and led by TRI. Each partner was represented by at least one individual. In addition, in attendance were representatives from 6 stakeholder groups (The Association of Document Validation Professionals, European Passenger Federation, Finland National Police Board, inandoutcomes, SITA, Polish Security Printing Works) and two of three members from the Ethics and Legal Advisory Group (Gemma Galdon Clavell from Eticas Consulting and Katerina Hadjimatheou from University of Essex). In total there were 43 consortium members in attendance. The workshop was held at Schiphol Airport in Amsterdam

on the premises of one of the consortium's end users, Royal Netherlands Marechaussee (RNM). The agenda for the day is attached In Annex A.

Figure 1: Some of the D4FLY participants at the workshop in Amsterdam.

Method and Agenda:

The method for the workshop was developed to emphasize and fulfil privacy-by-design goals. It is pedagogically more effective when learners become acquainted with a new subject matter and its relevance through prompting and active autonomous discovery rather than passively listening to a lecture.[22] Learners not only retain the material at a higher rate, they are motivated after the instruction to continue educating themselves in the relevant field.

Furthermore, learning about ethics (including privacy and data protection for the purposes of the project) cannot simply focus on memorizing certain ethical theories, and ethics and privacy principles. This sort of educational model could elicit a kind of moral fetishism—adhering to moral principles simply for the sake of adherence at the expense of consideration of nuance, of exceptions to the adopted principle, or the importance of avoiding harm and protecting values such as autonomy and dignity. As a worst case, the rote memorization of ethics and privacy theories and principles could even introduce in the individual a false confidence that one is performing the morally right action. Hannah Arendt writes of her shock upon hearing Adolf Eichmann profess during his trial that he was a Kantian[23]. Indeed, it is still difficult today to understand how someone who orchestrated the murder of millions of people could consider himself to be the follower of one of the West's most influential moral philosophers. Arendt concludes that Eichmann distorted Kant's categorical principle from "Act so that the principle of your actions is universalizable" to "Act as if the principle of your actions were the same as that of the legislator or of the law of the land".[24] Arendt's famous assertion that evil is "banal" is based on the unreflective or "unthinking" ways in which Eichmann

[22] Hannafin & Land, 1997; Wright, 2011.

[23] Arendt, 1964: 66.

[24] Ibid.

adopted moral principles to his situation and thereby justified his actions to himself. The lesson for us today, even in situations less morally urgent than those of Eichmann's victims, is that learning about morality and ethics ought not proceed by way of an unreflective memorization of principles, but through a process that encourages active and recurring moral reflection.

To that end, the purpose of the workshop was to collaborate with partners on the analysis and discussion of privacy and ethics principles, concerns, and opportunities that arise in the D4FLY project. In order to sincerely facilitate a privacy-by-design approach to the development of the project tools, D4FLY held a truly collaborative workshop.

Rather than mapping out the data flows prior to the workshop and simply presenting them to the partners, the partners worked on mapping the data flows together during the workshop. This activity demonstrated to the partners that mapping data flows can be complex, and that the D4FLY tools may engender privacy and data protection concerns that they had not previously considered.

TRI then gave a presentation on the nature of privacy, GDPR regulations, and ethical values including different ethical theories that are relevant to the project. This presentation included scholarship relevant to privacy and ethics as well as vignettes to clearly communicate to partners relevant scenarios that could arise in both the project phase and the deployment phase of D4FLY.

Subsequently, partners were organized into small groups comprised of at least one individual from a technical partner, one from an end user partner, and one from a stakeholder organization. They were asked to identify and discuss privacy, data protection, and ethics risks and opportunities corresponding to each of the tools being developed in the project.

Finally, everyone reassembled in the main room to give feedback on their group discussions and to discuss potential mitigations for the privacy and ethics concerns that had been identified.

In these ways, partners were integrated into the PIA in a collaborative approach. This integration and collaboration has continued past the workshop as TRI continues to participate in every WP TelCo and to discuss privacy related issues with partners.

4 CONDUCTING A PIA IN D4FLY

The D4FLY PIA analysis will achieve the following objectives (as described in the Grant Agreement):

1. Map the personal data flows between the D4FLY technologies, users and the services with which they interface.

2. Identify privacy and data protection risks associated with these data flows.

3. Report on past and continuing engagement with all partners to suggest possible technical or operational solutions, mitigation measures, and formulate recommendations to minimize unintended impacts.

4.1 Mapping Data Flows of D4FLY Tools in Project and Deployment Phases

Motivation for the D4FLY Tool Development

With the introduction of Automated Border Control (ABC) systems using digitally signed data from electronic passports and biometric recognition, classic document security may appear to lose its importance. Apart from standardization activities on machine readable security features by the International Civil Aviation Organization (ICAO) New Technologies Working Group [25] fewer innovations or improvements in the field of physical security feature verification have been made in recent years. Most systems still rely on simple image comparisons in visible, UV and infrared light, matching the questioned document against pictures of known good specimens. Fraud at ABC systems is related to certificate verification capabilities or attackers are using morphed face images in genuine electronic passports to trick the automated systems in accepting two persons for one reference image. Significantly, a recent evaluation of the current security system in place in Schiphol Airport (one of the Field Test sites in D4FLY) found clear cybersecurity vulnerabilities that ought to be addressed immediately.[26]

With increasing numbers of travellers (International Air Transport Association (IATA): near doubling between 2017 and 2030)[27] and the inability to increase space at border crossing points, speed of document and identity verification becomes a crucial issue. Thorough document and identity checks by border guards have to compete with travellers' need for reasonable processing times and limited space for queues. Public opinion on border control performance is constantly deteriorating [28] and this is increasing pressure on the border security authorities from press and politicians.

According to the IATA Global Passenger Survey 70.4% of travellers would willingly share additional data to speed up the border control process. FRONTEX has recently focused on

[25] Technical advisory group on machine readable travel documents (TAG/MRTD) twentieth meeting Montréal, 7 to 9 September 2011.

[26] DutchNews.nl, 2020.

[27] IATA, 2017.

[28] BBC, 2018.

biometrics on-the-move technologies[29] aiming at improvements on speed and convenience for citizens and third country nationals.[30] In addition to biometrics, and document fraud, attention should also be given to evaluating emerging technologies such as blockchain and distributed ledger technology to aid border control by providing a verifiable, immutable chain of trust.

Responding to a growing terrorist threat, the EC has adopted EU Regulation 2017/458 that reinforces checks at external borders with regard to persons enjoying the right of free movement under Union law. Reinforced checks create additional pressure on the flow of traffic, but this regulation gives member states the option to carry out these checks only on a targeted basis at specified border crossing points and only if these checks would not have a disproportionate impact on the flow of traffic. Obviously, this exception could create vulnerabilities in the EU borders if not managed carefully.

The D4FLY consortium has been established to respond to this challenge and is designing biometric tools to verify identity, AI algorithms for document authentication, and is exploring how blockchain technology might be implemented as a means to transfer digital information from travellers to border guards. It should be highlighted that with regard to the latter, this is in very early – as mentioned exploratory – phases rather than building for the purpose of using blockchain.

More efficient and accurate document and identity verification is a **privacy opportunity.** By aiming to decrease organized crime and human trafficking through better border security, victims', and potential victims', privacy and dignity can be better protected.

Mapping

When mapping data flows corresponding to D4FLY tools, two principal phases must be considered:

1. the project development phase, which encompasses the development of the tools and includes internal testing; and
2. the deployment phase, which imagines future scenarios if the tools were to be implemented by the EU as part of its border security processes.

This report will assess privacy and data protection concerns and opportunities in both of these phases, and consequently make relevant recommendations.

Considering the D4FLY tools in an imagined deployment phase raises two kinds of complexities. On the one hand, the scenarios are imaginary and, as a result, we must try to imagine all possible privacy risks and opportunities that could arise with the tool while simultaneously remaining realistic. On the other hand, the number of the privacy risks

[29] D4FLY adopts the term "biometrics on-the-move" or "on the fly" as defined by FRONTEX in April 2017 as "Acquisition of data (in particular biometrics) at a distance for the purpose of identity verification as a person walks by data capture equipment. The objective is to facilitate the processing of persons who cross a border crossing point, thus contributing to a smoother (and faster) flow" to supersede the phrase "biometrics on-the-fly".

[30] Frontex, 2018.

increases as we imagine widespread use of the tools, and a widespread number of actors that come into contact with the tools.

As stated above, it is crucial to keep in mind that in the context of border crossing, **travellers already consent to have their identity verified**. Currently, this process occurs by presenting a passport or other travel and/or identification document. As discussed in section 2, privacy is contextual, and the border crossing context is paramount to understanding privacy concerns in the D4FLY deployment phase. D4FLY tools are not going to collect data regarding an individual's identity outside of a context in which the individuals have already consented to having their identity verified. **The tools only seek to change the way identity is verified in the context of border crossings, but not expand the situations in which identity is verified.**

Finally, readers should also note that the project, and consequently this report, are divided into two main foci: identity verification and document authentication, both of which are the two pillars of D4FLY. Wherever possible in what follows, tools that share a similar data flow pattern have been grouped together for ease of communication.

Data Flow Maps: Project and Deployment Phases

Biometrics

We begin with data related to biometrics; these are defined as: biological measurements or physical characteristics that can be used to identify individuals.

Constraints on taking biometrics include: the trait or feature selected for identification purposes must be universal. Every person must have at least one, it must remain constant over time, and it must not commonly be "lost to accident or disease"[31]. The attribute should have properties unique to each individual. The feature or trait should be able to be easily measured without violating the privacy of the individual.[32]

Background – Schengen Information System[33] and Current Practices in EU

Launched in 1995 as a measure to support the abolishing of internal border controls within the Schengen area, the Schengen Information System (SIS) contributes to cooperation among law enforcement agencies in the Member States and to the reinforcement of external border controls.

SIS enables competent legal authorities, such as police and border guards, to input and receive alerts on certain categories of wanted criminals or missing persons.

Initially, consultations regarding person-related alerts could only be made on the basis of alphanumeric data (e.g., name, surname, date of birth). However, it became clear that this kind of search procedure had limitations as criminals often change identities or use different aliases.

In order to tackle such limitations, SIS currently offers the possibility to store, as part of person-related alerts, dactyloscopic data, including fingerprints/palm-prints and finger-marks/palm marks (the latter only in the case of alerts related to terrorist offences and other

[31] Sareen, 2014.

[32] Bergstedt et al., 2018.

[33] The following information is drawn from updates and reports from the EU Science Hub.

serious crimes) and facial images. Soon DNA profiles will also form part of the data in cases of missing persons.[34]

The forthcoming Entry and Exist System (EES) is meant to complement SIS by automating the monitoring of the border-crossing of third-country nationals. Collected data will include the name and date of birth of the traveller, as well as dates of entry and exit into/from the Schengen Area. In addition to these alphanumeric data points, it is planned to store biometric data like pictures and fingerprints.

Furthermore, it is currently common practice for EU Member States to issue electronic passports (ePassports). In order to cope with increasing security needs, a number of EU Member States have deployed Automated Border Control (ABC) systems that automate border checks for EU citizens in possession of an ePassport.

In practice, an ABC system[35] works by using the biometric data stored in the ePassport to verify a traveller's identity. The system verifies that the ePassport corresponds to its holder by comparing the individual's biometric characteristics with biometric data stored in the ePassport, checks the alphanumeric information against border control records, and finally determines eligibility for border crossings, without border guards intervening. Nevertheless, border guards can still intervene whenever something is wrong or does not go according to plans.

Finally, since D4FLY end users must be compliant with Regulation (EU) 2016/399-Schengen Borders Code, consortium partners are considering this regulation when designing tools, setting up field tests, and considering use cases.

Further applicable regulatory standards for using biometrics and automated border control in D4FLY include[36]:

- CEN/TS 17262:2018 'Personal identification - Robustness against biometric presentation attacks - Application to European Automated Border Control'
- Cen/Tc 224 'Personal Identification and Related Personal Devices With Secure Element, Systems, Operations And Privacy In A Multi-Sectorial Environment'
- CEN/TS 16634:2014 **'**Personal identification - Recommendations for using biometrics in European Automated Border Control'
- ISO/IEC JTC 1/SC 17 'Cards and security devices for personal identification'
- ISO/IEC JTC 1/SC 37 'Biometrics
- ISO/IEC WD TR 22604 'Biometric recognition of subjects in motion in access related systems'

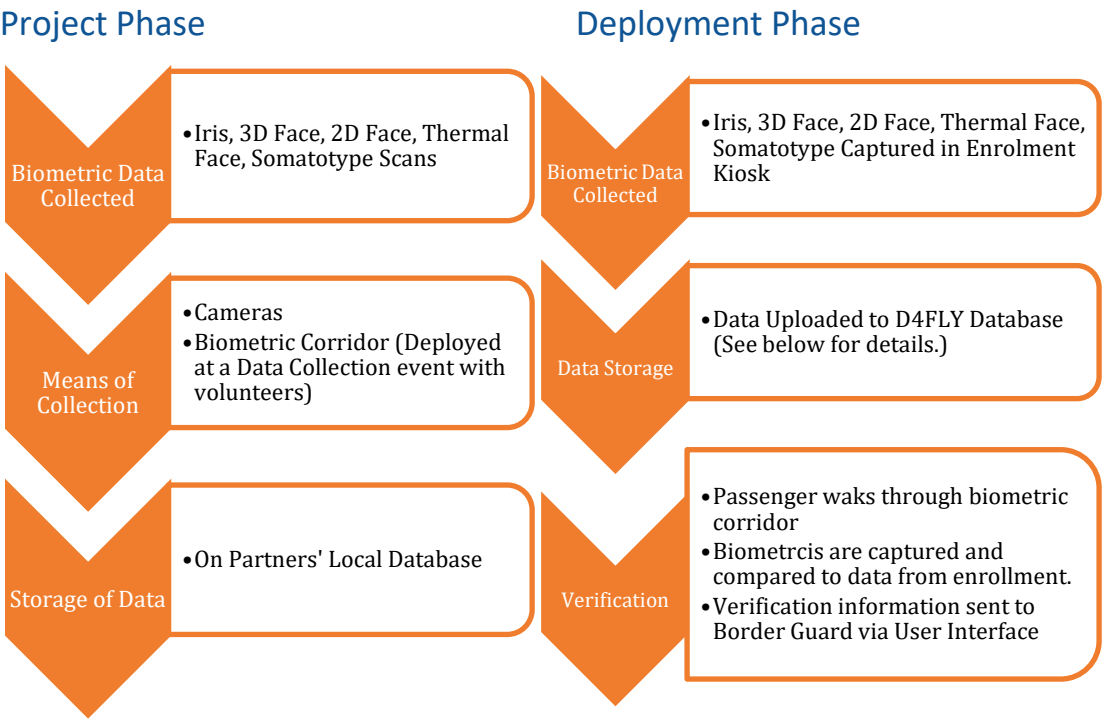
[34] Regulation (EU) 2018/1862

[35] For the official definition see: https://ec.europa.eu/home-affairs/what-we-do/networks/european_migration_network/glossary_search/automated-border-control-abc_en

[36] Trilateral is continuing to advise partners on the understanding and implementation of these regulations. When in doubt, they can contact TRI for assistance in understanding and implementing.

- ISO/IEC 19794 and ISO/IEC 39794 series 'Information technology — Biometric data interchange formats'

D4FLY is developing the following biometric technologies for identifying people on-the-move:



Privacy Concerns and Mitigations:

Despite foreseen adherence to the above-mentioned standards (and the relevant GDPR regulations presented in submitted deliverables D1.1 and D11.1-13) the following privacy and data protection concerns must be addressed.

Concern: Where biometrics are used in security contexts, there is a growing trend for the retention of full images (US-VISIT and EURODAC) or large samples (such as in the case of the UK National DNA Database). Arguably, the collection and storing of larger images is done to increase the effectiveness and accuracy of these systems. However, it could also be argued that a principal advantage of using smaller images is the reduced cost in terms of data storage owing to their being smaller in size. Nevertheless, with continued advances in storage capacity, it can be reasonably assumed that future biometric data collection will include the retention of the full image. Such trends significantly increase the potential chances of biometrics being jeopardized, an outcome not that unrealistic given that several governments have lost critical and large amounts of data on citizens.[37]

Mitigation: Partners must minimize the biometric data collected and implement a secure storage system with a strict time limit for data retention. For partners seeking further instruction, see Deliverables 11.1-Data Management Plan and D11.13 'A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants'. It is important to note, however, that the D4FLY Grant Agreement Article 18.1 requires partners to retain records and

[37] McCarthy, 2012.

documentation of their work for 5 years following receipts of final payments from the European Commission (EC). As such, personal data may be retained for this period of time where it is necessary to do so.

Concern: Biometric data is collected unknowingly from individuals.

Mitigation: During the project phase, the individuals providing their images are either employees of partners or volunteers from the consortium partner, University of Reading. During the deployment phase, as long as relevant GDPR legislation is followed concerning informed consent and security of data storage the risk to participants' privacy is low. Partners have been informed about the relevant legislation and principles (see Sections 3.2 and 3.3).

Concern: There is a risk that unexpected personal data, especially health data, could be communicated through the capture different biometrics. In an experiment carried out by Arora et al., researchers matched iris images captured before and after alcohol consumption. [38] The consumption of alcohol causes the pupil to dilate, which causes deformation in the iris patterns and, in turn, significantly affects the matching performance of iris scanners. Potentially, information about a traveller's on-board alcohol consumption could be communicated unintentionally to co-travellers by the automated system failing to make a match or perhaps by a border guard saying aloud why the verification failed.

Furthermore, during the Covid-19 pandemic (ongoing as of the writing of this report) thermal imaging has been utilized to detect whether individuals have a fever.[39] First, scientists are still undecided whether thermal imaging for fever detection is reliable. Consequently, "the World Health Organisation (WHO) has warned temperature screening for Covid-19 could yield false positives and is not effective for those who are asymptomatic."[40] Secondly, outside of the context of a pandemic, whether an individual has a fever is considered personal information. He/she may have a fever for non-contagious medical reasons, and may want these reasons kept private.

Mitigation: The mitigation for this risk is to train border guards to keep information concerning why the verification process failed confidential until the individual is in a private area separated from others.

Concern: There is a risk that the enrolment process will be biased against members of ethnic minorities thereby excluding them from participation. For example, in 2019 Joshua Bada, 28, from West London, tried to renew his passport and had his picture rejected as the online facial recognition technology confused his lips for an open mouth.[41] All users applying for a passport must submit an image of themselves with a plain expression and closed mouth. He had submitted a high-quality picture of himself

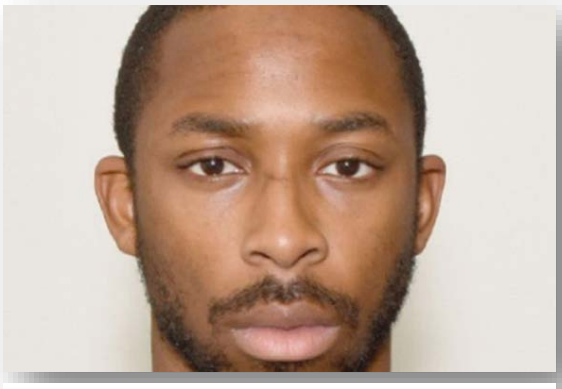


FIGURE 2: JOSHUA BADA, 2019

[38] Arora et al. 2012.

[39] BBC, 2020.

[40] Ibid.

[41] Cook, 2019. (Image also from this reference.)

on the system, which aims to inform users whether their pictures will meet the necessary criteria for a successful application.

Mitigation: The mitigation for this risk is to train the system used for enrolment on as diverse a dataset as possible. This case and similar problems were discussed in detail at the D4FLY Ethics Workshop and the technical partners believe they can address such potential problems. It was agreed that the primary step in mitigating bias is rigorous training of the tool with as diverse a dataset as possible and the immediate correction of any issues arising with the image capture of individuals. Research external to the D4FLY project has shown promise in this direction. For example, IBM report, "Unlike human beings, machines do not have inherent biases that inhibit D&I (diversity and inclusion). Rather, they are subject to the choices of data and algorithmic features chosen by the people building them. When appropriately developed and deployed, AI can remove the attributes that lead to biases and can learn how to detect potential biases, particularly those unconscious biases that are unintentional and hard to uncover in decision-making processes." [42] These claims must be met with scepticism. However, it may be that the use of AI could avoid many of the biases that humans hold, especially implicit biases. Technical partners should investigate this further.

Concern: There is a risk that an individual could undergo a significant change with his/her body between the time of enrolment and verification thereby excluding him/her from participation in the somatotype verification process. Such changes include a significant gain or loss of weight or a sex change. Imagine a pre-op transgender individual who does not want attention drawn to differences between his/her biological sex indicated on his/her travel document or his/her somatotype scan, and his/her outward appearance.[43] The individual may want such information to be kept confidential even from strangers who may be crossing the border at the same time.

Mitigation: The mitigation is the same as stated earlier: training. Border guards must be trained to keep confidential the reasons for a failed verification process. If the traveller cannot proceed through the automated border crossing, he/she will be redirected to an inspection by a human border guard. The reasons for this re-direction will be confidential and communicated only to the traveller and only in a private area. If the first line of the questioning occurs in the open area, the queue of travellers must remain at a certain distance to ensure privacy. Furthermore, it is not foreseen that somatotype biometrics alone could sufficiently verify an individual's identity. It is foreseen that it would be bundled with other biometrics to help verify identity.

Concern: The datasets could be accessed by unauthorized persons or stolen.

Mitigation: Partners are using databases with robust security measures.[44] For example:

WAT:

- The data set is anonymised and cannot be assigned to a specific subject by means of additional information or any other means (data is stored in

[42] Zhang et al, 2019: 6; Guenole & Feinzig, 2018.

[43] See Wilcox 2017 for a discussion about gender and biometrics at border crossings. Many of the points she identifies are societal and ethical issues rather than strictly privacy issues and, hence, will be addressed in the subsequent deliverable, D3.3.

[44] See D1.1 (DMP) for detailed information about the anonymization and pseudonymization techniques.

numbered folders, without names of persons). This treatment is permanent and irreversible.
- The data is processed by 3 people reported to access the database and authorized to process personal data.
- Access to the database is possible only from computers to which access (password protection) has only authorized persons.
- IP address filtering is used for all the computers having access to the data storage
- Access to data storage is password protected. Each person processing data uses personal login and password.
- The data storage is not accessible from the network outside the University (no public IP and public port).
- Access to the disk takes place in the internal network of the Military University of Technology. The area of the university is closed, with electronic based access control.
- The physical access to the disk is secured with a separate lockable room with a biometric access control system.

Veridos:

- Images of an iris from Veridos personnel will be used.
- These images will be anonymized, which means, only the iris part is stored, without the data subject's name.
- These images are stored locally on Veridos computers, which are access protected with username and password and can be accessed only by authorized persons. The data is not accessible from the network outside Veridos (firewall).
- Only a restricted and defined circle of people have password access to these images, additionally, the computers are located in an separate lockable room to which only around 10 people have access (entry Is key card controlled).
- Kiosk and BCP computers will have internet access in order to connect to a secure cloud storage database (Amazon workspace with restricted and dedicated authorized access).
- Any personal data will only be transferred and stored in encrypted form.
- For the data transfer only secure (encrypted) data transfer channels will be used.

UoR:

- Datasets (iris scans and multimodal biometric datasets) will be stored on GDPR compliant Microsoft SharePoint which is secure, and auto-backed up.
- Only authorised user accounts (currently only people who work on D4FLY project) can access the data and 2-factor authentication is required to login in to the SharePoint.
- When it is needed to share the dataset with project partners, access to the dataset will be granted upon submitting a request form and 2-factor authentication is also required to login and access the data.
- If biometric datascans will eventually be released to third parties for biometric competition and research, this information will be included in the information given to volunteers prior to consent. Volunteers retain the right to retract their initial consent at any time. All additional GDPR constraints will be followed as described in the DMP.

NTNU:

- The data sets are anonymised. Each identity is characterised by a numbered label and no other personal information can be derived. Note that the label is arbitrary and in no way related to any real information of the subject.
- The data is processed and can be accessed by only 2 individuals authorised to do so.
- Access to the database is possible only from the NTNU-provided laptops that belong to the 2 authorised individuals. Furthermore, the electronic access is password protected (both to the data set and to the laptops).
- The IP address is filtered in order to be identified as a NTNU address. For external access an authorised NTNU VPN connections should be established. Thus, the data sets are not accessible from networks outside NTNU.
- The data storage unit is secured in a lockable room, within the Computer Science Department, that can be unlocked only by the electronic access cards of the 2 authorised individuals. Those cards come with a password.
- Automated backups of the data sets are scheduled once a week on two additional storage means.

Algorithms-AI and Handcrafted Features

We next move onto elements of the D4FLY that relate to artificial intelligence (AI) algorithms and handcrafted algorithmic features. AI algorithms are computer software programs that are trained by machine learning to recognize patterns on their own without continuous human input. Recent scholarship has identified privacy concerns arising with the use of AI in the contexts of law enforcement[45], medicine[46], advertising[47], and border security[48] (to name only a few). One of the principal concerns arises due to the automation of decisions that affect individuals' rights and well-being. Handcrafted algorithmic features are manually engineered by a data scientist.

Regarding data protection, the provisions of the GDPR govern the data controller's duties and the rights of the data subject when personal information is processed. The GDPR therefore applies when artificial intelligence is under development with the help of personal data, and also when it is used to *analyse* or *reach decisions* about individuals.[49] The data protection principles and guidelines outlined in the Data Management Plan (D1.1-DMP) must be understood and adhered to. TRI has walked partners through the DMP and continues to advise in this capacity.

Regarding privacy, there have been several recent privacy violations by large companies using AI tools.[50] The Royal Free London NHS Foundation Trust, a division of the U.K.'s National Health Service based in London, provided Alphabet's DeepMind with data on 1.6 million

[45] Rowe and Muir 2019.

[46] Price et al. 2019.

[47] Estrada-Jiménez et al. 2019.

[48] Beduschi 2020.

[49] Norwegian Data Protection Authority, 2018: 15.

[50] Wiggers 2019.

patients without their consent.[51] Google abandoned plans to publish scans of chest X-rays over concerns that they contained personally identifiable information.[52] In 2019, Microsoft quietly removed a data set (MS Celeb) with more than 10 million images of people after it was revealed that some were not aware they had been included.[53]

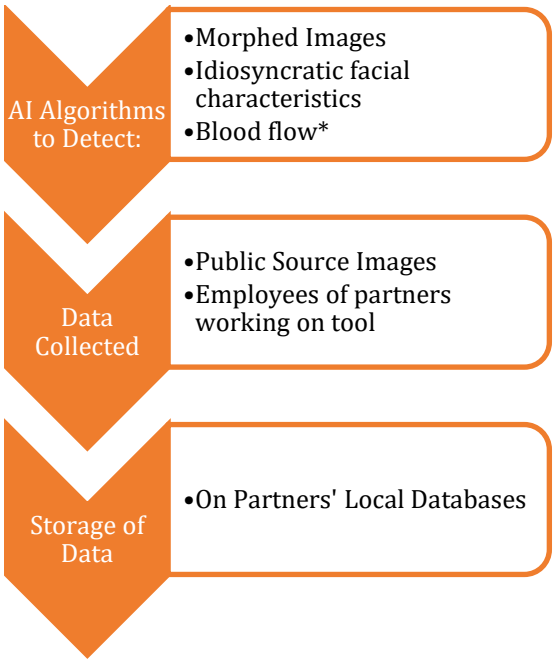
Nevertheless, AI and handcrafted algorithmic tools can help border guards identify fraudulent documents such as morphed photos thereby hindering criminals crossing borders for smuggling, trafficking, or terrorist purposes. The use of this technology at the border is quite new. Indeed, a number of governments around the world are now funding research on systems powered by artificial intelligence that can help to assess travellers at border crossings; they have not yet been implemented at the scale of biometric identity verification systems.[54] Some of the few systems already in use that automatically collect and share passenger data include select information given before flying (API - *Advance Passenger Information*) and reservation details (PNR - *Passenger Name Record*).

The D4FLY AI tools are intended to detect travel documents with a morphed photo so that two distinct individuals who look similar aim to use the same document, blood flow in faces for the purpose of detecting someone wearing a mask, and idiosyncratic facial characteristics to help border guards distinguish individuals who may appear similar to them. If successful, these tools would minimise cases of spoofing at EU borders.

Applicable standards include:

- Standardization in the area of Artificial Intelligence: ISO/IEC JTC 1/SC 42 'Artificial intelligence'

Project Phase



Deployment Phase

There are no relevant differences concerning the AI algorithms in the deployment phase, beyond that the data will be stored on the D4FLY database. If these tools have successfully included the privacy-by-design elements in the project phase as advised by TRI, there are no additional privacy risks that arise in the deployment phase. (Recall that ethics risks and opportunities will be addressed in the subsequent deliverable, D3.3.)

[51] Hern 2017.

[52] Nieva 2019.

[53] Roberts 2019.

[54] Kendrick, 2019.

*Blood flow detection will be accomplished by handcrafted algorithmic features rather than AI trained on machine learning.

Privacy Concerns and Mitigations:

Concern: The literature on privacy concerns arising from AI algorithms focuses prominently on the risk of stolen or hacked data.[55] Especially as large datasets are required to offset bias, the datasets used to fuel AI algorithms can be attractive targets for criminals.

Mitigation: As detailed above, given the robust security of the partners' databases, this risk is very low.

Concern: When information is provided by individuals, there is a general concern for data re-purposing. This concern often appears in AI algorithms designed for marketing purposes. For example, when an individual "likes" a post about French fries, he/she may later see an advertisement for a fast food chain.

Mitigation: With TRI's continued advice, partners must follow the GDPR policies explained in D1.1-Data Management Plan regarding limited purposes for data collection and use. According to GDPR Recital 33, data collected for one purpose cannot be repurposed without further consent. According to Recital 50, the following factors should be included when ascertaining whether the further processing of personal data is compatible with the original purpose:

- any connection between the original purpose and the purposes of the intended further processing
- the context in which the data was collected
- the data subject's relation to the controller and how this may affect the subject's reasonable expectations with regard to further processing
- the nature of the personal data
- the consequences for the data subject of the intended further processing
- whether the original processing operations and the new ones are subject to the appropriate safeguards.

This list is not exhaustive and all issues that are relevant in the individual case must be included in the appraisal.

Concern: As with any AI algorithm designed to identify individuals or their personal characteristics, there is a serious concern of bias, especially against women, members of ethnic minority groups, and transgender individuals.[56] Studies have found that for one-to-one facial matching algorithms, most systems had a higher rate of false positive matches for Asian and African-American faces over Caucasian faces, sometimes by a factor of 10 or even 100.

Mitigation: Technical partners have discussed these concerns with TRI. The mitigation planned is to "train" the algorithms on as diverse a dataset as possible. As this concern is more about ethics than privacy, it will be addressed in full in the next deliverable, D3.3, which focuses on ethics concerns and their mitigations.

[55] See, e.g. Tucker, 2019.

[56] Hao, 2019.

Concern: This last mitigation raises a further concern. In order to offset bias, AI algorithms must be trained on an immense set of data. The size of the dataset creates an apparent conflict with data minimisation and limited purpose regulations.

Mitigation: First, HHI will use publicly available datasets. Second, TRI and HHI will discuss the validity of the following research to see if it is applicable to D4FLY. If not, alternative methods must be identified.

Matthew Rosenquist, a Cybersecurity Strategist and Industry Advisor, offers the following technical suggestions:

> "**Federated learning** (aka collaborated learning) makes possible the training of algorithms without local data sets being exchanged or centralized. It's all about compartmentalization, which is great for privacy, but it difficult to set up and scale. Additionally, it can be limiting to data researchers that are desperate for massive data sets containing the rich information needed for training AI systems.
>
> **Differential privacy** takes a different approach, attempting to obfuscate the details by providing aggregate information but not sharing specific data, i.e., "describe the forest, but not individual trees". It is often used in conjunction with federated learning. Again, there are privacy benefits but it can result in serious degradation of accuracy for the AI system, thereby undermining their overall value and purpose.
>
> **Homomorphic encryption**, one of my favorites, is a promising technology that allows for data to remain encrypted yet still allow useful computations to be done as if they were unencrypted. Imagine a class of students being asked who is their favorite teacher: Alice or Bob. To protect the privacy of the answers, an encrypted database is created containing the names of individual students and the corresponding name of their favorite teacher. While in an encrypted state, calculations could be done, in theory, to tabulate how many votes there were for Alice and for Bob, without actually looking at the individual choices by each student. Applying this to AI development, data privacy remains intact while training can still proceed. Sounds great, but in real-world scenarios, it is extremely limited and takes tremendous computing power to accomplish. For most AI applications it is simply not a feasible way to train the system."[57]

Concern: The Black Box. "AI algorithms are capable of learning from massive amounts of data, and once that data is internalized, they are capable of making decisions experientially or intuitively like humans. However, it may be impossible to tell how an AI that has internalized massive amounts of data is making its decisions. There is no straightforward way to map out the decision-making process of these complex networks of artificial neurons."[58]

Mitigation: Transparency is achieved by providing data subjects with process details. Data subjects must be informed about how the information will be used, whether this information is collected by the data subjects themselves or by others (GDPR Articles 13 and 14). Additionally, the information must be easily available, on a home page for example, and be written in a clear and comprehensible language (GDPR Articles 12). This information shall enable the data subjects to exercise their rights pursuant to the GDPR.

[57] Rosenquist 2020.

[58] Bathaee 2018: 891-892.

Admittedly, it can be challenging to satisfy the transparency principle in the development and use of artificial intelligence. Firstly, this is because the advanced technology employed is difficult to understand and explain, and secondly because the black box makes it practically impossible to explain how information is correlated and weighted in a specific process.

The data controller must always provide information concerning: the identity of the data controller; how the data controller can be contacted; the purpose of processing; the legal basis for processing; the categories of personal data that are processed; and the data subjects' right to inspect the data.

In addition, an *extended duty to inform* (GDPR Article 22) will apply when personal data is collected for automated decision making. Article 22 prescribes that AI cannot be used as the sole decision-maker in choices that can have legal or similarly significant impacts on individuals' rights, freedoms and interests.

Therefore, border guards must be exceptionally well trained on how the AI tool works. Regarding automated decisions, the crucial determiner is at what stage of the automated decision making process was a decision made, and why. Given the black box problem, this information is not always discernible. Nevertheless, the oversight and understanding of the border guards of how this system works is a main mitigation for this concern. Secondly, data controllers and designers must perform regular checks on the system to ensure reliability and compliance.

Concern: Since AI tools will be used to detect individuals' temperature patterns, the same concern arises as discussed above with thermal face scan. Someone might be running a fever for non-contagious and private medical reasons.

Mitigation: The D4FLY tool is trying to detect temperature patterns that only spoofing attempts such as mask-wearing would disrupt. It does not record or detect the individual's particular temperature, but rather the temperature *pattern* throughout the face and body in the attempt to identify an artificial substance being used for spoofing.

Apps

Designing the following apps helps D4FLY achieve its goal of making border crossing more efficient and more accurate while leaving as much personal data and autonomous use in the hands of the traveller. As the following apps are very different in their purpose and in how they will collect data, this report separates them into distinct data flow maps.

Many countries (such as Canada and the USA)[59] have developed apps to make crossing their borders more efficient. EU projects such as PROTECT[60], TRESPASS[61] and SMILE[62] have also researched the use of mobile apps for border crossings and provide usable research that D4FLY partners are referencing.

The first D4FLY app is very much a research project to see if it is possible to verify individuals' identity through their sensory data.[63] The privacy opportunity that it affords is that the data would be stored on the user's phone and only sent to the border security terminal as the

[59] CanBorder, Mobile Passport Control (MBC)

[60] http://projectprotect.eu

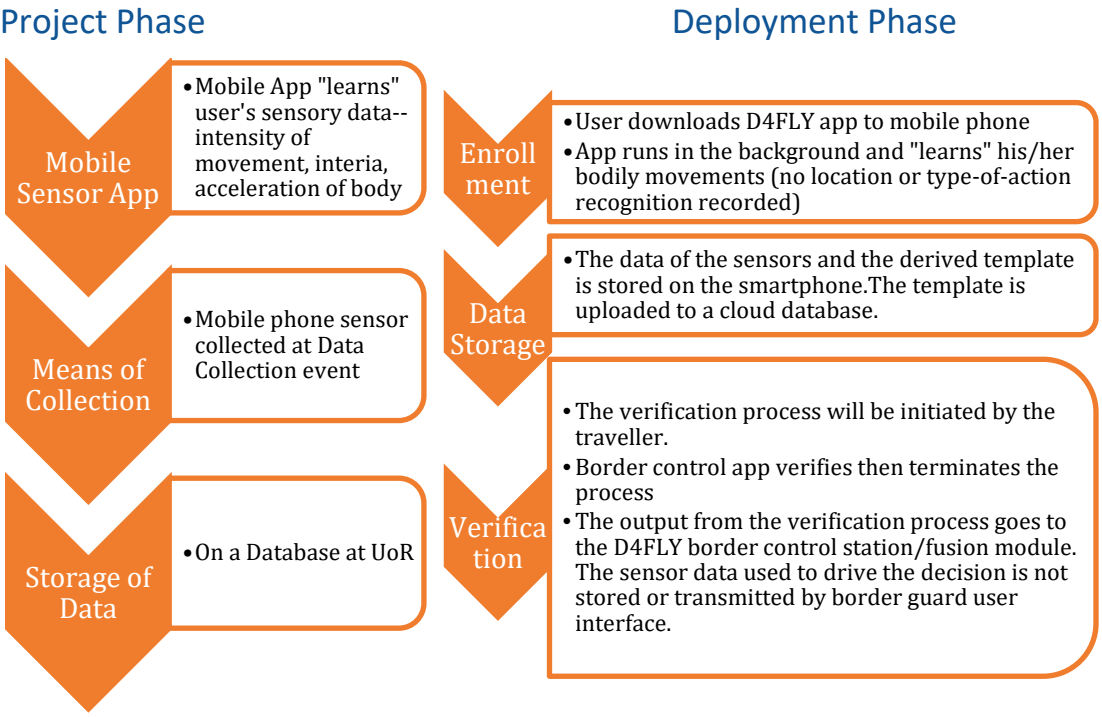
[61] http://www.trespass-project.eu

[62] https://smile-h2020.eu/smile/

[63] Gadaleti and Rossi, 2016.

traveller walks toward the border crossing point. Further, this tool would arguably maximize bona fide traveller verification speed as well as improving verification quality in difficult situations.

Smart Phone Sensors to Support Identification at Border



Privacy Concerns and Mitigations:

Concern: Potentially sensitive information about individuals can be discovered through access to sensory data and this cannot easily be protected using traditional privacy approaches.[64]

Mitigation: Malekzadeh et al. (2018a) claim they have developed: 'a privacy-preserving sensing framework for managing access to time-series data in order to provide utility while protecting individuals' privacy. We introduce *Replacement AutoEncoder*, a novel algorithm which learns how to transform discriminative features of data that correspond to sensitive inferences, into some features that have been more observed in non-sensitive inferences, to protect users' privacy. This efficiency is achieved by defining a user-customized objective function for deep autoencoders. Our replacement method will not only eliminate the possibility of recognizing sensitive inferences, it also eliminates the possibility of detecting the occurrence of them'. Please see full article listed in References for technical details.

Further, TRI will discuss with partners the applicability to D4FLY of Malekzadeh et al. (2018b). The authors claim to have developed a 'feature learning architecture for mobile devices that provides flexible and negotiable privacy-preserving sensor data transmission by appropriately transforming raw sensor data. The objective is to move from the current binary setting of

[64] Malekzadeh et al., 2018a. 2018b

granting or not permission to an application, toward a model that allows users to grant each application permission over a limited range of inferences according to the provided services. The internal structure of each component of the proposed architecture can be flexibly changed and the trade-off between privacy and utility can be negotiated between the constraints of the user and the underlying application.' This sort of flexibility provides more opportunities for autonomous choice by the user.

Concern: Use of third-party software: Most mobile apps are written by combining various functions, developed by other companies (and not the app developer). These third-party libraries help developers, for example, track user engagement (analytics), connect to social networks and generate revenues by displaying ads. However, in addition to the provided services, libraries may also collect personal data for their own use. The owners of the libraries can use this information to build detailed digital profiles of users by combining the data they collect from different mobile apps. For example, a user might give one app permission to collect his/her location, and another app access to his/her contacts. If both apps used the same third-party library, the library's developer could link these two pieces of data together. Furthermore, these libraries are often proprietary and closed-source, and cannot be easily analysed. As a result, it is common that a mobile app developer does not fully understand what data these services are actually collecting.[65] While not a security risk as such, combining sources of data can lay the ground for an attack.[66]

Mitigation: As the app is still in the research phase, the mitigation for this concern has not been described. However, partners have been informed and they intend to address the concern fully.

Concern: One might be concerned that this app tracks the individual's location or specific action types without his/her consent. Individuals might be tracked to a place of worship or entertainment venue they do not want others to know about. Or one might not have anything to hide but feel watched, and as discussed in Section 2, this could be a violation of human dignity.

Mitigation: The data will be collected from volunteers who have given their informed consent, and the data will be securely stored. It is pivotal to note that the app does not record the user's location. Nor will the app make action recognition. That means it cannot identify which type-of-action or particular kind of movement the user is making. It does collect data on acceleration, intensity of movement, and inertia of the body.

Concern: It might be seen as a privacy risk that the app is learning the traveller's behaviour. Individuals may want to keep certain kinds of actions or action-types private, or as mentioned above, it may be intrinsically a violation of privacy and, consequently, an affront to an individual's dignity.

Mitigation: This app does not learn location nor can it recognize particular actions or kinds of actions. Hence personal data in addition to identity is not recorded. See also, first mitigation for this tool.

Concern: An individual could suffer an accident which might alter his/her movements such that he/she can no longer be identified by the mobile sensor app.

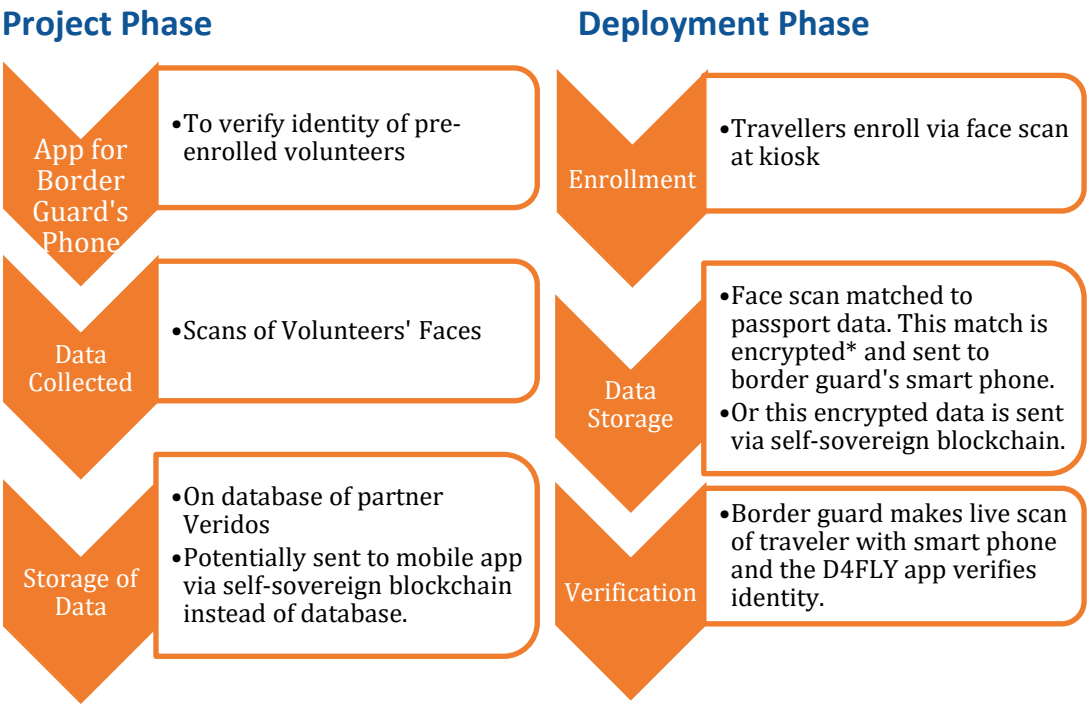
Mitigation: The mitigation is that this app is not foreseen to be a stand-alone identity verifier. It is too inaccurate to date, can easily be spoofed, and individual's movements can change in

[65] Vallina-Rodriguez, 2016.

[66] ENISA, 2017.

ways their faces or irises cannot. The app is foreseen as one among many biometrics that could be bundled together to potentially help verify a traveller's identity.

Border Guard App to be used in a confined space, e.g. a coach



* The pre-enrolled data is downloaded and stored on the smartphone in an internal database in encrypted format. The database resides in internal memory on the smartphone and can only be accessed by the app and not from outside. The data content inside the database will remain encrypted at all times. The storage of the data in the database is temporary, as soon as the process of verification ends, the encrypted data content is deleted from the database.

Privacy Concerns and Mitigations:

Concern: As with any biometric collection technology, developers must ensure that the tool works equally well on any person regardless of ethnicity, gender, age, or disability.

Mitigation: See mitigation described under biometric section.

Opportunity: The use of blockchain may enhance the privacy protection of the data. In theory, self-sovereign blockchain could be a better privacy-option than storing the data on a centralized database. A centralized management model of user's identity when accessing and using systems is not necessarily applicable at least as a single usable option, due to many reasons. One of the reasons is privacy. More precisely, the user should be given back the control of his anonymity and his own data. Another reason to transfer into decentralized systems is security-related: general security is also improved due to lack of single point of failures and trust into systems added because of the immutable nature of decentralization. Furthermore, when individuals remain in possession of their own personal data, the right to be forgotten is promoted as there exists no sharing of the data with companies or other entities.

However, when applying distributed ledger technology, there are naturally various implementation variations already inside one technology, not to mention choosing between

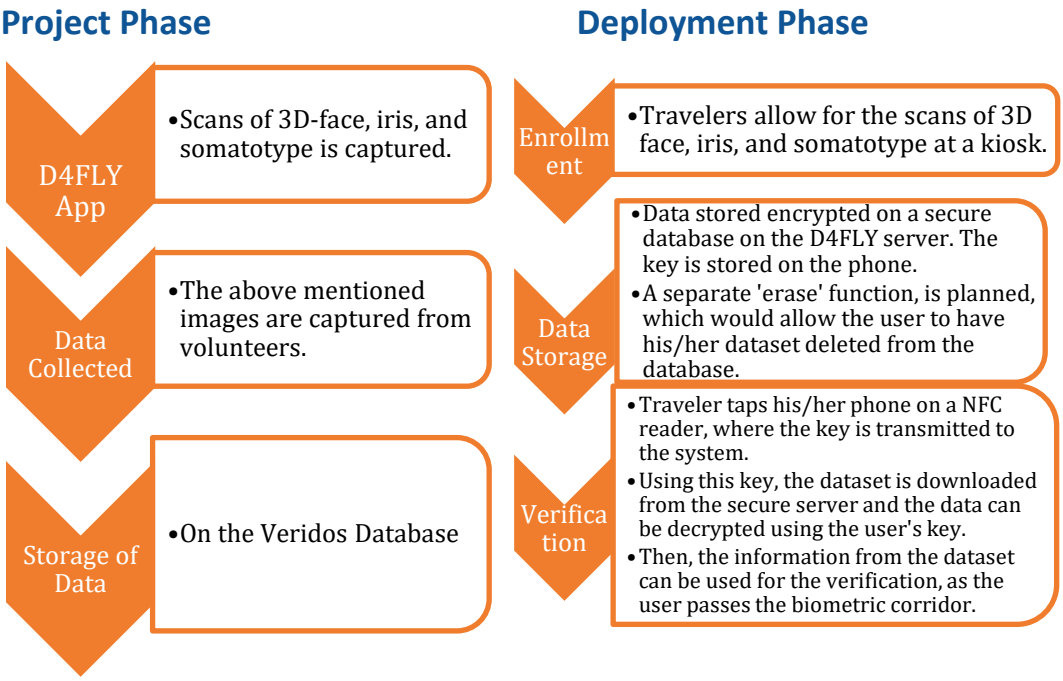
several technology platforms. Decisions have to be made regarding for instance the used programming language, necessary interfaces and so on. When these are applied in the context of digital identity, the considerations that most affect the end result are first of all with the choice between public vs private blockchain and the choice of used consensus mechanism. Awareness towards the type of data that will be stored on the blockchain is also necessary (D6.4).

Concern: Despite the opportunities it affords regarding autonomous use, blockchain elicits serious security concerns. Researchers have detected that it is possible for an attacker "to take control of one node's communications and fool it into accepting false data that appears to come from the rest of the network can trick it into wasting resources or confirming fake transactions."[67]

Mitigation: No personal identity information should be sent via blockchain. With TRI, partners should discuss the merits of including malicious nodes detection and restricting access technology for the network layer, transaction mixing technology, encryption technology and limited release technology for the transaction layer, and some defense mechanisms for blockchain applications layer.[68]

Smartphones as alternative carrier for identity data

This task will research and develop smartphone applications that can be used as alternative data carriers to travel documents. The term 'alternative' in this statement does not relate to 'data', but to 'data carriers'; it seeks to provide an alternative to passports, but not an alternative to the data stored on / in the passport.



* In the envisioned deployment phase, the data storage shall be on a dedicated secure server, and not - as planned for the prototype during the project – on a "D4FLY server", which is an Amazon service.

[67] Orcutt, 2018.

[68] Liehuang et al., 2017.

*As this app collects biometric data, it shares the same concerns and mitigations described under the biometrics section.

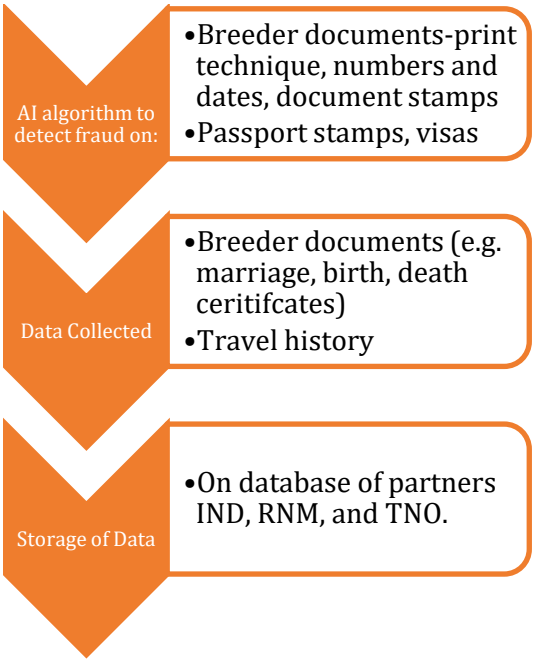
*It shares the concerns and mitigations corresponding to the other apps such as avoiding the use of third-party library software.

Document Authentication

The principal aim of this tool is to detect forgeries or fakes among travel documents and "breeder" documents (documents such as birth certificates used to apply for travel documents or residency permission). The end user, IND, reports receiving 44,400 applications for naturalisation, 85,940 applications for residency, and 29,340 applications for asylum in the Netherlands in 2019.[69] Approximately 1/5 of applications are suspect. The development of this tool would automate the verification of the non-suspicious cases, thereby freeing up more capacity for the human inspection of the suspicious cases. The AI tool can detect fraudulent travel stamps, altered dates and photos, incorrect information concerning issuing authorities, as well as inconsistencies throughout the entire document. Please see D1.1-DMP for details concerning the anonymisation of the data.

AI Algorithms

Project Phase



Deployment Phase

There are no relevant differences concerning the AI algorithms to check travelers' documents in the deployment phase. If these tools have successfully included the privacy-by-design elements in the project phase, there are no additional privacy risks that arise in the deployment phase. (Recall that ethics risks and opportunities will be addressed in the subsequent deliverable, D3.3.)

Privacy Concerns and Mitigations:

[69] https://ind.nl/en/news/Pages/Annual-report-2019-number-of-applications-increases-sharply.aspx

Concern: Partners are gathering personal data including: digital scans (images) of breeder documents (e.g., marriage certificates, birth certificates) and travel documents. Breeder documents are documents that can support the identification of a person, but later also travel documents (e.g., passport). The breeder documents can be related to identity, birth, marital status, death, place of residence, issuer. The documents contain personal information of individuals, such as names, gender, address (from the past), photograph, date of birth, place of birth, identification number, nationality, name of partner, names of children. Personal data, when shared, is always a concern. Additionally, data could be lost or stolen.

Mitigation: Partner TNO has created an anonymization tool for the documents that they intend to use. The anonymization is for dissemination/communication purposes, but is not possible to implement prior to their processing of the documents or their data. Nevertheless, TNO, IND, and RNM have Dutch legal authority to collect and process personal data from the above-mentioned documents.

Algemene machtigingsregeling IND (General authorization scheme IND):

The Minister gives a mandate to the head of the Immigration and Naturalization Service to take all decisions on his behalf, to finalize all documents, to sign all outgoing letters with regard to all matters arising from the implementation of the Aliens Act 2000 and the Kingdom Act on the Dutch nationality, unless otherwise provided by law or if the nature of the competence precludes this.

Schengen Border Code & Aliens Act 2000:

The RNM is in charge of (Schengen) border control within The Netherlands and is charged with supervision of the observance of the statutory provisions relating to aliens.

Vreemdelingencirculaire 2000 Paragraaf C1/2 (Foreigner Act implementation guidelines 2000 Paragraph C1/2):

The official in charge of border control or the supervision of aliens who has found evidence at the foreigner, informs the foreigner that the foreigner must provide proof of authenticity, which are not documents for crossing the border and / or identity documents, to the IND (BDOC).

TNO law:

The Dutch research institute TNO is an institute founded by Dutch law that has the legal task to conduct scientific research on behalf of governmental institutions.

Further, these partners have completed a DPIA for the tasks related to this tool, As such they, along with oversight and advice from ethics partner TRI, have considered the privacy and data protection risks and mitigations associated with the development of this tool. TRI has also advised the partners on which regulations are relevant for the sharing of this data among the relevant and authorized partners. Readers should refer to this document, as it does not need to be replicated here in full.

Additional Tests and Tools

*It deserves special mention that each of these tools will also be used in field tests, at the locations of the end users. These are: Lithuanian/Belarus land border, Schiphol airport in Amsterdam Piraeus Cruise Ship Port in Greece, Eurostar Arrival Station in St. Pancras, UK and Coquelles, France.

The maps of the data flows, and the noted privacy concerns and mitigations are the same in the field test environments. The field tests will also involve prior enrolment in a D4FLY kiosk. This enrolment procedure ensures informed consent for the field tests and is included in the data maps corresponding to the Deployment Phase.

*D4FLY partner OVDK is developing a new kinegram to be embedded in travel documents. A kinegram is a moving hologram embedded in bank notes or travel documents. It contains no personal data and has no impact on privacy beyond adding additional security measures to travel documents, the advantages of which were described at the beginning of Section 3 of this report.

5 FUTURE RECOMMENDATIONS

Although some recommendations have already been discussed in sections 3 and 4 and in the DMP (D1.1), they are summarized here, and additional recommendations have been made. As the project is currently in M8, it is emphasized that this report is preliminary and will be updated as the project continues. All ethics-related recommendations will be included in the subsequent deliverable D3.3, as described in the Grant Agreement.

Monitoring the implementation of these recommendations will continue to proceed through TRI's use of a ledger system that has already been used with success to monitor activities with human subjects. TRI will make an Excel table containing all of the recommendations of this report, containing columns displaying 3-month intervals throughout the project. TRI and partners will discuss during WP meetings which recommendations have been or will be implemented.

To emphasise privacy and data protection throughout all stages of the project, we have developed guidelines to be embedded in the design process. These are a direct response to the risks raised in Section 3 of this report. As the D4FLY tools will improve the accuracy and efficiency of identity verification and document authentication for EU border crossing points, the recommendations could also serve as a basis for future best practice guidelines in the area of automated digital identification and authentication more broadly.

The following recommendations are based on an analysis of possible solutions to the identified risks as well as end user needs and case reports drafted in D2.1 (M12), which has already been shared internally in the consortium. This step involves developing strategies to eliminate, avoid, reduce or transfer the privacy and data protection risks. These strategies could include technical solutions, operational and/or organisational controls and/or communication strategies (e.g., to raise awareness). The primary aim of these solutions is to ensure that the D4FLY tools respect the privacy of citizens and residents while meeting end user needs. At times, the recommendations are intentionally non-specific and refer solely to a principle that the technology should respect, to allow the technology developers flexibility as to implementation. Thus, these recommendations represent a suggestion and/or set of suggestions, but those building the D4FLY tools will develop the means. They will be used as a springboard for elucidating what technical solutions can best achieve the recommendations within the confines of the D4FLYproject. These recommendations will also be revisited by the authors of this PIA as the project progresses and will be further elaborated and implemented in conjunction with the technology partners in preparation for the pilot and testing phases.

We make the following **recommendations:**[70]

P1 Each partner must understand that it is responsible for its actions, for compliance with the GDPR and safeguarding EU fundamental rights.

P2 Partners must ensure the highest possible security standards for the storage of personal data.

P3 Partners will follow good data governance practices and will collect no more personal data than is necessary. D4FLY partners will ensure that only adequate, relevant, and

[70] Because this PIA is separate from the subsequent EIA and the legal analysis, we adopt the system of using P1…Pn for privacy recommendations, E1…En for ethics recommendations, etc.

limited to what is needed for their tasks. As such, partners will not collect extraneous information from participants where such personal data is not required to complete the task at hand. If additional personal data beyond what is necessary is provided to partners, they will destroy it as soon as is practicable.[71] Additionally, data collected for one purpose cannot be repurposed without further consent.[72]

P4 Privacy-aware control structures for the operation of data collection tools will be implemented to ensure that data collection tools operate in compliance with the defined legal and ethical guidelines and reflect the technical specifications provided by the D4FLY tools. (See D1.1-DMP.)

P5 D4FLY must abide by the principle of non-discrimination to the extent that partners do not detect any inherent bias towards race, gender, age, location, etc. in the biometric and AI tools developed. The principle of non-discrimination guards against adverse distinction in the treatment of different groups or individuals based on race, colour, sex, gender, age, language, religion, political or other opinion, national or social origin, property, birth, disability, health, sexual orientation or other status. Partners ought to conduct regular audits on the system to test whether such biases have been developed.

P6 When D4FLY partners process data by volunteer human subjects, they must receive the subjects' informed consent. This means that users understanding why data is being collected, how it will be used and stored, by whom, and for what purpose. This also means that anyone who volunteered data for the project can withdraw their consent at any time, and elect to have their data destroyed, deleted and returned to them. To meet this end, TRI has developed an understandable and easily accessible information and consent template for partners to use, keeping in mind they must individualize it to their activities. (See DMP, D1.1.)

P7 While privacy notices or terms of use may set out the basis for data processing practices, such texts are not always understandable or effective. Most commonly, privacy notices and terms of use are simply accepted by users without any engagement. [73] D4FLY should therefore pursue alternative ways of presenting privacy related information.

Suggestions include:

(a) Explaining how information is disclosed and the opt-out choices for each kind of disclosure. This should come directly under the caption. Other elements such as data collected should come later in the notice.
(b) A checkbox to indicate whether the consumer does or does not have an opt-out choice for each category of data disclosure.
(c) Use the phrase "opt-out" rather than "choice".
(d) Create a dialogue rather than notification structure.

P8 In order to ensure that both end users and travellers are aware of how travellers' personal data is collected and stored, the project will itself need a clear picture of the data that is being collected, and how this data is being used. D4FLY will need to

[71] Art.5(1)(b), GDPR

[72] Recital (33), GDPR

[73] Bogdanovic et al., 2009; Coles-Kemp & Kani-Zabihi, 2010.

determine what data must be collected in order to achieve the end goals. Subsequently, the end users and travellers need to be made aware of exactly what data will be collected, how it will be used, how it will be shared and the details of how it will be stored. This information needs to be presented in a clear and concise way, in order to ensure that travellers actually read and engage with the information in order to truly consent. Under Art. 7 EU Data Protection Directive, and similarly under the GDPR Art. 4(11), personal data can only be processed if the data subject has unambiguously given consent. (See also D1.1-DMP).

P9 As a first step and before any biometric data is collected, partners must be aware of potential privacy risks especially regarding why particular biometric data could violate someone's right to privacy.

P10 Partners designing biometric data collection devices ought to ensure that the tool does not inadvertently collect additional personal data, such as health data, unnecessary for verifying an individual's identity.

P11 Any biometric capture must work equally well on all persons regardless of racial or ethnic background, complexion, age, sex, gender, disability or other characteristics.

P12 Before any AI algorithm is used, D4FLY design partners should:

(a) Take any possible measures to minimise algorithmic discrimination and bias and to develop a strong and common ethical framework for the transparent processing of personal data and automated decision-making that may guide data usage and the ongoing enforcement of EU law.
(b) Where possible implement innovative techniques to develop auditable machine learning algorithms. Internal and external audits should be undertaken with a view to explaining the rationale behind algorithmic decisions and checking for bias, discrimination and errors.
(c) Moreover, education on the privacy aspects in the design of algorithms should be included in the training of developers.
(d) Consider making anything about the assumptions being built into the algorithm made visible/transparent to users so they can also better judge when those assumptions don't meet the situation on the ground.
e) Consider and discuss with TRI the recommendations concerning finding a balance between the black box problem and the need for transparency.
f) Determine, with help from TRI, the appropriate size of the dataset such that training the AI will avoid bias, but so that the dataset does not violate the principle of data minimisation.

P13 App developers must minimize as much as possible use of a third-party library that could inadvertently pull and combine personal data from multiple different apps.

P14 D4FLY end users (i.e. border authorities) should be trained on the technology, and trained how to interact with travellers engaging with the tools. The use of short but informative videos may be preferable over long documents, which users may lose interest in. These videos could be shown during a training workshop.

P15 As the precise details of how blockchain technology will be implemented with the D4FLY tools still need to be made clearer, partners need to be very cautious about its planning and its implementation. Owing to its still experimental stage, it has both a high potential for increased security of personal data, and yet, also a higher potential for identity theft.

P16 Do not put unencrypted personal data on blockchain.

P17 Accountability: Reflection, redress, response are the three Rs of a privacy (and ethical) impact assessment to help facilitate accountability. Are there avenues for the partners, designers, engineers, and others to reflect on what they are doing and whether it is compliant and ethical? Are there avenues to redress mistakes? Are there avenues for responding to stakeholders' concerns and proactively communicate the procedure and concerns?

P18 Technical partners ought to regularly audit their tools along the lines of these recommendations.

6 NEXT STEPS

This report does not finalise the impact assessment process, but plays a role in the continued assessment of privacy and data protection issues raised by the design, development and implementation of D4FLY tools. TRI will continue to collaborate with partners on establishing the recommendations in the design and use of the tools. To this end, the project will monitor the implementation of these recommendations as the technology is developed and undergoes changes, and/or as new risks arise and become apparent. As output of this process, the D4FLY project includes the following tasks and deliverables: “Task 3.3 Social and Ethical Impact Assessment” (M26); “Task 3.4 Border control legal analysis and policy recommendations” (M28); and “Task 3.5 Guidelines for future development” (M32).

7 CONCLUSIONS

This document is a report of the first part of the PIA+ conducted on the D4FLY tools. The first part of the PIA+ has focused exclusively on privacy and data protection concerns related to the project and its tools. Societal and ethical concerns and opportunities emerging with the development and potential deployment of the tools will be the focus of the subsequent deliverable, D3.3.

This report provided an explanation of privacy, presented applicable EU policies and legislation for partners to consider, mapped the data flows of each D4FLY tool, identified privacy and data protection concerns and mitigations, reported on the day-long privacy, data protection and ethics workshop for partners, reported on additional engagement with partners, and made preliminary future recommendations. The assessment will be revisited and updated accordingly throughout the project.

REFERENCES

Arendt, H 1964, *Eichmann in Jerusalem*, New York, Penguin.

Arora S. S., M. Vatsa, R. Singh, A. Jain 2012, 'Iris recognition under alcohol influence: A preliminary study', Proc. 5th IAPRInt. Conf. Biometrics (ICB), pp. 336-341.

Bathaee Y 2018, 'The Artificial Intelligence Black Box and The Failure of Intent and Causation', *Harvard Journal of Law & Technology* Vol 31, No. 2, pp. 890-938.

BBC 2018, 'BA boss demands action on Heathrow queues', 6 August 2018, https://www.bbc.co.uk/news/education-45080539

BBC 2020, 'Coronavirus: Amazon using thermal cameras to detect Covid-19', BBC.com, 20 April 2020, https://www.bbc.com/news/technology-52356177.

Becker, M 2019, 'Privacy in the digital age: comparing and contrasting individual versus social approaches towards privacy', *Ethics and Information Technology* 21, pp. 307–317.

Beduschi A 2020, International migration management in the age of artificial intelligence', *Migration Studies.*
https://academic.oup.com/migration/article/doi/10.1093/migration/mnaa003/5732839

Bergstedt K, T Tran, and K Waller 2018, 'Biometrics: Ethical Implications of Future Authentication Systems', *Medium*, 27 December 2018. https://medium.com/var-city-uw/biometrics-ethical-implications-of-future-authentication-systems-b0ac833b53a7

Bloustein, E 1964, "Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser", *New York University Law Review*, 39: 962–1007

Bogdanovic, D, C. Crawford, and L. Coles-Kemp 2009, 'The need for enhanced privacy and consent dialogues', *Information Security Technical Report* Vol. 14 no. 3, pp. 167– 172.

Cavoukian, A 2011, 'Privacy by Design. The 7th Foundational Principles', [Online] https://www.ipc.on.ca/wpcontent/uploads/Resources/7foundationalprinciples.pdf

Cohen, J 2002, *Regulating Intimacy: A New Legal Paradigm*, Princeton: Princeton University Press.

Coles-Kemp L & E Kani-Zabihi 2010, 'On-line privacy and consent: a dialogue, not a monologue', *Proceedings of the 2010 workshop on New security paradigms*, pp. 95–106.

Cook, J 2019, '"Racist" passport photo system rejects image of a young black man despite meeting government standards', *The Telegraph* (Online), 19 September 2019. https://www.telegraph.co.uk/technology/2019/09/19/racist-passport-photo-system-rejects-image-young-black-man-despite/

Directorate-General for Communication Special Eurobarometer 431 Report on Data Protection 2015, Survey Conducted by TNS Opinion & Social: https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf

DutchNews.nl 2020, Schiphol passport security needs stronger protection against cyber attacks', Dutchnews.nl, 20 April 2020, https://www.dutchnews.nl/news/2020/04/schiphol-passport-security-needs-stronger-protection-against-cyber-attacks/

ENISA, 2017, 'Privacy and data protection in mobile applications: A study on the app development ecosystem and the technical implementation of GDPR.'

Estrada-Jiménez J et al. 2019, 'On the regulation of personal data distribution in online advertising platforms', *Engineering Applications of Artificial Intelligence* 82, pp. 13-29.

European Convention of Human Rights, Art. 8.

European Parliament and Council, Regulation (EU) 2016/679 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, OJ L 119, Vol.59, 4 May 2016 (General Data Protection Regulation).

Finn, R, D Wright, and M Freidewald 2013, *European Data Protection: Coming of Age.* S. Gutwirth et al. (eds.). Dordrecht: Springer.

Fried, C 1970, *An Anatomy of Values*, Cambridge: Harvard University Press.

Frontex 2018, Industry Day-Biometric on the Move, https://frontex.europa.eu/research/invitations/industry-day-biometric-on-move-Voqk1n

Gadaleta M & M. Rossi, 2016, 'IDNET: Smartphone-based Gait Recognition with Convolutional Neural Networks'

Gerety, T 1977, 'Redefining Privacy', *Harvard Civil Rights-Civil Liberties Law Review*, 12: 233-96.

Gerstein, R 1978, 'Intimacy and Privacy', *Ethics*, 89: 76–81.

Guenole, N & S Feinzig, S 2018, 'The business case for AI in HR', Retrieved from https://www.ibm.com/downloads/cas/AGKXJX6M

Hannafin, M J., & S M. Land 1997, 'The foundations and assumptions of technology-enhanced student-centered learning environments', *Instructional Science* 25, pp. 167-202.

Hao, K 2019, 'This is how AI bias really happens—and why it's so hard to fix,' *MIT Technology Review* (Online). Feb. 4, 2019. https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

Hern, A 2017, 'Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind', The Guardian, 3 July 2017, https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act

IATA 2017, '2036 Forecast Reveals Air Passengers Will Nearly Double to 7.8 Billion', *Press Release* no. 55, 24 October 2017. https://www.iata.org/en/pressroom/pr/2017-10-24-01

ICO 2020, *Conducting privacy impact assessments code of practice.* [Online]

https://ico.org.uk/media/about-the-ico/consultations/2052/draft-conducting-privacy-impact-assessments-code-of-practice.pdf

Information Commissioner's Office 2008, "Privacy by Design."

International Conference of Data Protection and Privacy Commissioners 2009, *International Standards on the Protection of Personal Data and Privacy: The Madrid Resolution*. Madrid, 22(f). https://dig.watch/sites/default/files/2009_M1.pdf

Karanja, Stephen Kabera 2008, 'Privacy and Protection of Marginalized Social Groups'. *Studies in Ethics, Law, and Technology*, Vol 2 no 3.

Kendrick, Molly 209, The border guards you can't win over with a smile', BBC, 17 April 2019. https://www.bbc.com/future/article/20190416-the-ai-border-guards-you-cant-reason-with

Kroener, I & D Wright 2014, 'A strategy for operationalising privacy by design', *The Information Society* vol. 30 no. 5, pp. 355-365.

Kupfer, J 1987, "Privacy, Autonomy and Self-Concept", *American Philosophical Quarterly*, 24: 81–89

Liehuang, Z 2017, 'Survey on Privacy Preserving Techniques for Blockchain Technology', *Journal of Computer Research and Development* Vol 54, no. 10, pp. 2170-2186.

Malekzadeh M, R Clegg & H Haddadi 2018a, 'Replacement AutoEncoder: A Privacy-Preserving Algorithm for Sensory Data Analysis', https://arxiv.org/pdf/1710.06564.pdf

Malekzadeh M, A Cavallaro R Clegg, & H Haddadi 2018b, 'Protecting Sensory Data against Sensitive Inferences', W-P2DS'18: 1st Workshop on Privacy by Design in Distributed Systems, April 23–26, 2018, Porto, Portugal. https://doi.org/10.1145/3195258.3195260

McCarthy, P 2012, 'Biometric Technologies, Ethical Implications'.

Nieva, R 2019, 'Google ditched project to release 100,000 X-ray images, amid privacy concerns', CNet.com, 15 November 2019, https://www.cnet.com/news/google-ditched-project-to-release-100000-x-ray-images-amid-privacy-concerns/

Nissenbaum, H 2009, *Privacy in context: Technology, policy and the integrity of social life*, Stanford, Stanford University Press.

Norwegian Data Protection Authority 2018, Artificial intelligence and privacy Report, January 2018. https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

Orcutt, M 2018, 'How Secure is Blockchain Really?' *MIT Technology Review*, 25 April 2018, https://www.technologyreview.com/2018/04/25/143246/how-secure-is-blockchain-really/

Price N et al. 2019, Privacy in the age of medical big data', *Nature Medicine* 25, pp. 37-43.

Regan, P 1995, *Legislating Privacy*, Chapel Hill, NC: University of North Carolina Press.

Roberts, J 2019, 'Microsoft Removes Face Recognition Photos Amid Privacy Controversy'. Fortune.com, 7 June 2019, https://fortune.com/2019/06/07/microsoft-facial-recognition/

Rosenquist, M 2020, 'There is no easy fix to AI privacy problems', *HelpNetSecurity.com*. 23 January 2020. https://www.helpnetsecurity.com/2020/01/23/ai-privacy-problems/

Rowe, M & R Muir 2019, 'Big Data Policing: Governing the Machines?' *Policing and Artificial Intelligence,* Taylor & Francis, London: UK.

Sareen, P 2014, "Biometrics - Introduction, Characteristics, Basic Technique, Its Types and Various Performance Measures," International Journal of Emerging Research in Management & Technology, vol. 3, no. 4, pp. 109–119.

Solove, D 2008, *Understanding Privacy*, Cambridge, MA: Harvard University Press.

Tucker, C 2019, 'Privacy, Algorithms, and Artificial Intelligence", *The Economics of Artificial Intelligence: An Agenda*, eds. A Agrawal et al., London: The University of Chicago Press.

Vallina-Rodriguez, N et al. 2016, 'Tracking the Trackers" Towards Understanding Mobile Advertising and Tracking Ecosystem.' *NSF Haystack Project.*

Wiggers, K 2019, 'AI has a privacy problem but these techniques could fix it', *Vulturebeat.com*, 21 December 2019. https://venturebeat.com/2019/12/21/ai-has-a-privacy-problem-but-these-techniques-could-fix-it/

Wilcox, L 2017, 'Gendered bodies in securitized migration regimes', Handbook on Migration and Security, ed. Phillippe Bourbeau, Cheltenham, UK, Edward Elgar Publishing, pp. 87-104.

Wright, D 2012, 'The state of the art in privacy impact assessment', *Computer Law and Security Review* 28, pp. 54-61

Wright, G B 2011, 'Student-Centered Learning in Higher Education', *International Journal of Teaching and Learning in Higher Education*, vol 23 no. 1, pp. 92-97.

Zhang H 2019, 'The role of AI in mitigating bias to enhance diversity and inclusion', *IBM Smarter Workforce Institute*. https://www.ibm.com/downloads/cas/2DZELQ4O

ANNEX A

D4FLY: Privacy And Ethical Impact Assessment Workshop

22 January 2020

Time	Item	Participants	Room #
09:30 – 09:45	**Arrival & Registration**	All	TBC
09.45 – 11.15	**Part I** • Agenda & Introductions • Introduction: *What is an E/PIA?* • Data flows of D4FLY Tools	All	TBC
11.15 – 11:30	**Coffee break**	All	TBC
11:30 – 12:30	**Part II** • Thinking about ethics, privacy, and data protection • Methodology, principles, and conflicts	All	TBC
12:30 – 13:30	**Lunch** **&** **Meet the Stakeholders**	All	TBC
13:30-15:00	**Part III** • Ethics, Privacy, and Data Protection Concerns and Opportunities: group exercise	All	TBC

Time	Item	Participants	Room #
15:00 – 15:15	**Refreshment break**	All	TBC
15:15 – 16.15	**Part IV** • Mitigations of concerns: group exercise		
16.15 – 16.30	**Part V** • Conclusion & next steps	All	TBC