Data Pitch

H2020-ICT-2016-1

Project number: 732506

D3.9 Legal and privacy aspects of transnational, cross-sector data sharing in open innovation

Coordinators: Professor Sophie Stalla-Bourdillon and Dr Laura Carmichael

With contributions from: Dr Pei Zhang (Developer)

Quality reviewer: Jérémy Decis

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Nature	Report
Work package	3
Contractual delivery date:	31 December 2019
Actual delivery date:	31 December 2019
Version:	1.0
Keywords:	Automated decision-making, controllers, data ethics, data protection, data transit, GDPR, law, personal data, processors, profiling, regulatory compliance, restricted transfers, territorial scope.

Table of Contents

List of figures and tables

Abbreviations

AEPD = Agencia Española de Protección de Datos

AI = Artificial Intelligence

APEC = Asia-Pacific Economic Cooperation

Art.29 WP = Article 29 Data Protection Working Party

B2B = Business to Business

BCR = Binding Corporate Rules

CBPR = Cross Border Privacy Rules system

CMS = Content Management System

CNIL = Commission Nationale de l'Informatique et des Libertés

D. = Deliverable

DPA = Data Protection Authority

DPIA = Data Protection Impact Assessment

EC = European Commission

EDPB = European Data Protection Board

EDPS = European Data Protection Supervisor

EEA = European Economic Area

ENISA = European Union Agency for Network and Information Security

ESRC = Economic and Social Research Council

EU = European Union

GDPR = General Data Protection Regulation

H2020 = Horizon 2020

IAPP = International Association of Privacy Professionals

ICO = Information Commissioner’s Office

ICT = Information and Communications Technologies

IPR = Intellectual Property Rights

ISO = International Organization for Standardization

MCE = Model-Centric Explanation

NZ = New Zealand

ODI = Open Data Institute

OECD = Organisation for Economic Co-operation and Development

PIA = Privacy Impact Assessment

SCE = Subject-Centric Explanation

SME = Small and Medium Enterprises

UK = United Kingdom

USA = United States of America

UKAN = UK Anonymisation Network

V1 = Version 1

V2 = Version 2

Abstract

The D3.9 report is conceived as a supplement to the Legal and Privacy Toolkit v1 and v2 reports (available at: https://datapitch.eu/deliverables/) – and provides the last update to the Legal and Privacy Toolkit ("the toolkit") during the concluding stages of the Data Pitch programme. The objective of this report is to provide those directly involved with open innovation guidance on the key legal and privacy aspects of transnational, cross-sector data sharing. It offers practical guidance that can be understood by non-data protection specialists, including four legal decision-trees – an interactive version of these is provided through a prototype e-learning tool. Part A provides an outline of transnational, cross-sector data sharing regulation, and the different approaches taken across countries and industry sectors. Part B focuses on the key data protection aspects of transnational, cross-sector data sharing from an EU perspective.

Disclaimer: The content of the Legal and Privacy Toolkit (including any related training resources, such as the prototype e-learning tool) does not constitute legal advice. If in doubt, you should always contact a lawyer.

Executive summary

D.3.9 toolkit update: The Legal and Privacy Toolkit ("the toolkit") is a crucial component of the Data Pitch programme. The first (D3.1) and second (D3.5) versions of the toolkit were published on the Data Pitch website in June 2017 and June 2018 respectively: https://datapitch.eu/deliverables/. The toolkit covers a wide-range of key legal considerations that are likely to occur in the course of an open innovation programme – with particular focus given to data protection law. This deliverable report (D3.9) is the third and final toolkit update of the Data Pitch programme.

Objective: In accordance with the Grant Agreement, the central focus of Data Pitch Deliverable 3.9 is as follows:

> *"Legal and privacy aspects of transnational, cross-sector data sharing in open innovation."*

Definitions: For the purpose of this deliverable, the terms used by the Grant Agreement objective are defined as follows:

Data sharing. One or more parties *"communicate, disclose or otherwise make particular data available"** to one or more other parties. Data sharing can (i) occur within a specific organisation or between third party organisations; and (ii) be "systematic"** or "exceptional"**. [*EDPS Position Paper on the transfer of personal data to third countries and international organisations by EU institutions and bodies, 2014 (p. 7); **ICO Data Sharing Code of Practice, 2011 (p. 7).]

Transnational data sharing. One or more parties communicate, disclose or otherwise make particular data available to one or more other parties in third countries or international organisations.

Cross-sector data sharing. One or more parties communicate, disclose or otherwise make particular data available to one or more other parties in distinct areas (e.g. another industry).

Open innovation. Organisations work with external partners (e.g. data providers and SMEs) and/or obtain insights from external sources in order to develop high impact, cutting-edge ideas, methods, products and/or services (e.g. in the course of an accelerator).

Focus on transnational, cross-sector data sharing: Open innovation acceleration programmes strive for the development of high impact, cutting-edge ideas, methods, products and/or services. In order to bring these innovative ideas to fruition, participants are required to share and (re)use data. Given that transnational, cross-sector data sharing is a core component of the modern economy, it is highly likely open innovation programmes will involve data sharing that crosses national borders and sectors within the EU and beyond. Guidance is therefore given on the key legal and privacy aspects of transnational, cross-sector data sharing in open innovation that can be understood by non-legal specialists through devised training materials.

This D3.9 report is divided into three parts:

- **Part A – An outline of transnational, cross-sector data sharing regulation**. The regulation of (transnational, cross-sector) data sharing and re-usage varies across the world – from complete absence of regulation to legislation used to constrain and/or incentivise data sharing. Part A therefore raises awareness of how different types of law act concurrently in relation to a data sharing arrangement, and how legal rights and duties may differ between sectors and countries (including within the EU bloc).
- **Part B – Key data protection aspects of transnational, cross-sector data sharing: an EU perspective**. Given that data shared (i) is *or* could likely become personal data and (ii) are likely to traverse national borders and/or sectors within an open innovation environment, it is of paramount importance that those involved with open innovation remain compliant with the General Data Protection Regulation (GDPR). Part B therefore focuses on some of the key data protection aspects of transnational, cross-sector data sharing from an EU perspective:

1. Restricted transfers to third countries and international organisations outside the European Economic Area (EEA) under the GDPR.
2. The territorial scope of the GDPR.
3. Automated decision-making and profiling.

Guidance is provided on these three areas through: (i) an overview of existing, authoritative guidance; and, (ii) the creation of four legal decision-trees that aim to help raise-awareness by communicating these three key aspects of the GDPR in a simple way to non-data protection specialists.

- **Part C – The development of training materials.** Part C explains how Part B led to the development of further legal training materials in the form of an update to the prototype e-learning tool on data protection and the basics of mapping data flows created as part of the Legal and Privacy Toolkit v2.

Conclusions. This report concludes by providing an overview of the key legal and privacy aspects related to transnational, cross-sector data sharing, including **a quick checklist on transnational, cross-sector data sharing in open innovation for Data Providers and Participating SMEs to consider**:

a) **Map your data flows** to help determine whether your planned data processing involves any transnational and/or cross-sector data sharing. See Appendix B to this report for more information.
b) **Assess what regulatory constraints apply** to your planned data processing – such as applicable laws and jurisdictions. Consider any blanket bans and sector/process/service-specific constraints and requirements (e.g. sector-specific codes of best practice, industry standards, and other legislative requirements).
c) **Ensure you have the authority to share data and/or rights to (re)use data** for your planned activity. Appropriate and effective rights management and clearance is imperative.
d) **Check how any legal agreements and/or licensing arrangements** (that you plan to enter into) may constrain (transnational, cross-sector) data sharing and re-usage – and how this could potentially limit your planned data processing activity.
e) Identify and ensure that all **personal data transfers that take place to third countries and international organisations** in the course of an open innovation programme (and elsewhere) **are lawful**.
f) Identify and ensure that all **cross-border data sharing within the EEA** in the course of an open innovation programme (and elsewhere) **are lawful** – i.e. comply with any applicable national data protection variations and/or pertinent sectoral laws.
g) Identify and ensure that all other transnational data sharing activities are lawful.
h) **Determine which parties are the controllers and processors for a particular data processing activity** in order to assess the possible extent of your legal responsibilities and liabilities for the planned data processing in question.
i) **Make sure you are familiar with the minimum legal standards for privacy information** outlined by Articles 13, 14 and 15 of the GDPR. Inform data subjects about your processing activities, including any transnational transfers of personal data.
j) Identify and ensure that all **profiling and automated decision-making activities** in the course of an open innovation programme (and elsewhere) **are lawful**.
k) **Do not engage in any activities that involve solely automated decision-making, including profiling that produces legal or similarly significant effects** unless a legal exemption applies.
l) **Open innovation activities that involve automated decision making and/or profiling should be explained** to the Consortium, data subjects and where applicable the Data Provider involved. At minimum, you should provide information about: (i) purpose and legal basis for your activities; including, the use or derivation of sensitive data, characteristics and/or preferences; (ii) the meaningful logic involved, including its rationale; and, (iii) the significance and envisaged consequences of the profiling, including any intended future processing and ways in which automated decision-making may affect the data subject. Note: there is no need for complex mathematical explanations – subject-centric explanations should be used where possible.
m) **Adhere to data ethics principles** specified by the open innovation programme and from authoritative sources.
n) **Keep up-to-date with authoritative guidance** on transnational, cross-sector data sharing, including automated-decision making and profiling.

Appendices: Given this report comprises the last update to the toolkit, the appendices summarise the main findings from its three toolkit reports. Firstly, **Appendix A** provides a toolkit overview, including its objectives and final configuration. Secondly, **Appendix B** presents a summary of the key legal and privacy aspects of data sharing and (re)usage for anyone involved in a data sharing scheme to consider. Finally, **Appendix C** provides copies of the key supporting documents used by the Data Pitch programme as part of its compliance strategy.

Disclaimer: The content of the Legal and Privacy Toolkit (including any related training resources, such as the prototype e-learning tool) does not constitute legal advice. If in doubt, you should always contact a lawyer.

1. Introduction

1.1 D3.9: Legal and Privacy Toolkit update

The Legal and Privacy Toolkit ("the toolkit") [3], [4] is a crucial component of the Data Pitch programme.[1] The primary focus of the toolkit is to provide an overview of the key aspects of the legal and regulatory framework that applies to data sharing and re-use principally for closed data within the open innovation acceleration environment. The toolkit therefore covers a wide-range of key legal considerations that are likely to occur in the course of an open innovation programme – with particular focus given to data protection law. The first (D3.1) and second (D3.5) versions of the toolkit were published on the Data Pitch website in June 2017 and June 2018 respectively: https://datapitch.eu/deliverables/.[2] This deliverable report (D3.9) constitutes the third and final toolkit update of the Data Pitch programme.

Note: Given this report comprises the last update to the toolkit, the appendices summarise the main findings from its three toolkit reports. Firstly, **Appendix A** provides a toolkit overview, including its rationale, configuration and relation to other parts of the Data Pitch Consortium strategy for handling data processing issues under the programme. Secondly, **Appendix B** presents a summary of the key legal and privacy aspects of data sharing and (re)usage for anyone involved in a data sharing scheme to consider. Finally, **Appendix C** provides copies of the key supporting documents used by the Data Pitch programme as part of its compliance strategy.

1.2 Objective

In accordance with the Grant Agreement, the central focus of Data Pitch Deliverable 3.9 is as follows:

"Legal and privacy aspects of transnational, cross-sector data sharing in open innovation." [Underlining added for emphasis.]

1.2.1 Objective definitions

For the purposes of the deliverable, the terms used by the Grant Agreement objective are defined as follows:

Data sharing: One or more parties *"communicate, disclose or otherwise make particular data available"* [1, p. 7] to one or more other parties. Data sharing can (i) occur within a specific organisation or between third party organisations; and (ii) be "systematic" or "exceptional" [2, p. 9].[3]

Transnational data sharing: One or more parties communicate, disclose or otherwise make particular data available to one or more other parties in third countries or international

[1] The principal motivation for the Data Pitch programme is to support successful applicants (i.e. start-ups) with their high-impact, innovative and data-centric business ideas, products and services that directly respond to the specific challenges defined by the programme [122]. This support is given in numerous forms, mentoring and training services to financial assistance and rights to re-use valuable data that would otherwise remain inaccessible i.e. closed data. For more information about closed, shared and open datasets – refer to the data spectrum: [117].

[2] In brief, the Legal and Privacy Toolkit v1 [3] is to ensure that all those involved with Data Pitch: (a) are made aware of the key legal rights that arise in relation to data sharing as part of the programme; and, therefore (b) adhere to any legal obligations that concern these data sharing activities. This legal guidance is achieved through the provision of an overview concerning the legal and regulatory framework that applies to data sharing and data reuse. This overview: (i) sets out the key considerations that govern the data sharing arrangements between the parties involved in the programme; (ii) maps the relevant legal issues that arise in the context of data sharing; and (iii) outlines a methodology for handling these legal issues in a suitably risk-averse manner. This framework aims to treats data ethically and responsibly, with comprehensive, yet pragmatic guidance on data disclosure and its handling. The Legal and Privacy Toolkit v2 is a toolkit update that extends the data protection guidance provided in the first version (at the half-way point – M18 – of the Data Pitch programme). In accordance with the Grant Agreement, the central focus of Data Pitch Deliverable 3.5 is as follows: *"A data situation model to assess anonymisation practices of data owners that can be used as a guide by data owners themselves before releasing their data."* The Legal and Privacy Toolkit v2 focuses on data mapping as an effective and practical approach to the creation of data situation models.

[3] Note that this definition is based on: (i) Information Commissioner's Office (ICO) Data Sharing Code of Practice [2, p. 9]: *"the disclosure of data from one or more organisations to a third party organisation or organisations, or the sharing of data between different parts of an organisation […] two main types of data sharing: [/] systematic […] and, [/] exceptional […]"*; and, (ii) [1, p. 7]. For more information on the key types of business-to-business (B2B) data sharing models see: [150, p. 5] and [151, pp. 60-65].

organisations.[4]

Cross-sector data sharing: One or more parties communicate, disclose or otherwise make particular data available to one or more other parties in distinct areas (e.g. another industry).[5]

Open innovation: Organisations work with external partners (e.g. data providers and SMEs) and/or obtain insights from external sources in order to develop high impact, cutting-edge ideas, methods, products and/or services (e.g. in the course of an accelerator).[6]

1.3 Report overview

Open innovation acceleration programmes strive for the development of high impact, cutting-edge ideas, methods, products and/or services. In order to bring these innovative ideas to fruition, participants are required to share and re-use data. Given that transnational, cross-sector data sharing is a core component of the modern economy,[7] it is highly likely – as an organisation involved in an open innovation programme – you will need to share data with and/or (re-)use data from other individuals and organisations, across national borders and sectors within the EU and beyond. It is therefore crucial that you are able to understand the key legal and privacy aspects pertinent to transnational, cross-sector data sharing activities.

This report is conceived as a supplement to v1 and v2 of the Legal and Privacy Toolkit. It aims to raise-awareness of the key legal and privacy aspects of transnational, cross-sector data sharing in open innovation by offering practical guidance that can be understood by non-legal specialists through devised training materials.

This D3.9 report is divided into the following parts:

- **Part A – An outline of transnational, cross-sector data sharing regulation**. The regulation of (transnational, cross-sector) data sharing and re-usage varies across the world – from complete absence of regulation to legislation used to constrain and/or incentivise data sharing. Part A therefore raises awareness of how different types of law act concurrently in relation to a data sharing arrangement, and how legal rights and duties may differ between sectors and countries (including within the EU bloc).
- **Part B – Key data protection aspects of transnational, cross-sector data sharing: an EU perspective**. Given that data shared (i) is *or* could likely become personal data and (ii) are likely to traverse national borders and/or sectors within an open innovation environment,**[8]** it is of paramount importance that those involved with open innovation remain compliant with the General Data Protection Regulation (GDPR). Part B therefore focuses on some of the key data protection aspects of transnational, cross-sector data sharing from an EU perspective:

  1) Restricted transfers to third countries and international organisations outside the European Economic Area (EEA) under the GDPR.
  2) The territorial scope of the GDPR.
  3) Automated decision-making and profiling.

  Guidance is provided on these three areas of key data protection considerations through: (i) an overview of existing, authoritative guidance; and, (ii) the creation of four legal decision-trees that aim to help raise-awareness by communicating these three key aspects of the GDPR in a simple

[4] For instance, Lexico [148] defines "transnational" as: *"[e]xtending or operating across national boundaries"*.

[5] For instance, "cross-sectoral" is defined by Lexico [149] as: *"[r]elating to or affecting more than one group, area, or section"*. E.g. see [157] for further background information on cross-sector data sharing – in particular data collaboratives.

[6] For instance, "innovate" is defined by Lexico [153] as: *"[m]ake changes in something established, especially by introducing new methods, ideas, or products"*. Furthermore, Henry Chesbrough [154, p. 1] defines open innovation as follows: *"the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively"*. Note that "open" innovation does not necessarily equate to "free" innovation – as it is common for some open innovation programmes to involve licensing fees and other financial agreements [152, p. 9]. For further information on open innovation see e.g.: [113] for a brief history of open innovation; [110] on innovation accelerators; [118] for an European Commission (EC) list of online resources on open innovation; [115] for an interview with Henry Chesbrough on open innovation; and, [155] & [156] on open innovation in practice.

[7] Transnational data sharing has become an essential aspect of the global economy, as summarised by J. Manyika et al. [8, p. 11]: *"Flows of physical goods and finance were the hallmarks of the 20th-century global economy, but today those flows have flattened or declined. Twenty-first-century globalization is increasingly defined by flows of data and information."*

[8] E.g. as is the experience of Data Pitch.

way to non-data protection specialists.

- **Part C – The development of training materials.** Part C explains how Part B led to the development of further legal training materials in the form of an update to the prototype e-learning tool on data protection and the basics of mapping data flows created as part of the Legal and Privacy Toolkit v2.

This report then concludes by summarising the key points from Parts A-C.

1.3.1 Disclaimer: important notice to readers

Disclaimer: The content of the Legal and Privacy Toolkit (including any related training resources, such as the prototype e-learning tool) does not constitute legal advice. If in doubt, you should always contact a lawyer.

2. Part A – An outline of transnational, cross-sector data sharing regulation

2.1 Brief overview

The regulation of transnational, cross-sector data sharing and re-usage varies across the world.[9] In some countries, there is a complete absence of regulation for transnational, cross-sector data sharing [5, p. 17]. Whereas, in other countries, legislators may decide to constrain certain types of transnational, cross-sector data sharing [5, p. 17].[10] Such regulatory constraints may take the form of a "blanket ban" on transnational, cross-sector data sharing [6]. In other cases, regulatory constraints can target a specific sector – e.g. *"personal, health, accounting, tax, gambling, financial, mapping, government, telecommunications, e-commerce, and online publishing data"* [6] – or a specific process or service – e.g. *"online publishing, online gambling, financial transaction processing, and apps that provide services over the Internet (thereby bypassing traditional distribution)"* [6]. Many domains therefore shape the regulation of transnational data flows, including *"trade, internet administration, domestic public policy and human rights"* [7, p. 125].

2.2 Key reasons for restricting transnational, cross-sector data sharing

Legislators may decide to constrain certain types of transnational, cross-sector data sharing for a variety of reasons, in particular:

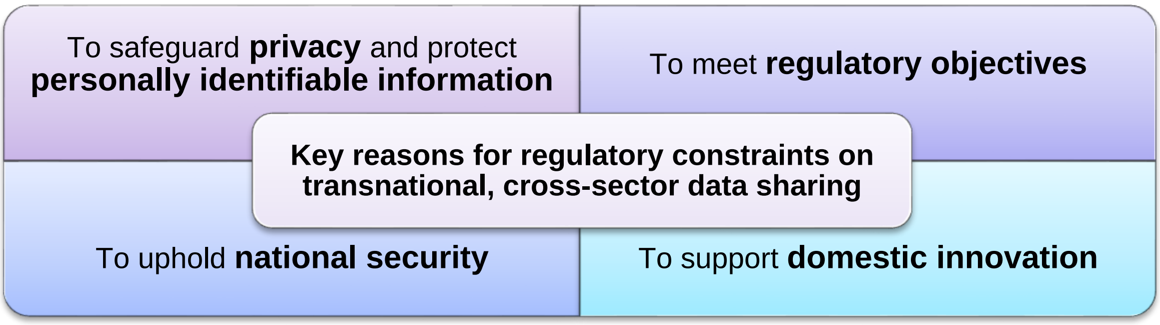

To safeguard **privacy** and protect **personally identifiable information**
To meet **regulatory objectives**
Key reasons for regulatory constraints on transnational, cross-sector data sharing
To uphold **national security**
To support **domestic innovation**

Figure 1 Four key reasons for regulatory constraints on transnational data flows. Toolkit diagram based on analysis from Francesca Casalini and Javier López González, 2019 [8, p. 13].

[9] For instance, Francesca Casalini and Javier López González [5, p. 17] provide an indicative taxonomy of approaches to cross-border data flows (see original for full information): (a) *"no regulation"*, (b) *"free flow"*, (c) *"free flow conditional on safeguards"*, and (d) *"flow conditional, including on ad hoc authorisation"*.

[10] This can be referred to as data localisation. For further background information on data localisation see: [145], [114], [147], [146]. Narrow data localisation focuses on particular sectors and broad data localisation policies [104]. Also see [103] for different types of data localisation policies.

2.2.1 To safeguard privacy and protect personally identifiable information

Legislators may constrain certain types of transnational, cross sector data sharing to safeguard privacy and protect personally identifiable information [5, p. 13]. For instance, Article 44 of the General Data Protection Regulation (GDPR) restricts the transfer of personal data to third countries and international organisations outside the European Economic Area (EEA)[11] unless controllers and processors comply with the provisions of Chapter V of the GDPR.[12] The global economy depends on transnational data sharing [8] yet there is no international legal framework for data protection and privacy [9].[13]

2.2.2 To meet regulatory objectives

Legislators may constrain certain types of transnational, cross-sector data sharing in order to meet regulatory objectives – such as a legal requirement to *"access information for audit purposes"* [5, p. 13]. These regulatory objectives can be *"sector-specific"* – e.g. there may be a legal requirement to access *"banking data"* [5, p. 13].

2.2.3 To uphold national security

Legislators may constrain certain types of transnational, cross sector data sharing in order to uphold national security – e.g. to prevent the transnational flow of highly sensitive data outside national borders, and to enable *"national security services to access and review data"* [5, p. 13].

2.2.4 To support domestic innovation

Legislators may constrain certain types of transnational, cross sector data sharing in order to enable domestic innovation as part of a *"digital industrial policy"* – e.g. data are stored locally to enable access first by *"national producers or suppliers"* in order *to "encourage or help develop domestic capacity in digitally intensive sectors" "* [5, p. 13]. Such constraints can be (non-)sector-specific [5, p. 13].

2.3 Other constraints on transnational, cross-sector data sharing

Other areas may (in)directly constrain transnational data flows, such as *"intellectual property, censorship or technology issues"* [7, p. 126]. For instance, the laws applicable to data are numerous – as illustrated by following six key legal areas relevant to (transnational, cross-sector) data sharing and re-usage described by v1 of the toolkit (for further detail see **Appendix B** to this report):

Six key legal areas relevant to data sharing and re-usage

- ***"Data protection laws** – these set the rules for processing personal data.*
- ***Electronic privacy laws** – these govern the protection of privacy in the electronic communications sector.*
- ***Intellectual property laws** – these encompass a number of different rights that may be asserted of a more proprietary type, including in association with the use of data and its expression.*
- ***Competition laws** – these aim to prevent anti-competitive harm that might result from commercial activities, including from the sharing of information.*
- ***Laws of confidentiality** – these protect confidential information.*
- ***Contract laws** – these govern the ways in which private parties can agree to work together, including in respect of data sharing agreements that include certain rights and obligations regarding data usage and access. Ultimately, if terms in agreed contracts are broken, contracting parties could try to enforce such terms in a court of law."*

[11] The EEA comprises the contracting parties to the EEA Agreement, which currently consists of the 28 EU member states and three European Free Trade Association (EFTA) member states – Iceland, Liechtenstein and Norway [202]. (Note that the GDPR is binding within the EEA.)

[12] Note that the EU framework for transnational data sharing is described by Lingjie Kong [107, p. 442] as "strict" and "regionally orientated".

[13] For instance, Svetlana Yakovleva [101, p. 487] highlights that: "*there is no international intergovernmental organization explicitly mandated to create unified international privacy and* data *protection standards."* Furthermore, any move towards a global framework is likely to be hampered, as the three principle "data realms" – China, Europe and the USA – have different approaches to data protection and privacy [204, p. 247]. For more information on global data protection laws see: [205], [206], [207], [37], and [203].

Source: Legal and Privacy Toolkit v1 [3, p. 21] – available at https://datapitch.eu/deliverables/

Furthermore, these legal rights and duties may differ between sectors and countries (including within the EU bloc).

2.3.1 Check legal agreements and licensing arrangements

An example of a common contractual constraint on data sharing and re-usage is where only named individuals and/or organisations (data users) party to a data sharing agreement are: (i) granted access to re-use a specific dataset; and (ii) prohibited from sharing these data with any third parties.[14]

Appropriate and effective rights management and clearance is imperative – you need to ensure that you have the authority to share data and/or rights to re-use data for your planned activity.[15] As part of this assessment, it is further essential that you check whether and, if so, the extent in which any legal agreements and licensing arrangements (that you plan to enter into) may constrain transnational, cross-sector data sharing. It is useful here to repeat the first point of the *"quick checklist to help confirm whether data sharing and reuse is lawful"* from v1 of the toolkit:

Quick checklist to help confirm whether data sharing and reuse is lawful – Point 1

"Make sure you check the legal agreement you have signed. In the case of Participating SMEs, check the nature of the limitations included which could reduce your planned re-usage of the data being shared with you and in particular, the allocation of intellectual property rights between the data provider and the data recipient. One key distinction to bear in mind in order to adequately allocate intellectual property rights is the distinction between the algorithm produced to process the shared data and the output of the processing activity through the means of the algorithm, which could be described as enriched or output data."

Source: Legal and Privacy Toolkit v1 [3, p. 58] – available at https://datapitch.eu/deliverables/

2.4 Enabling transnational, cross-sector data sharing through regulation

Legislators may decide to enable certain types of transnational, cross-sector data sharing through regulation. For instance, a chief aim of the Digital Single Market strategy is to facilitate the free-flow of data throughout EU member states where possible [10]. As part of this building a European data economy initiative, the Regulation on a Framework for the Free Flow of Non-Personal Data in the European Union entered into force in May 2019 [10], [11]. Furthermore, the GDPR intends to support the free-flow of personal data throughout EU member states where possible i.e. the title of the GDPR contains the phrase *"on the free movement of such data"*.

[14] Note that, in some cases, legal agreements and licensing arrangements can also enable (transnational, cross-sector) data sharing and re-usage, such as international open licences.

[15] For instance, Data Providers to the Data Pitch programme were required to complete a Data Provide Questionnaire (a copy is located in Appendix C to this report), including questions about the data that they planned to share as part of the accelerator. These questions also focused on potential transnational data sharing: *"[…] Please provide details of the following, including the relevant legal jurisdiction: [/] 1. Any IP rights applying to the data you would be happy to share under the project. If so, please specify what type they are and what they apply to. [/] 2. Ownership of such IP rights (including territory of right) [/] 3. Any registrations of such IP rights (including the territory of registration and the relevant market of registration) [/] 4. Any anticipated difficulty in granting licences (or sub-licences) to use such IP rights under the project (for example, on-going litigation)? 5. Whether any of the data you would be happy to share under the project was given in confidence. Under UK law, for example, a common law duty of confidence will arise where the information in question has a quality of confidence about it and was given in circumstances importing an obligation of confidence."* – Extract from Data Pitch: Data Provider Questionnaire.

2.5 Summary: regulation of transnational, cross-sector data sharing

2.5.1 Box A.1 – Some key points on the regulation of transnational, cross-sector data sharing (Part 1 of 2)

Box A.1

- **The legal framework for data sharing and (re)usage is complex and multi-layered** in that: (a) different types of law act concurrently in relation to a data sharing arrangement – e.g. from IPR and contractual rights and duties to data protection; and, (b) legal rights and duties can differ between sectors and countries (including within the EU bloc).
- **Legislators may decide to constrain certain types of transnational, cross-sector data sharing**, in particular to: (a) safeguard privacy and protect personally identifiable information; (b) meet regulatory objectives; (c) uphold national security; and, (d) support domestic innovation.
- Such regulatory constraints on transnational, cross-sector data sharing may: (i) take the form of **blanket bans**; (ii) target data flows that take place within **specific sectors**; and/or (iii) focus on data flows that occur as part of **processes and/or services**.
- **Other legal factors can constrain transnational, cross-sector data sharing**, such as intellectual property law (e.g. trade secrets) and contract law.
- **Legislators may also decide to support transnational, cross-sector data sharing through regulatory instruments,** such as the Regulation on a Framework for the Free Flow of Non-Personal Data in the European Union.

2.5.2 Box B.1 – Quick checklist on transnational, cross-sector data sharing in open innovation (Part 1 of 2)

Box B.1

- **Map your data flows** to help determine whether your planned data processing involves any transnational and/or cross-sector data sharing. See Appendix B to this report for more information.
- Assess what **regulatory constraints apply** to your planned data processing – such as applicable laws and jurisdictions. Consider any blanket bans and sector/process/service-specific constraints and requirements (e.g. sector-specific codes of best practice, industry standards, and other legislative requirements).
- Ensure you have the **authority to share data** and/or **rights to (re)use data** for your planned activity. Appropriate and effective rights management and clearance is imperative.
- Check how any **legal agreements and/or licensing arrangements** (that you plan to enter into) may constrain (transnational, cross-sector) data sharing and re-usage – and how this could potentially limit your planned data processing activity.

3. Part B – Key data protection aspects of transnational, cross-sector data sharing: an EU perspective

3.1 Brief overview

The General Data Protection Regulation (GDPR) [12] applies to (transnational, cross-sector) data sharing and (re)usage that involves information pertaining to an identified or identifiable natural person.[16] Given that data shared is *or* could likely become personal data within an open innovation environment, it is of paramount importance that those involved with open innovation remain

[16] For further guidance on data protection law, see Appendix B to this report, which explores the following topics: (i) defining personal data, (ii) high-risk personal data processing, (iii) appropriate safeguards and controls for personal data processing, and (iv) anonymisation assessment.

compliant with the GDPR.

The GDPR specifically-governs transnational, cross-sector data sharing in two main ways:

a. **Restricted transfers.** Article 44 of the GDPR restricts the transfer[17] of personal data to third countries and international organisations outside the European Economic Area (EEA)[18] unless controllers and processors[19] are able to comply with the provisions of Chapter V of the GDPR. It is therefore important to identify and ensure that all personal data transfers that take place to third countries and international organisations in the course of an open innovation programme (and elsewhere) are lawful.[20]

b. **Territorial scope.** The territorial scope of the GDPR is broad in that it applies to EEA and (in certain circumstances) non-EEA controllers and processors.[21] It is thus vital to determine whether the GDPR is applicable to non-EEA controllers and processors, and where relevant, how this applicability is likely to impact on the legal responsibilities and liabilities of EEA controllers and processors in the course of an open innovation programme (and elsewhere).

The principal purpose for data sharing within the course of open innovation is so that data can be analysed as part of an innovation challenge. Data sharing and re-usage is not only increasingly transnational and cross-sector in nature, but driven by data analytics, such as machine learning.[22] As (transnational, cross-sector) data sharing is a key driver of open innovation, it is further important to examine the following key data protection aspects of data-driven re-use:

c. **Automated individual decision-making, including profiling.** Given the proliferation of data-driven personal data processing activities over recent years, the GDPR "specifically addresses" [13, p. 5] two key types of (in some cases overlapping) automated personal data processing: (i) profiling; and (ii) automated decision-making. For instance, Article 22 of the GDPR places a "general prohibition" on solely automated decision-making, including profiling that has legal or similarly significant effects [13, p. 19]. It is therefore crucial to identify and ensure that all personal data processing that involves profiling and/or automated individual decision-making in the course of an open innovation programme (and elsewhere) are lawful.

Part B provides guidance on these three key data protection aspects of transnational, cross-sector data sharing (as outlined by points a.-c. above).

3.2 Restricted transfers

As aforementioned, it is important to identify and ensure that all personal data transfers that take place to third countries and international organisations in the course of an open innovation programme (and elsewhere) are lawful.

3.2.1 Personal data transits and transfers

Chapter V of the GDPR specifically applies to transfers of personal data to third countries or international organisations – i.e. transnational data sharing.[23] The European Data Protection Supervisor (EDPS) [1, p. 7] provides four examples of transfers of personal data:[24]

[17] Note that the GDPR does not refer to "transnational data sharing" but "data transfers". See [123] for further background information on the data sharing and (re)usage terminology used by the GDPR.

[18] The European Economic Area (EEA) comprises the contracting parties to the EEA Agreement, which currently consists of the 28 EU member states and three European Free Trade Association (EFTA) member states – Iceland, Liechtenstein and Norway [202]. Note that the GDPR is binding within the EEA.

[19] Refer to section 3.3.1 of this report for more information on controllers and processors.

[20] For instance, as part of an open innovation programme, an EEA data provider who plans to transfer personal data to a non-EEA SME needs to ensure such a transfer would be lawful.

[21] Paul de Hert and Michal Czerniawski [27, p. 242] state: *"Data controllers and processors outside the EU are brought under the GDPR via criteria that allow for identification of potential links between processing operations and the EU law. The overall logic here is based on targeting: if you target EU data subjects then the Regulation reaches out to you. This targeting or destination approach proposed in the GDPR is, however, not absolute and Article 3(2) tries to set clear limits to the extraterritorial scope of the General Regulation*."

[22] See [130] for more information on the increased role of machine learning as part of algorithmic decision-making.

[23] See [208] for a comparison between the regulation of transnational transfers of personal data under the repealed Data Protection Directive (DPD) and, its successor, the GDPR.

[24] In addition, W. Kuan Hon et al. [209, p. 263] provide three key ways in which transfers of personal data may take place: (i) *"transmission*

Four examples of transnational personal data transfers

"[S]ending of personal data by an EU institution or body (data controller) to a non-EU recipient by post or e-mail; "push" of data from an EU data controller's data base to a non-EU recipient; granting access of an EU data controller's data base ("pull") to a non-EU recipient; direct on-line collection of individual's data in the EU by a non-EU processor acting on behalf of an EU data controller; publication of personal data on the internet by an EU data controller."

Source: EDPS [1, p. 7]

Chapter V of the GDPR does not apply to **mere transits of personal data** that occur outside the EEA – i.e. a mere transit does not amount to transnational data sharing.[25] For illustration, the ICO [14] offers the following example of a mere transit of personal data:[26]

Example of a mere transit of personal data

"Personal data is transferred from a controller in France to a controller in Ireland (both countries in the EU) via a server in Australia. There is no intention that the personal data will be accessed or manipulated while it is in Australia. Therefore the transfer is only to Ireland."

Source: ICO [14]

This distinction is therefore important because *"there is no restriction on transferring data that is simply passing through a non-EEA country 'in transit' from one EEA member state to another"* [15]. For that reason, **it is vital that you are able to distinguish between transfers and transits of personal data when considering any planned data sharing activities.**

Data transfer assessment

A key issue is that the GDPR does not provide a legal definition of transfer of personal data.[27] Authoritative guidance from the EDPS [1, p. 7] outlines the elements that are required for a transfer to take place:

Elements required for a transnational personal data transfer

"[C]ontrollers should consider that this term [transfer] would normally imply the following elements: communication, disclosure or otherwise making available of personal data, conducted with the knowledge or intention of a sender subject to the Regulation that the recipient(s) will have access to it."

Source: EDPS [1, p. 7]

The definition provided by the EDPS therefore raises two distinct and required elements. First, the material element – the sender shares personal data (in some form) with one or more recipients outside the EEA i.e. through *"communication, disclosure or otherwise making available*". Second, the mental element – the sender has the *"knowledge or intention"* that the recipient will *"have access"* to these personal data. Conversely, the mere transit of personal data therefore takes place on the proviso that there is no: (i) intention to access or manipulate these data; or, (ii) actual access to or

of data to a 'third country' outside the EEA"; (ii) *"remote access to such data from a third country"*; and (iii) *"physical transportation to a third country of hardware storing such data"*.

[25] For instance, the ICO [210, p. 6] clearly distinguishes the concepts of transit and transfer: *"Transfer does not mean the same as mere transit"*. Elizabeth Brownsdon [112] further states: *"The DPA does not define "transfer" but it is not the same as the mere transit of data from country to country which is allowed under the DPA."* Note that Lexico [211] defines the verb "to transit" as *"Pass across or through (an area)"*.

[26] Note that the ICO [214, p. 2] provides the following variation of this example: *"[I]f data is transferred electronically from country A to country B, via a server in country C, there is no transfer to country C, only transit. For there to be a transfer data must have been sent to that country in order for something to happen to it there."* Helen Rowe [105] also re-iterates this definition of data transit: *"the fact that the electronic transfer of personal data may be routed through a third country on its way from the UK to another EEA country does not bring such transfer within the ambit of the Eighth Principle unless some substantive processing operation was being conducted upon the personal data in the third country in question."* Sana Khan [108, p. 11] also offers a comparative example of data transfer and data transit: *"For example, if a voice interview is carried out using telecommunication methods and is transmitted via a non-EEA country, this falls within the transit exception. However, if notes of the same interview are created and transferred into the database of a non-EEA country, this will constitute a transfer of data and as such only acceptable under the Directive if adequate safeguards have been put in place."*

[27] This is despite requests for greater clarity on the distinction between transfer and transit [9, p. 174], including calls for a legal definition of personal data transfer e.g. [212, p. 18] and [213, pp. 76-77].

manipulation of these data [14], [15].

It is further important to consider the potential risks when sharing data, as Christopher Kuner [9, p. 175] states:

Consider the risks of a planned data transit

"Mere transit of personal data should not be regarded as an exception to the definition of a data transfer, since data in transit are still flowing. Rather, the focus should be on the fact that mere transit does not pose a substantial risk of harm to the data, since they are simply being passed on to the next connection point. The fact that data are merely in transit should be an element to be taken into account when a risk analysis of transborder data flows is performed."

Source: Christopher Kuner [9, p. 175]

Exercise 1: Test your knowledge on the difference between transit and transfer

Instructions

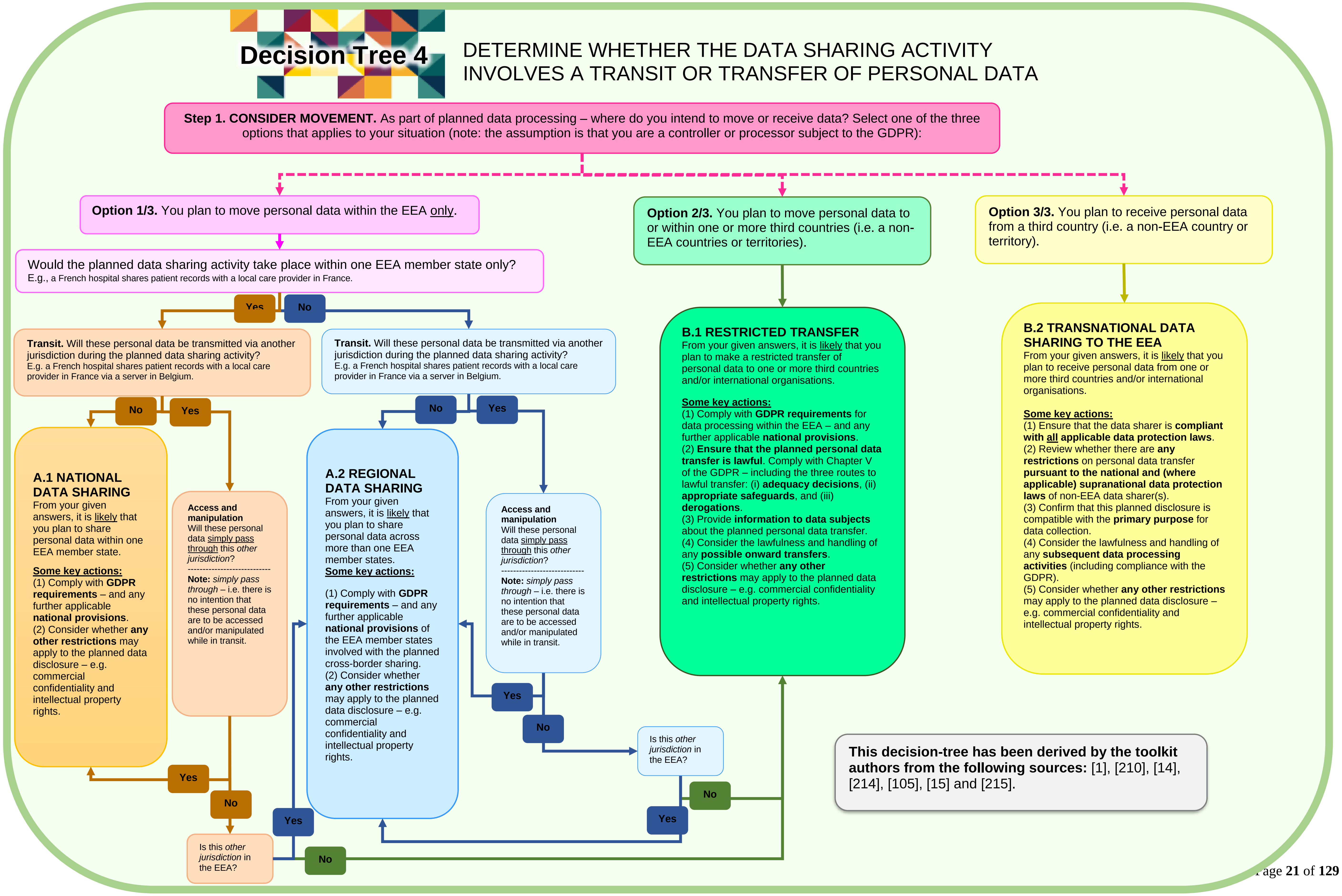
Use the following decision-tree (on the next page) to indicate whether a data sharing activity involves a transit or transfer of personal data.

You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

a) The head of HR based in the company headquarters in South Korea plans to send employee records to a HR manager based in Norway.

b) A French hospital plans to share pseudonymised patient records with a start-up based in Italy in order to investigate the effectiveness of certain treatments.

c) A German company plans to send customer profiles to a global marketing company based in Australia.



Decision Tree 4

DETERMINE WHETHER THE DATA SHARING ACTIVITY INVOLVES A TRANSIT OR TRANSFER OF PERSONAL DATA

Step 1. CONSIDER MOVEMENT. As part of planned data processing – where do you intend to move or receive data? Select one of the three options that applies to your situation (note: the assumption is that you are a controller or processor subject to the GDPR):

Option 1/3. You plan to move personal data within the EEA only.

Would the planned data sharing activity take place within one EEA member state only?
E.g., a French hospital shares patient records with a local care provider in France.

Transit. Will these personal data be transmitted via another jurisdiction during the planned data sharing activity?
E.g. a French hospital shares patient records with a local care provider in France via a server in Belgium.

A.1 NATIONAL DATA SHARING
From your given answers, it is likely that you plan to share personal data within one EEA member state.

Some key actions:
(1) Comply with **GDPR requirements** – and any further applicable **national provisions**.
(2) Consider whether **any other restrictions** may apply to the planned data disclosure – e.g. commercial confidentiality and intellectual property rights.

Access and manipulation
Will these personal data simply pass through this *other jurisdiction*?

Note: *simply pass through* – i.e. there is no intention that these personal data are to be accessed and/or manipulated while in transit.

Is this *other jurisdiction* in the EEA?

Transit. Will these personal data be transmitted via another jurisdiction during the planned data sharing activity?
E.g. a French hospital shares patient records with a local care provider in France via a server in Belgium.

A.2 REGIONAL DATA SHARING
From your given answers, it is likely that you plan to share personal data across more than one EEA member states.
Some key actions:

(1) Comply with **GDPR requirements** – and any further applicable **national provisions** of the EEA member states involved with the planned cross-border sharing.
(2) Consider whether **any other restrictions** may apply to the planned data disclosure – e.g. commercial confidentiality and intellectual property rights.

Access and manipulation
Will these personal data simply pass through this *other jurisdiction*?

Note: *simply pass through* – i.e. there is no intention that these personal data are to be accessed and/or manipulated while in transit.

Is this *other jurisdiction* in the EEA?

Option 2/3. You plan to move personal data to or within one or more third countries (i.e. a non-EEA countries or territories).

B.1 RESTRICTED TRANSFER
From your given answers, it is likely that you plan to make a restricted transfer of personal data to one or more third countries and/or international organisations.

Some key actions:
(1) Comply with **GDPR requirements** for data processing within the EEA – and any further applicable **national provisions**.
(2) **Ensure that the planned personal data transfer is lawful**. Comply with Chapter V of the GDPR – including the three routes to lawful transfer: (i) **adequacy decisions**, (ii) **appropriate safeguards**, and (iii) **derogations**.
(3) Provide **information to data subjects** about the planned personal data transfer.
(4) Consider the lawfulness and handling of any **possible onward transfers**.
(5) Consider whether **any other restrictions** may apply to the planned data disclosure – e.g. commercial confidentiality and intellectual property rights.

Option 3/3. You plan to receive personal data from a third country (i.e. a non-EEA country or territory).

B.2 TRANSNATIONAL DATA SHARING TO THE EEA
From your given answers, it is likely that you plan to receive personal data from one or more third countries and/or international organisations.

Some key actions:
(1) Ensure that the data sharer is **compliant with all applicable data protection laws**.
(2) Review whether there are **any restrictions** on personal data transfer **pursuant to the national and (where applicable) supranational data protection laws** of non-EEA data sharer(s).
(3) Confirm that this planned disclosure is compatible with the **primary purpose** for data collection.
(4) Consider the lawfulness and handling of any **subsequent data processing activities** (including compliance with the GDPR).
(5) Consider whether **any other restrictions** may apply to the planned data disclosure – e.g. commercial confidentiality and intellectual property rights.

This decision-tree has been derived by the toolkit authors from the following sources: [1], [210], [14], [214], [105], [15] and [215].

Figure 2 Legal Decision-Tree 4: Determine whether the data sharing activity involves a transit or transit of personal data

3.2.2 Transfers of personal data to third countries or international organisations

A lawful transnational transfer of personal data from the EEA can only take place if you adhere to the following two-step approach [16]:[28]

Two step approach to lawful transnational transfers of personal data

Step 1. – The transfer complies with *"requirements for data processing within the EU"* [16].

Step 2. – The transfer complies with *"the conditions laid down in Art. 44 et seq. GDPR in order to ensure an adequate level of data protection"* [16] – i.e. one of the three following routes laid down by the GDPR as follows.[29]

Source: Paul Voigt and Axel von dem Bussche [16]

Route A: Adequacy decisions

The planned transnational transfer of personal data is based on an adequacy decision (Article 45 and Recitals 104-107).[30]

An adequacy decision means that the EC has determined that (specified sectors within) a third country or an international organisation provide(s) an adequate level of protection for personal data comparable to that in the EEA [17].[31] As a result, an adequacy decision means that transnational transfers of personal data can take place to (specified sectors within) a third country or an international organisation without any "any specific authorisation" (Article 45(1)). Therefore, it is first vital to check whether an adequacy decision is in place that would cover a planned transnational data transfer – the EC website [18] maintains a list of adequacy decisions.[32]

Route B: Appropriate safeguards

The planned transnational transfer of personal data is subject to appropriate safeguards (Articles 46-47).

If the planned transnational transfer of personal data is not subject to an adequacy decision, the next step is to check whether any of the appropriate safeguards outlined by Article 46 apply to the planned transnational transfer. These appropriate safeguards include: binding corporate rules,[33] standard data protection clauses (adopted by the EC or a supervisory authority),[34] compliance with an

[28] Note that the European Data Protection Supervisor (EDPS) [1, pp. 29-31] provides a useful checklist to consider before making a transfer. See [216] and [116] for further background on the GDPR and transfers of personal data.

[29] For further background information see: [102].

[30] The European Commission (EC) defines the term *adequacy decision* [17] as follows: **"***An adequacy decision is a decision taken by the European Commission establishing that a third country provides a comparable level of protection of personal data to that in the European Union, through its domestic law or its international commitments. As a result, personal data can flow safely from the European Economic Area (EEA) (the 28 EU Member States as well Norway, Liechtenstein and Iceland) to that third country, without being subject to any further safeguards or authorisations."*

[31] Pursuant to Article 45(2) of the GDPR, the EC takes a number of elements into consideration *"[w]hen assessing the adequacy of the level of protection"*, such as: *"the rule of law"*, *"human rights"*, *"the existence and effective functioning of one or more independent supervisory authorities in the third country"*, and *"international commitments"*. Refer to Article 45(2) of the GDPR and [121] for more information. The EC continually monitors and periodically reviews each adequacy decision *"at least every four years"* (Article 45(3)).For further information, refer to: Article 45(3) and Recital 106 of the GDPR. Furthermore, the EC is able to *"repeal, amend or suspend"* an adequacy decision where it considers that the specific territory, specified sector(s) within a third country and international organisation no longer provides an adequate level of protection (Article 45(5)). For further information, refer to: Article 45(5) and Recital 107 of the GDPR.

[32] According to the European Commission website [18] (correct at the time of writing – November 2019), the following eleven countries/territories have been granted an adequacy decision: (i) Andorra, (ii) Argentina, (iii) Faroe Islands, (iv) Guernsey, (v) Israel, (vi) Isle of Mann, (vii) Japan, (viii) Jersey, (ix) New Zealand, (x) Switzerland, and (xi) Uruguay. Furthermore, the following two countries have been granted partial adequacy decisions: (xi) Canada – applies to commercial organisations; and, (xii) United States of America – limited to Privacy Shield framework [18]. Moreover, adequacy decisions are not absolute – e.g. Schrems I and II [19, p. 8]. [217] Moreover, talks are on-going with (xiii) South Korea to potentially grant an adequacy decision [18]. It also important to note that there is uncertainty over the status of the UK when it officially withdraws from the EU. For more information see: [218], [19, p. 8], [219], and [120].

[33] See [220] and [106] for more information on binding corporate rules.

[34] See [221] for more information on standard data protection clauses adopted by the European Commission (EC).

approved code of conduct and adherence to an approved certification mechanism.

Route C: Derogations

A derogation applies to the specific situation of the planned transnational transfer of personal data (Article 49 of the GDPR).

If the planned transnational transfer of personal data is not subject to an adequacy decision and no appropriate safeguards are applicable, the derogations outlined by Article 48 are a "last resort" [19, p. 9]. There are eight exceptions under which a transfer could be made: (i) data subject explicitly consents; (ii) performance of a contract; (iii) conclusion or performance of a contract; (iv) important reasons of public interest; (v) establishment, exercise or defence of legal claims; (vi) vital interests; (vii) public register; and, (viii) compelling legitimate interests.[35]

Need for "a global data protection programme"

In the words of Sahar Bhaimia [20, p. 25]: *"The default rule of thumb is that no personal data can be transferred to a third country (outside the EEA) unless it has been formally designated as "adequate' or the organisation uses one of an exhaustive list of mechanisms which provide necessary safeguards."* Therefore, if these three routes to lawful transnational transfer do <u>not</u> apply, personal data is not to be transferred in this instance.

As transnational data sharing is so ingrained in business activity, Eduardo Ustaran [21, p. 8] states: *"to ensure compliance, organisations are strongly advised to develop a viable global data protection compliance programme in line with the adequacy criteria devised by the European Commission, and commit to abiding by it through either a contractual mechanism or a set of [binding corporate rules] BCRs."*

3.2.3 Accessing or receiving personal data from third countries or international organisations

The GDPR is predominantly concerned with transfers of personal data to third countries and international organisations outside the EEA. In the words of Mira Burri [22, p. 70] states: *"the new generation of Internet controls seeks to keep information from going out of a country, rather than stopping it from entering the sovereign state space."*[36] **Given transnational data sharing is multi-directional, it is important that you consider the legal implications of accessing or receiving personal data from persons and/or organisations outside the EEA.**

Many non-EEA countries restrict the restrict transfers of personal data outside their borders – such as some South American countries [23] and Hong Kong [24].[37] It is therefore crucial that you are able to effectively identify and therefore comply with any applicable laws (including those of the EU and beyond) that concern the transnational data sharing activity in question.[38]

3.2.4 Cross-border personal data sharing within the EEA region

While the GDPR aims to better-harmonise data protection laws across the EEA, it does not provide for complete uniformity amongst member states [25, p. 11], [26]:[39]

[35] For more information, refer to: [164].

[36] Note the word "out" in this quote was italicised in the original article.

[37] For more information on the restrictions on transfers of personal data overseas from the USA, Malta, Norway, Turkey, China, Mexico, UK, Romania, Italy, the Netherlands, Brazil, Indonesia, India, Israel, Singapore, France, Germany, Switzerland and Ecuador, refer to [222].

[38] For instance, Outlaw.com [215] (Pinsent Masons) outlines four key questions a UK data controller should consider if they plan to receive data from overseas (i.e. outside the EEA): Q1: *"Is the UK entity only acting as a data processor on behalf of the overseas entity?"* Q2: *"If the UK entity does exercise control over the processing of the data, is the overseas entity complying with the laws of its own country?"* Q3: *"Are there any restrictions on transfer from that country?"* Q4: *"Similarly, will the UK data controller be using the data in a way compatible with the purposes for which it was originally collected?"*

[39] For instance, the divergences between member states on the implementation of the GDPR, such as criminal offences for re-identification under the UK Data Protection Act 2018.

The GDPR and national variations

"[…] Member States should be allowed to maintain or introduce national provisions to further specify the application of the rules of this Regulation […] Member States have several sector-specific laws in areas that need more specific provisions. This Regulation also provides a margin of manoeuvre for Member States to specify its rules, including for the processing of special categories of personal data ('sensitive data') […]"

Source: Recital 10 of the GDPR [12]

In consequence, **it is vital that you consider any national legal data protection variations[40] – this includes sector-specific legislation – that may exist alongside the GDPR before a cross-border transfer of personal data takes place within the EEA.[41]**

3.3 Territorial scope

The territorial scope of the GDPR is broad in that it is applicable to both EEA, and in specific circumstances, non-EEA controllers and processors. As aforementioned, while point-to-point transfers of data still take place, it is now common for data to be shared and (re-)used as part of "complex information chains" [2, p. 4] that involve multiple partners within distributed networks [9, p. 2]. In complex chains of data, it can be more difficult to distinguish controllers and processors [27, p. 220]. It is thus vital to determine whether the GDPR is applicable to non-EEA controllers and processors, and where relevant, how this applicability is likely to impact on the legal responsibilities and liabilities of EEA controllers and processors in the course of an open innovation programme (and elsewhere).

3.3.1 Controllers and processors

In order to assess whether the GDPR applies to a planned data processing activity, it is important that you first determine which entities are controllers and processors – i.e. the entities legally responsible and liable for personal data processing [28, p. 4].

The GDPR applies to data processing undertaken by two main actors: (1) controllers and (2) processors. **It is therefore essential that you understand the distinction between controllers and processors in order to assess the possible extent of your legal responsibilities and liabilities for the planned data processing in question.**

Legal definition of controller

The GDPR defines a controller as follows:

Legal definition of controller

"the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law"

Source: Article 4(7) of the GDPR [12]

A fundamental aspect of this legal definition is that the controller determines the purposes and means of the processing of personal data i.e. *the why* and *the how* of the planned data processing in question. Data owners are *"de facto"* controllers if they determine the purpose(s) (i.e. *"the why"* [29, p. 13], [30, p. 6]) for a specific (planned) data processing activity [29, pp. 14, 15].

The *means* of the planned data processing activity under consideration is also referred to as *"the*

[40] According to Amberhawk Training [223]: *"[p]rovisions that allow Member States law to modify the GDPR provision can be found in the following Articles: 4(7), 4(9), 6(2), 6(3)(b), 6(4), 8(1), 8(3), 9(2)(a), 9(2)(b), 9(2)(g), 9(2)(h), 9(2)(i), 9(2)(j), 9(3), 9(4), 10, 14(5)(b), 14(5)(c), 14(5)(d), 17(1)(e), 17(3)(b), 17(3)(d), 22(2)(b), 23(1)(e), 26(1), 28(3), 28(3)(a), 28(3)(g), 28(3)(h), 28(4), 29, 32(4), 35(10), 36(5), 37(4), 38(5), 49(1)(g), 49(4), 49(5), 53(1), 53(3), 54(1), 54(2), 58(1)(f), 58(2), 58(3), 58(4), 58(5), 59, 61(4)(b), 62(3), 80, 83(5)(d), 83(7), 83(8), 85, 86, 87, 88, 89, 90."*

[41] For instance, a cross-border personal data transfer occurs where a French Data Provider submits a pseudonymised dataset to a Spanish SME as part of an open innovation challenge.

how" [29, p. 13], [30, p. 6] and *"the manner in which data are processed"* [30, p. 6]. The *means* has a number of *"essential elements"* [29, p. 14] that only a controller can determine: (i) *"which data shall be processed?"*; (ii) *"for how long shall they be processed?"*; and, (iii) *"who shall have access to them?"* In addition, the ICO provides a further list of seven decisions that made only by a controller:

Seven decisions to be made only by a controller

"To determine whether you are a data controller you need to ascertain which organisation decides: [/] to collect the personal data in the first place and the legal basis for doing so; [/] which items of personal data to collect, ie the content of the data; [/] the purpose or purposes the data are to be used for; [/] which individuals to collect data about; [/] whether to disclose the data, and if so, who to; [/] whether subject access and other individuals' rights apply ie the application of exemptions; and [/] how long to retain the data or whether to make non-routine amendments to the data."

Source: ICO [30, pp. 6-7]

Controllers therefore determine the elements *"essential to the core of lawfulness of processing"* [29, p. 15] – i.e. the essential elements that concern how data are processed as part of a specific (planned) data processing activity [29, p. 14].

Three types of controllers

More than one controller is able to determine the purposes and means of a planned data processing activity [29, p. 18]. Furthermore, multiple controllers are able to process the same personal dataset for different purposes and means simultaneously [31]. Hence, another crucial feature of the legal definition of controller is that the controller jointly or alone determines the purposes and means of a specific (planned) data processing activity:

Joint controllers

"Where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers. They shall in a transparent manner determine their respective responsibilities for compliance with the obligations under this Regulation […]"

Source: Article 26(1) of the GDPR [12]

It is important to recognise *"in the context of joint control the participation of the parties to the joint determination may take different forms and does not need to be equally shared"* [29, p. 19]. Moreover, given that joint controllers can be held either jointly or severally liable for data breaches [32], it is crucial that data owners understand what type of role they are likely to have in any given data processing activity.

Figure 3 Table 1. An overview of three types of controllers

Table 1. An overview of three types of controllers

	Type of controller	Brief overview	Example
1	**Single controller**	A single controller determines the purposes and means of the processing in question alone. They will have no relationship of any kind with any other controllers for the specific data processing activity in question i.e. there is no collaboration [33].	A retail chain gathers and processes data from its customers in-house.
2	**Controllers in common**	A controller in common determines the purposes and means of the processing in question alone. However, they will work alongside other controllers often by sharing a pool of data that they process independently [31].	A travel agency and hotel exchange personal travel information, but process these data independently [33].

3	**Joint controller**	A joint controller determines the purposes and means of the processing in question with others.	Two universities jointly collaborate on a research project to gather and process personal data [33].

The GDPR defines processor as follows:

Legal definition of processor

"[A] natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller"

Source: Article 4(8) of the GDPR [12]

A fundamental aspect of this legal definition is that a processor processes data on behalf of the controller as *"separate legal entity with respect to the controller"* [29, p. 25]. Controllers are able to delegate elements which are not essential to the core of the lawfulness of processing – i.e. the technical and organisational questions that relate to the manner in which data are processed as part of a specific (planned) data processing activity, such as *"which hardware or software shall be used?"* [29, p. 14]. The ICO provides the following list of further examples:

Non-essential decisions that can be taken by processors

"Within the terms of the agreement with the data controller, and its contract, a data processor may decide: [/] what IT systems or other methods to use to collect personal data; [/] how to store the personal data; [/] the detail of the security surrounding the personal data; [/] the means used to transfer the personal data from one organisation to another; [/] the means used to retrieve personal data about certain individuals; [/] the method for ensuring a retention schedule is adhered to; and [/] the means used to delete or dispose of the data."

Source: ICO [30, p. 7]

Processors are therefore able to utilise their technical and organisational knowledge to decide how to process personal data on behalf of the controller short of determining any of the essential elements [30, p. 7]. In the words of Article 29 of the GDPR: *"The processor and any person acting under the authority of the controller or of the processor, who has access to personal data, shall not process those data except on instructions from the controller, unless required to do so by Union or Member State law."*

Exercise 2: Test your knowledge on controllers and processors

In order to assess which parties are controllers and/or processors in relation to a specific data processing activity, you must take into consideration the individual circumstances of this activity [34, p. 12]. In other words, an assessment must take place on *a "case-by-case basis"* [35, p. 3]. Therefore, the particular circumstances of each planned processing activity may give rise to a different assignment of roles. In the words of Bridget Treacy [36, p. 5]: *"Parties will need to analyse very carefully their respective data processing obligations in the knowledge that, in relation to a particular data set, they may be a controller for certain processing, and a […] processor for other processing."*

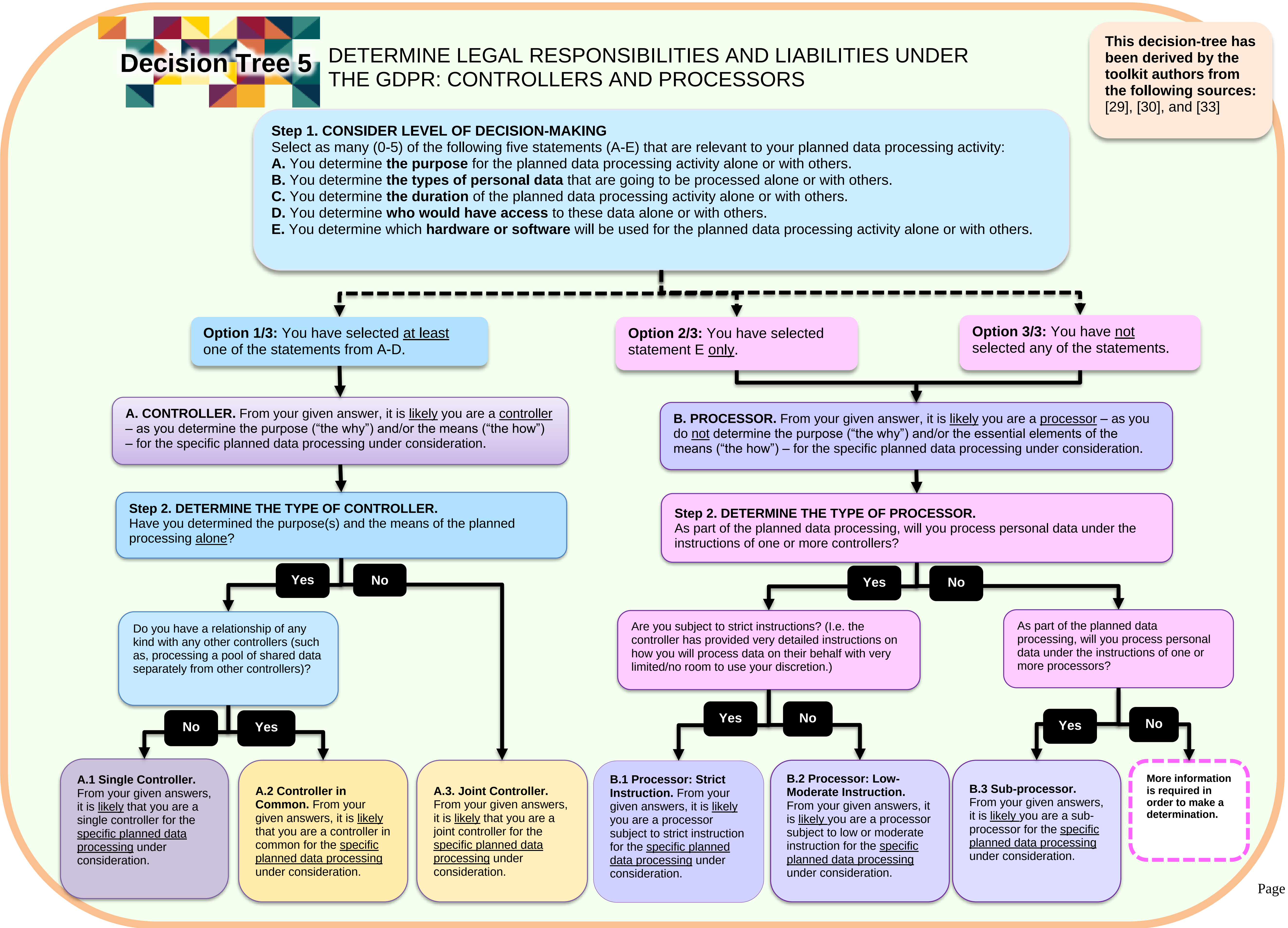
Instructions

Use the following decision-tree (on the next page) to indicate your potential legal responsibilities and liabilities for a specific data processing activity.

You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

(a) A large multi-national retail chain gathers customer insight data and shares these data with three external marketing companies for analysis. One of the external marketing companies further shares these data with a data analytics company.

(b) A hospital shares some patient data with a medical technology company in order to improve existing patient care.



Decision Tree 5 DETERMINE LEGAL RESPONSIBILITIES AND LIABILITIES UNDER THE GDPR: CONTROLLERS AND PROCESSORS

This decision-tree has been derived by the toolkit authors from the following sources: [29], [30], and [33]

Step 1. CONSIDER LEVEL OF DECISION-MAKING
Select as many (0-5) of the following five statements (A-E) that are relevant to your planned data processing activity:
A. You determine **the purpose** for the planned data processing activity alone or with others.
B. You determine **the types of personal data** that are going to be processed alone or with others.
C. You determine **the duration** of the planned data processing activity alone or with others.
D. You determine **who would have access** to these data alone or with others.
E. You determine which **hardware or software** will be used for the planned data processing activity alone or with others.

Option 1/3: You have selected at least one of the statements from A-D.

Option 2/3: You have selected statement E only.

Option 3/3: You have not selected any of the statements.

A. CONTROLLER. From your given answer, it is likely you are a controller – as you determine the purpose ("the why") and/or the means ("the how") – for the specific planned data processing under consideration.

Step 2. DETERMINE THE TYPE OF CONTROLLER.
Have you determined the purpose(s) and the means of the planned processing alone?

Yes

No

Do you have a relationship of any kind with any other controllers (such as, processing a pool of shared data separately from other controllers)?

No

Yes

A.1 Single Controller. From your given answers, it is likely that you are a single controller for the specific planned data processing under consideration.

A.2 Controller in Common. From your given answers, it is likely that you are a controller in common for the specific planned data processing under consideration.

A.3. Joint Controller. From your given answers, it is likely that you are a joint controller for the specific planned data processing under consideration.

B. PROCESSOR. From your given answer, it is likely you are a processor – as you do not determine the purpose ("the why") and/or the essential elements of the means ("the how") – for the specific planned data processing under consideration.

Step 2. DETERMINE THE TYPE OF PROCESSOR.
As part of the planned data processing, will you process personal data under the instructions of one or more controllers?

Yes

No

Are you subject to strict instructions? (I.e. the controller has provided very detailed instructions on how you will process data on their behalf with very limited/no room to use your discretion.)

Yes

No

As part of the planned data processing, will you process personal data under the instructions of one or more processors?

Yes

No

B.1 Processor: Strict Instruction. From your given answers, it is likely you are a processor subject to strict instruction for the specific planned data processing under consideration.

B.2 Processor: Low-Moderate Instruction. From your given answers, it is likely you are a processor subject to low or moderate instruction for the specific planned data processing under consideration.

B.3 Sub-processor. From your given answers, it is likely you are a sub-processor for the specific planned data processing under consideration.

More information is required in order to make a determination.

Figure 4 Legal Decision-Tree 5: Determine legal responsibilities and liabilities under the GDPR: controllers and processors

3.3.2 Three main ways the GDPR applies to the processing of personal data

The territorial scope of the GDPR is broad – as summarised by Clare Sullivan [37]*: "[t]he GDPR […] has extensive extraterritorial provisions that apply to processing of personal data outside the EU regardless of place of incorporation and geographical area of operation of the data controller/ processor."* The GDPR therefore can apply to data processing undertaken by both EEA and non-EEA bodies that involve personal data, which meet one of the three criteria for territorial scope under Article 3 of the GDPR:[42]

a) **The establishment criterion – Article 3(1) of the GDPR:** *"This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not."* The GDPR therefore applies to controllers and processors established[43] in the EEA – this includes non-EU companies that (i) exercise effective and real activity through (ii) stable arrangements – e.g. a branch, office, and/or single employee in the EEA [38].[44] It is important that controllers and processors established in the EEA fulfil their GDPR obligations for all personal data processed in the context of their activities – i.e. it is irrelevant whether these data relate to EEA or non-EEA data subjects [39].

b) **The targeting criterion – Article 3(2) of the GDPR:** *"This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to: (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or (b) the monitoring of their behaviour as far as their behaviour takes place within the Union."* The GDPR therefore applies to non-EEA companies (who are not established in the EEA pursuant to Article 3(1)) that target goods and services or monitor behaviour within the Union. It is important to note that if Article 3(2) applies, you must designate a representative.[45]

c) **"Processing in a place where member state law applies by virtue of public international law" – Article 3(3) of the GDPR**. It is also important to note this third (more limited) way in which the GDPR applies to the processing of personal data.

Relationships between EU and non-EU controllers and processors

A planned data processing activity may involve more than one party (e.g. a controller and a processor) – it is important to note that simply because the GDPR applies to one party does not mean it will automatically apply to all other parties involved in the planned data processing activity:

[42] For more information refer to EDPS guidance: On 16 November 2018, the European Data Protection Board adopted guidelines on the territorial scope of the GDPR. The following section aims to highlight some of the key points from this guidance – please refer to original guidelines for full information. Some comments on these guidelines: [127].

[43] While the legal text of GDPR does not offer a definition of "established in the Union", Recital 22 of the GDPR offers some further clarification: *"[…] Establishment implies the effective and real exercise of activity through stable arrangements. The legal form of such arrangements, whether through a branch or a subsidiary with a legal personality, is not the determining factor in that respect."* In many cases, it is obvious that the GDPR applies to a planned processing activity that involves personal data, such as those carried out by EEA bodies in the context of their activities (i.e. those entities incorporated and registered in the EEA); regardless of whether these data are processed within or outside the EEA [38].

[44] It is important to note that "established in the Union" is broadly-interpreted – i.e. the establishment criterion applies to entities based outside the Union where there is "*any real and effective activity — even a minimal one — exercised through stable arrangements"* [28, p. 5]. Furthermore, in the case of exclusive online services, the threshold for stable arrangements is "quite low" [28, p. 5] – e.g. the stable presence of a single employee/agent of the non-EU entity [28, p. 5]. However, the mere accessibility of a non-EU entity's website in the Union alone would not amount to an establishment in the Union [28, p. 5]. For more information see EDPB.

[45] Article 4(17) of the GDPR defines representative as: *"a natural or legal person established in the Union who, designated by the controller or processor in writing pursuant to Article 27, represents the controller or processor with regard to their respective obligations under this Regulation".*

Relationship between EU and non-EU controllers and processors

"The existence of a relationship between a controller and a processor does not necessarily trigger the application of the GDPR to both, should one of these two entities not be established in the Union."

Source: EDPB [28, p. 9]

Guidance given by the EDPB [28, p. 9] focuses on the following two scenarios:[46]

- **Scenario 1: "*Processing by a controller in the EU using a processor not subject to the GDPR"*** [28, p. 9]. The controller is subject to the GDPR controller obligations – and must ensure that the processor is compliant with the GDPR for the planned data processing activity [28, p. 9]. The non-EEA processor(s) therefore will be indirectly subject to the GDPR through a controller-processor contract (which must cover Article 28 obligations) [28, p. 9]. Chapter V of the GDPR on transfers of personal data to third countries or international organisations may apply.
- **Scenario 2: "*Processing in the context of the activities of an establishment of a processor in the Union"*** [28, p. 10]. The non-EU controller that instructs a processor in the Union will not be subject to the GDPR where:
  - (i) The activities of the controller does not fall under the scope of Article 3(1), (2) or (3) [28, p. 10];
  - (ii) *"The processing is carried out in the context of the controller's own activities"* [28, p. 10]; and
  - (iii) *"The processor is merely providing a processing service which is not "inextricably linked" to the activities of the controller"* [28, p. 10].

The processor established in the Union remains subject to the GDPR for this processing activity however – and comply with GDPR processor obligations.[47]

Exercise 3: Test your knowledge on the application of the GDPR

Instructions

Use the following decision-tree (on the next page) to indicate your potential legal responsibilities and liabilities for a specific data processing activity.

You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

(a) A large multi-national retail chain gathers customer insight data and shares these data with three external marketing companies for analysis in Sweden, Brazil and Mexico. One of the external marketing companies further shares these data with an Indian data analytics company.

(b) A Canadian hospital shares some patient data with a Portuguese medical technology company in order to improve existing patient care.

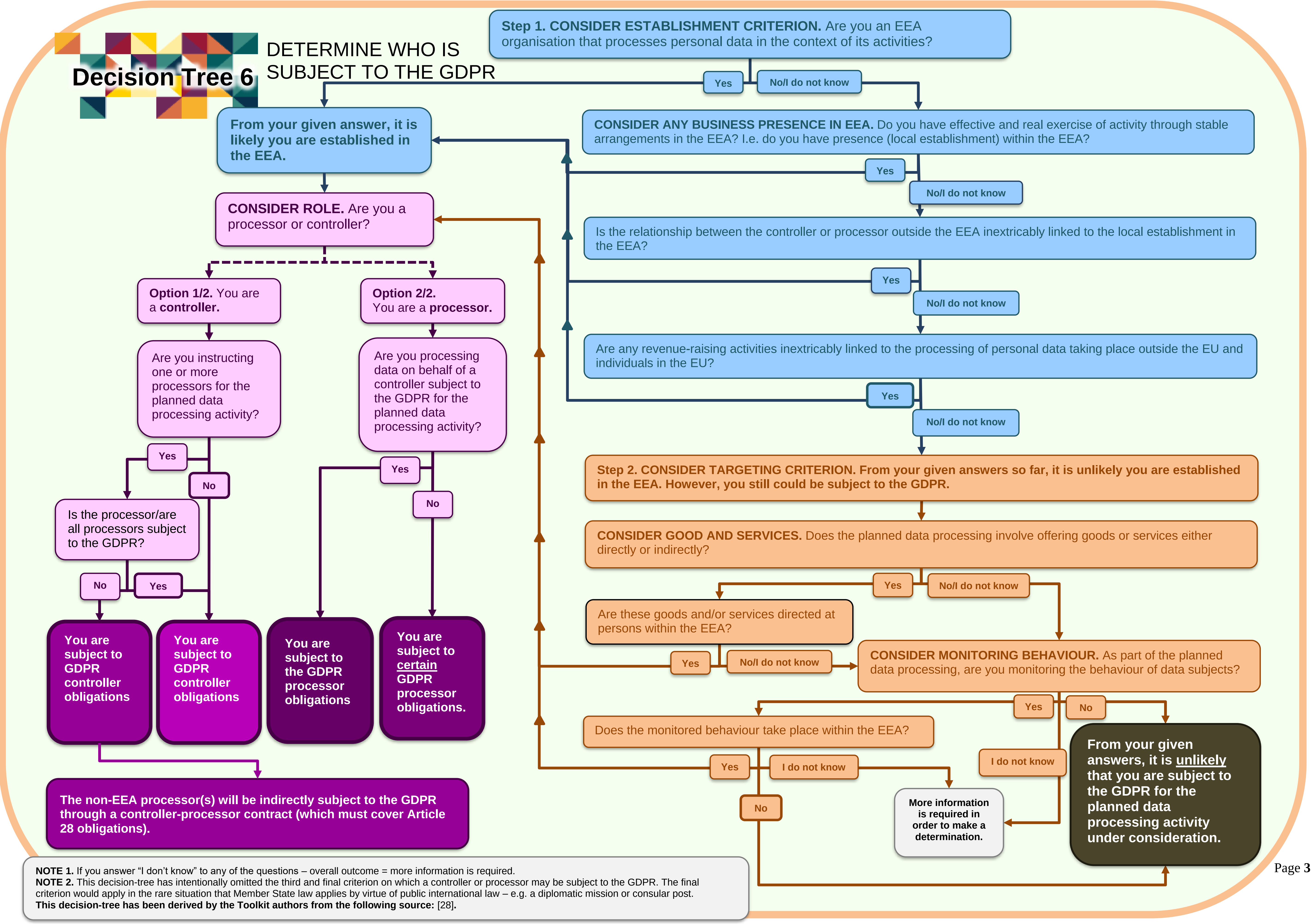
(c) The head of HR based in the company headquarters in South Korea plans to send employee records to a HR manager based in Norway.

(d) A French hospital plans to share pseudonymised patient records with a start-up based in Italy in order to investigate the effectiveness of certain treatments.

(e) A German company plans to send customer profiles to a global marketing company based in Australia.

[46] Note that this EDPB does not directly address that relationship between restrictions on transnational transfer and the extraterritorial provisions of the GDPR – for further information see [127], [224], and [225] .

[47] For further information, see [28, pp. 11-12].



Decision Tree 6 DETERMINE WHO IS SUBJECT TO THE GDPR

Step 1. CONSIDER ESTABLISHMENT CRITERION. Are you an EEA organisation that processes personal data in the context of its activities?

Yes

No/I do not know

From your given answer, it is likely you are established in the EEA.

CONSIDER ROLE. Are you a processor or controller?

Option 1/2. You are a **controller.**

Are you instructing one or more processors for the planned data processing activity?

Yes

No

Is the processor/are all processors subject to the GDPR?

No

Yes

You are subject to GDPR controller obligations

You are subject to GDPR controller obligations

The non-EEA processor(s) will be indirectly subject to the GDPR through a controller-processor contract (which must cover Article 28 obligations).

Option 2/2. You are a **processor.**

Are you processing data on behalf of a controller subject to the GDPR for the planned data processing activity?

Yes

No

You are subject to the GDPR processor obligations

You are subject to <u>certain</u> GDPR processor obligations.

CONSIDER ANY BUSINESS PRESENCE IN EEA. Do you have effective and real exercise of activity through stable arrangements in the EEA? I.e. do you have presence (local establishment) within the EEA?

Yes

No/I do not know

Is the relationship between the controller or processor outside the EEA inextricably linked to the local establishment in the EEA?

Yes

No/I do not know

Are any revenue-raising activities inextricably linked to the processing of personal data taking place outside the EU and individuals in the EU?

Yes

No/I do not know

Step 2. CONSIDER TARGETING CRITERION. From your given answers so far, it is unlikely you are established in the EEA. However, you still could be subject to the GDPR.

CONSIDER GOOD AND SERVICES. Does the planned data processing involve offering goods or services either directly or indirectly?

Yes

No/I do not know

Are these goods and/or services directed at persons within the EEA?

Yes

No/I do not know

CONSIDER MONITORING BEHAVIOUR. As part of the planned data processing, are you monitoring the behaviour of data subjects?

Yes

No

I do not know

Does the monitored behaviour take place within the EEA?

Yes

I do not know

No

More information is required in order to make a determination.

From your given answers, it is <u>unlikely</u> that you are subject to the GDPR for the planned data processing activity under consideration.

NOTE 1. If you answer "I don't know" to any of the questions – overall outcome = more information is required.
NOTE 2. This decision-tree has intentionally omitted the third and final criterion on which a controller or processor may be subject to the GDPR. The final criterion would apply in the rare situation that Member State law applies by virtue of public international law – e.g. a diplomatic mission or consular post.
This decision-tree has been derived by the Toolkit authors from the following source: [28].

Figure 5 Legal Decision-Tree 6: Determine who is subject to the GDPR

3.4 Obligation to provide privacy information to data subjects

3.4.1 Minimum standard for privacy information

The GDPR places an obligation on controllers to inform data subjects[48] about their processing activities. Articles 13, 14 and 15 of the GDPR sets out a minimum standard for this privacy information by outlining specific types of information, such as *"the identity and the contact details of the controller and, where applicable, of the controller's representative"* – Articles 13(1)(a) and 14(1)(a).[49] It is therefore that you are familiar with Articles 13, 14 and 15 – and all privacy information that you are legally required to provide to data subjects.[50]

3.4.2 Information about transnational personal data transfers

As part of this privacy information, controllers must provide information to data subjects on transnational transfers of personal data:

Obligation on controller to provide privacy information to data subjects on transfers of personal data

"[…] where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second subparagraph of Article 49(1), reference to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available […]"

Source: Articles 13(1)(f) and 14(1)(f) of the GDPR [12]

Furthermore, data subjects have a right to be informed about the appropriate safeguards used by a controller as part of a transnational transfer of personal data:

The right of data subjects to be informed about appropriate safeguards for transfers of personal data

"Where personal data are transferred to a third country or to an international organisation, the data subject shall have the right to be informed of the appropriate safeguards pursuant to Article 46 relating to the transfer."

Source: Article 15(2) of the GDPR [12]

3.5 Data-driven innovation: profiling and automated decision-making

Open innovation is becoming increasingly data-driven. For instance, as part of their open innovation activities Participating SMEs may:

- **Develop (big data) analytical techniques** – such as artificial intelligence systems, including machine learning – as part of their open innovation activities. E.g. where a Participating SME re-uses data in order to train a machine-learning algorithm.

[48] While Articles 13, 14 and 15 specifically-focus on informing data subjects, other key stakeholders will require information about your data processing activities. For instance, in terms of a machine learning initiative, various stakeholders – such as "AI engineers", "software developers and/or integrators", "business owners" and "end-users" – will require privacy information [138]. Furthermore, other transparency requirements may require controllers and processors to provide information to data protection supervisory authorities, auditors, and the courts. Privacy information must therefore suit the context and targeted audience.

[49] Refer to Articles 13, 14 and 15 of the GDPR for full information.

[50] Note that there are specific circumstances in which the obligation to provide privacy information may be limited – see Article 13(4) and 14(5) of the GDPR for full information.

- **Apply (big data) analytical techniques** to mine data for hidden value. E.g. where a Participating SME utilises data derived from a data mining process as part of a new analysis.

Attention must be given therefore to how the GDPR governs the utilisation of such (big data) analytical techniques,[51] especially its regulation of profiling and automated-decision making.

What is profiling?

The GDPR defines profiling as:[52]

Legal definition of profiling

"any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".

Source: Article 4(4) of the GDPR [12]

From this legal definition, data processing is very likely to constitute profiling where all three following three factors are present: (i) there is some form of automation (the planned processing activity is either partly or fully automated); (ii) personal data is involved;[53] and, (iii) an evaluation of personal aspects that relate to a natural person [13, pp. 6-7].[54] ICO [40] provides four keys ways in which an organisation may carry out profiling:[55]

Four key types of profiling

"You are carrying out profiling if you: [/] collect and analyse personal data on a large scale, using algorithms, AI or machine-learning; [/] identify associations to build links between different behaviours and attributes; [/] create profiles that you apply to individuals; or [/] predict individuals' behaviour based on their assigned profiles."

Source: ICO [40]

Profiling is used across a number of sectors for diverse purposes [41], [42].[56] For instance, in 2018, the European Agency for Fundamental Rights published a handbook – *"Preventing unlawful profiling today and in the future: a guide"* [43] – which specifically-focuses on the use of profiling within law enforcement and border management.

What is automated decision-making?

There is no explicit legal definition of automated decision-making given by the GDPR. Automated decisions can be made with or without profiling (hence the reference to *"including profiling"*) [13, p.

[51] In many cases, personal data processing that involves profiling and/or automated decision-making is invisible or incomprehensible to data subjects – i.e. "opaque" [13, p. 5]. Furthermore, there is the potential for such personal data processing to cause significant harm to individuals, such as by perpetuating *"existing stereotypes and social segregation"*, *"inaccurate predictions"* and *"unjustified discrimination"* [13, pp. 5-6].

[52] Fedelma Good et al. [109, p. 7] state: *"In simpler terms, profiling is the automated processing of personal data to analyse or to make predictions about individuals' behaviour and the type of person they are."*

[53] See Appendix B to this report for more information.

[54] The Article 29 Working Party [13, p. 7] maintain the evaluation of personal aspects involves *"some form of assessment or judgement about a person"*. It is important to note therefore *"a simple classification of individuals based on known characteristics such as their age, sex, and height does not necessarily lead to profiling"* [13, p. 7] where the purpose is not to evaluate the personal aspects that relate to a natural person [13, p. 7]. In some cases, it is clear that the planned data processing is profiling – i.e. where it *"draws inferences"* from related personal aspects connected to a natural person [40] (such as the application of machine learning to *"predict patients' health"* [40]). In other cases, it may be less clear – i.e. where the planned data processing "draws inferences" from seemingly unrelated personal aspects connected to a natural person [40] (such as, an insurance company analysing words and phrases related to (un)safe driving in social media posts to assign risk levels to individuals and therefore determine driver insurance premiums [40]).

[55] Frederike Kaltheuner and Elettra Bietti [132, pp. 4-6] outline four key purposes of profiling: (i) *"to infer or predict information"* (refer to [132, pp. 4-5]); (ii) *"to score, rank, evaluate and assess people"* (refer to [132, p. 5]); (iii) *"inform a decision about an individual"* (refer to [132, pp. 5-6]); and (iv) *"to make or inform a decision that personalises an individual's environment"* (refer to [132, p. 6]).

[56] In general, organisations may utilise profiling to: *"find something out about individuals' preferences; [/] predict their behaviour; and/or [/] make decisions about them"* [41]. Samantha Sayers and James Drury-Smith [125, p. 3] provide some further high-level examples of potential profiling activities: (i) *"Analysing an individual's credit history for the purposes of credit scoring"*; (ii) *"Creating a single customer view in order to track customer behaviour across products/services"*; (iii) *"Monitoring of employees"*; (iv) *"Conducting big data analytics"*; (v) "*Undertaking cross device tracking*"; and (vi) *"Conducting behavioural marketing"*.

8]. The ICO [44] define automated decision-making as *"the process of making a decision by automated means without any human involvement."* An automated decision usually involves: (i) factual data, (ii) digitally created profiles and/or (iii) inferred data [44].

In some cases, automated decision-making will not involve profiling.[57] For instance, where automated decisions are based on factual data, such as the use of an automated system to mark multiple choice exam papers[58] or an automated system to impose speeding fines solely based on factual data obtained by speed cameras[59]. Furthermore, it is important to note that a simple automated decision-making process can be developed into an automated decision-making process based on profiling.[60] For instance, an examination board aims to integrate this automated exam marking process within an "intelligent tutoring system" that is able to *"adapt to the needs of individual students by systematically and dynamically providing prompts, scaffolding, and feedback based on their ability to detect, track, and model key [self-regulated learning] SRL processes"* [45, p. 107]. In other cases, automated decision-making will involve profiling, such as an individual classification based on interests used to target products and services. COMPAS is as an existing example of an automated decision-making system for criminal sentencing in the US [46].

General prohibition on solely automated decision-making, including profiling

It is very important that you are aware of the general prohibition placed on solely automated decision-making, including profiling that has legal or similarly significant effects by the GDPR [13, p. 19]:

General prohibition on solely automated decision-making, including profiling

"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

Source: Article 22(1) of the GDPR [12]

This general prohibition applies where the following three criteria apply: [61]

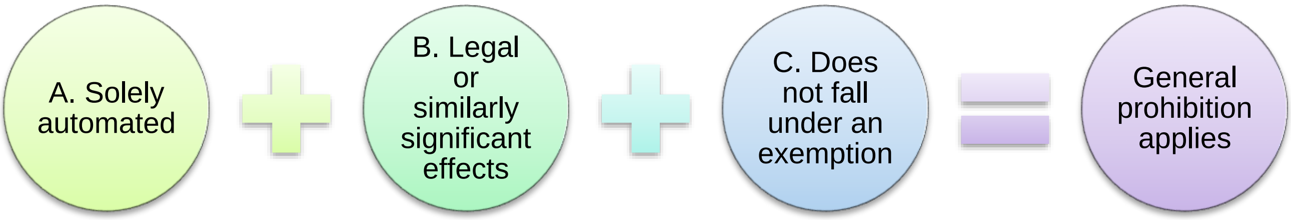

A. Solely automated
B. Legal or similarly significant effects
C. Does not fall under an exemption
General prohibition applies

Figure 6 Criteria for general prohibition on solely automated decision-making. Toolkit diagram based on analysis from the Article 29 Data Protection Working Party [16, pp. 19-24].

(a) The automated decision-making is solely automated

Automated decision-making, including profiling is solely automated where there is **<u>no</u> meaningful human involvement** [13, p. 20].[62] Controllers cannot circumvent the general prohibition *"by*

[57] . Paul B. de Laat [130] outlines the three main phases of automated decision making – "Phase 1: Data Collection"; "Phase 2: Modern Construction"; and, "Phase 3: Model Use".

[58] Note this example has been paraphrased, see [44] for more detail.

[59] Note this example has been paraphrased, see [13, p. 8] for more detail.

[60] The Article 29 Working Party also give an example about speed cameras.

[61] Note that the ICO provides a GDPR-compliance check-list for all automated decision-making, including profiling on its website, see [40].

[62] Note that the scope of the general prohibition is not without criticism. First, the applicability of Article 22 to solely automated decisions may be extremely limited in practice [49]. For instance, it seems to be uncommon for solely automated decision-making systems to make decisions with significant effects without any meaningful human involvement [49]. Rather, automated decision-making is often used as part of "decision support systems" that assist people to make decisions with significant effects, e.g. for "welfare benefit" [49]. Second, meaningful human involvement is no panacea for responsible and fair decision-making [132, pp. 13-14]. Third, the general prohibition overlooks issues with human decision-making, such as opacity, bias and prejudice [132, pp. 13-14]. Fourth, the GDPR aims to be technologically neutral – for instance, Recital 15 of the GDPR states: *"In order to prevent creating a serious risk of circumvention, the protection of natural persons should be technologically neutral and should not depend on the techniques used. The protection of natural persons should apply to the processing of personal data by automated means, as well as to manual processing, if the personal data are contained or are intended to be contained in a filing system. […]."* However, Article 22 appears to be out of step with this technologically

fabricating human involvement" [13, p. 21]. According to Article 29 Working Party [13, p. 21], in order for human involvement to be meaningful it must meet the following criteria:

(i) Human involvement must be more than a *"token gesture"* [13, p. 21].
(ii) The person(s) involved must have *"the authority and competence to change the decision"* [13, p. 21].
(iii) The person(s) involved must *"consider all the relevant data"* [13, p. 21].
(iv) The controller *"should identify and record the degree of any human involvement"* within their data protection impact assessment (DPIA) [13, p. 21].

(b) The solely automated decision-making produces legal or similarly significant effects

The general prohibition applies to decisions based on solely automated decision-making that have **"serious impactful effects"** [13, p. 21].[63] Some examples of significant legal effects, include *"the freedom to take legal action"* and "*denial of particular social benefit"* [13, p. 21].[64] Moreover, some examples of similarly significant effects include "automatic refusal of credit application or e-recruiting practices without any human intervention" (recital 71 of the GDPR.)[65]

(c) The solely automated decision-making produces legal or similarly significant effects is not exempt

The GDPR provides three exemptions to the general prohibition on solely automated decision-making produces legal or similarly significant effects:

Exemptions to the prohibition

"Paragraph 1 shall not apply if the decision: [/] (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; [/] (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or [/] (c) is based on the data subject's explicit consent."

Source: Article 22(2) of the GDPR

Other considerations: sensitive data and regional variations

It is important to note that any automated decision-making and/or profiling that involves sensitive data must meet the additional requirements under Article 9(2) and Article 6 of the GDPR [13, p. 22]. Furthermore, controllers need to comply with any regional variations between the EU members states interpretation of Article 22 [47].

neutral stance, as it exclusively applies to a very narrow category of processing: solely automated decision-making, including profiling with significant effects. It does not apply to every decision with significant effects for data subjects, such decision support systems and other types of human decision-making. The GDPR therefore appears to hold solely automated decision-making with significant effects to a higher standard than other forms of decision-making with human involvement that result in significant effects [226, p. 11]. Decision-making with human involvement is not without bias and inaccuracy. Nick Wallace and Daniel Castro [226, p. 24] therefore recommend that any right to review or explanation should be technologically neutral and depend on the *"nature and the seriousness of the decision in question, not simply whether the decision was made by a human or an algorithm"* – and further argue that this focus on automated decision-making may create a disincentive to use AI [226, p. 11].

[63] As a result, it would be extremely unlikely that *"recommendations for new television programmes based on an individual's previous viewing habits"*[40] would constitute a serious impactful effect and therefore fall under the scope of Article 22.

[64] The Article 29 Working Party [13, p. 21] offers the following examples of legal effects – the planned automated decision-making would affect: (i) the legal rights of individuals – e.g. *"the freedom to associate with others, vote in an election, or take legal action "* ; (ii) the legal status of individuals – e.g. *"entitlement to or denial of a particular social benefit granted by law, such as child or housing benefit; refused admission to a country or denial of citizenship"*; and (iii) rights under a contract – e.g. *"cancellation of a contract"*. Also see [40] for more information.

[65] The planned automated decision-making would *"significantly affect the circumstances, behaviours or choices of the individuals concerned"* [13, p. 21] – individually or in groups (such as, minority groups and vulnerable adults) – e.g. *"automatic refusal of a credit application or e-recruiting practices without any human intervention"* – Recital 21 of the GDPR**.** The planned automated decision-making would have *"a prolonged or permanent impact on the data subject"* [13, p. 21] – e.g. *"decisions that affect someone's access to health services"* [13, p. 22]. The planned automated decision-making would, at its most extreme, lead to the *"exclusion or discrimination of individuals"* [13, p. 21] – e.g. *"decisions that affect someone's access to education"* [13, p. 22].

Practical guidance on creating privacy information about profiling and automated decision-making

In many cases, personal data processing that involves profiling and/or automated decision-making is invisible or incomprehensible to data subjects – i.e. "opaque" [13, p. 5]. Furthermore, there is the potential for such personal data processing to cause significant harms to individuals, such as perpetuating *"existing stereotypes and social segregation"*, *"inaccurate predictions"* and *"unjustified discrimination"* [13, pp. 5-6].[66] **It is therefore important that you inform data subjects about your processing activities – in particular, where you use profiling and/or automated decision-making.**

Right to explanation debate

Much of the privacy information – required by Articles 13, 14 and 15 of the GDPR – appears to be largely straightforward (e.g. contact details and purposes for processing). However, one area that has caused some uncertainty is the provision of privacy information concerning profiling and automated decision-making to data subjects.

Obligation on controller to provide privacy information to data subjects on automated decision-making, including profiling

"the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject"

Source: Articles 13(2)(f), 14(2)(f) and 15(1)(h) of the GDPR [12]

The GDPR does not provide an exact formula for the construction of privacy information for automated decision-making, including profiling.[67] This ambiguity has led to considerable debate over how to meet the legal standard for meaningful information,[68] what constitutes a useful explanation in practice, and furthermore whether the GDPR gives rise to a legal right of explanation for data subjects.[69] It is therefore important to examine authoritative guidance on the minimum standard prescribed by law, such as that provided by Article 29 Working Party [13] and national data protection supervisory bodies such as ICO.[70]

Provide privacy information about automated decision-making and/or profiling

It is important that controllers *"explain clearly and simply to individuals how the profiling or automated*

[66] For further background information see, e.g.: [129] for an outline of six types of ethical concerns raised by algorithms – (i) "inclusive evidence", (ii) "inscrutable evidence", (iii) "misguided evidence", (iv) "unfair outcomes", (v) "transformative effects", and (vi) "traceability"; [119] for a CNIL report on the ethical matters raised by algorithms and artificial intelligence; [136] for exploration of algorithmic fairness and transparency; and [131] for four key barriers to algorithmic transparency.

[67] For instance, Cliff Kuang [142] states: *"The law [GDPR] was written to be powerful and broad and fails to define what constitutes a satisfying explanation or how exactly those explanations are to be reached. It represents a rare case in which a law has managed to leap into a future that academics and tech companies are just beginning to devote concentrated effort to understanding."*

[68] Note that Gianclaudio Malgieri and Giovanni Comandé [128, p. 257] maintain that legislators intentionally selected this term as *meaningful* is a polysemous word in English (i.e. it has a double meaning). The term *meaningful* therefore should be interpreted as information about automated decision-making, including profiling that is both: (i) *"complete"* – i.e. *"relevant, significant, important"*; and, (ii) *"comprehensible"* – i.e. *"intended to show meaning"* [128, p. 257].

[69] A core focus of the right to explanation debate is that the non-legally binding (yet informative) text of Recital 71 of the GDPR states that a data subject has *"the right […] to obtain an explanation of the decision reached after such an assessment"*. However, this reference to a right to explanation is absent from the legally binding text of Article 22(3) of the GDPR. For further information on the right to explanation debate, refer to, e.g.: (i) Bryce Goodman and Seth Flaxman [135] who ignite the wider debate over the scope and existence of a right to explanation; (ii) Sandra Wachter et al. [124] who argue there is no right to explanation under the GDPR rather a narrower right to be informed; (iii) Lillian Edwards and Michael Veale [50] who maintain that a right to explanation is unlikely to provide a "complete remedy" – could actually be distracting and create a "transparency fallacy"; (iv) Andrew D. Selbst & Julia Powles [54] who contend that the provision of meaningful information does gives rise to a right to explanation; and (v) Gianclaudio Malgieri and Giovanni Comandé [128] who argue a right to legibility of automated decision-making exists in the GDPR that exceeds a right to explanation. Also see, e.g.: [140], [141], [111], [226], [49], [57], [137], [139].

[70] The Article 29 Working Party guidance on meaningful information has come under some criticism – e.g. [49], [226, p. 10]. On one hand, this may be beneficial – as there is room for flexibility [54]. On the other hand, this lack of legal certainty may cause problems for companies – as its interpretation may lead to litigation [226, p. 9]. Veale & Edwards [49] contend that the Article 29 Working Party guidelines [13] seem to provide implicit support to the stance taken by Wachter et al. [124] – i.e. meaningful information is about "general oversight" rather than "an explanation of a specific decision". Michael Veale and Lillian Edwards [49] further state: *"little tangible help is given [by the Article 29 Working Party guidelines] in relation to whether that elusive right [of explanation] can be derived from art 22 or recital 71, and (perhaps unsurprisingly) no help is given at all on what kind of elements might go into such explanations."*

decision-making process works" [13, p. 16].[71] In the case of solely automated decision-making, including profiling that produces legal or similarly significant effects (falls under an Article 22 exemption), controllers are required to:

Mandatory privacy information about solely automated decision-making, including profiling that produces legal or similarly significant effects:

- *"tell the data subject that they [the controller] are engaging in this type of activity;*
- *provide meaningful information about the logic involved; and*
- *explain the significance and envisaged consequences of the processing."*

Source: Article 29 Working Party [13, p. 13]

It is important to note that the Article 29 Working Party [13, p. 25] also recommend that the abovementioned privacy information be provided to data subjects for all other types of profiling and automated decision-making.

(1) Tell data subjects you plan to use/are using profiling and/or automated decision-making

i. **Be upfront.** If you plan to use AI for solely automated decision making, including profiling that have legal or similarly significant effects, you must tell data subjects (at minimum): (a) you are undertaking this activity, (b) your purposes for this activity, (c) information about the logic involved, and (d) the envisaged consequences of the processing [48]. It is also *"good practice"* [13, p. 29] to provide this information for other types of processing that involve automated decision-making and/or profiling.

ii. **Draw attention to any involvement of sensitive data, preferences and/or characteristics.** You should tell data subjects whether you plan to process or derive sensitive data as part of your open innovation activities [13, p. 22].

iii. **Consider visual techniques and ways to communicate privacy information.** It is also important to consider the ways in which you provide privacy information – e.g. through "layered" notices, "dashboards", "just-in-time notices", "icons", and "mobile and smart device functionalities" [48]. Visual techniques can also be used – e.g. the Article 29 Working Party offer the example of providing an app to compare the level of insurance payments between fictional drivers to show how an individual could lower their insurance rates [13, p. 26].

(2) Provide meaningful information about the logic involved

iv. **Explain rationale and criteria.** The GDPR provides no explicit definition of logic involved. The Article 29 Working Party [13, p. 25] interprets the phrase logic involved, as information about: (i) the rationale for a decision,[72] and (ii) the criteria relied on to reach a decision, including their source and relevance.[73]

[71] Furthermore, Recital 58 of the GDPR [12] states: *"The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used. Such information could be provided in electronic form, for example, when addressed to the public, through a website. This is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising. Given that children merit specific protection, any information and communication, where processing is addressed to a child, should be in such a clear and plain language that the child can easily understand."* For examples of open access reports on algorithms and analytics models used by government see: [55] for a review on the types of algorithms and safeguards used by government departments in New Zealand; and, [158] for an assessment concerned with the use and quality assurance of business-critical analytic models used by the UK government.

[72] The OED Online [144] defines the term *rationale* as *"a set of reasons or a logical basis for a course of action or belief"*. For example, the rationale behind a credit scoring system used to assess and accept or reject an individual's loan application could be explained as: a "*process helps them make fair and responsible lending decisions"* [13, p. 25]. Furthermore, the rationale behind a credit scoring system used to assess and accept or reject an individual's credit terms for purchases could be explained as: a process "*to decide whether or not to offer credit terms for purchases […] The retailer explains that the buyer's past behaviour and account transaction history indicates the most appropriate payment mechanism for the individual and the retailer"* [143].

[73] These criteria may include the type of information collected and (re)used to create a profile and/or automated decision, including their source and relevance [143], [13, p. 26]. The Article 29 Working Party [13, p. 31] provides examples of required information: [/] *"the categories of data that have been or will be used in the profiling or decision-making process;* [/] *"why these categories are considered pertinent;* [/] *"how any profile used in the automated decision-making process is built, including any statistics used in the analysis;* [/] *"why this profile is relevant to the automated decision-making process; and [/] "how it is used for a decision concerning the data subject."* For example, the criteria relied on by company employing a credit scoring system used to assess and accept or reject an individual's loan application could be explained as follows: (i) *"the information provided by the data subject on the application form"*; (ii) *"information about*

v. **No need for complex mathematical explanations.** Controllers need to provide clear, meaningful information to data subjects rather than complex mathematical explanations [13, p. 31]. However, it is important to note that controllers must provide meaningful information about the logic involved even where the data processing in question is extremely complex [13, p. 25] (e.g. neural networks).[74]

vi. **Where possible use subject-centric explanations.** According to Veale and Edwards [49], the Article 29 Working Party's interpretation of "logic involved" appears to be one of "general oversight" rather than "an explanation of a particular decision" – e.g. model-centric explanations (MCEs) that provide *"[…] broad information about an [machine learning] ML model which is not decision or input-data specific"* – e.g. "setup information" and "training data" [50, pp. 55-56]. Yet, in some cases, MCEs are too limited in scope to provide enough meaningful information about an individual decision in practice, you should therefore consider the use of subject-centric-explanations. SCEs are *"[…] built on and around the basis of an input record. They can only be provided in reference to a given query – which could be real or fictitious or exploratory"* [50, p. 55].[75] There are four key types of SCEs:[76]

Four key types of subject-centric explanations (SCEs)

- *"Sensitivity-based subject-centric explanations: what changes in my input data would have made my decision turn out otherwise? […]*
- *Case-based subject-centric explanations: which data records used to train this model are most similar to mine? […]*
- *Demographic-based subject-centric explanations: what are the characteristics of individuals who received similar treatment to me? […]*
- *Performance-based subject-centric explanations: how confident are you of my outcome? Are individuals similar to me classified erroneously more or less often than average? […]"*

Source: Lilian Edwards and Michael Veale [50, p. 58]

(3) Explain the significance and envisaged consequences of the processing

The GDPR provides no explicit definition of *envisaged consequence*. The Article 29 Working Party [13, p. 26] interprets the phrase *envisage consequences*, as information about:

vii. **Any intended or future processing** [13, p. 26]**.** For instance, an automated decision making process may involve on-going monitoring, such as an automated decision making process for motor insurance premiums that monitors driver behaviour in order to determine the insurance payment levels [13, p. 26].

viii. **The ways in which automated decision-making may affect the data subject** [13, p. 26]**.** It is important that you give *"real, tangible examples of the type of possible effects"* [13, p.

previous account conduct, including any payment arrears"; and (iii) *"official public records information such as fraud record information and insolvency records"* [13, p. 26]. Whilst not explicitly mentioned by the Article 29 Working Party, Michael Veale and Lilian Edwards [49] contend that information about training data (such as their collection and cleansing) would also be important information to provide to data subjects.

[74] Again, see Recital 58 of the GDPR.

[75] For instance, Finale Doshi-Velez et al. [140, p. 3] further maintain that *"the content an explanation should contain, we offer the following: an explanation should permit an observer to determine the extent to which a particular input was determinative or influential on the output".* At least the observer should know at least one of the following: (i) "the main factors in a decision"; (ii) if "changing a certain factor" would have influenced the decision"; and (iii) if "two similar-looking cases" get similar or different results [140, p. 3].

[76] Note that counterfactual explanations also fall under the scope of SCEs, and provide another useful way in which data subjects can obtain meaningful information relating to individual decisions that result from automated decision-making systems [133, p. 844]. Sandra Wachter et al. [133, p. 848] define a counterfactual explanation as: *"Score p was returned because variables V had values (v1, v2 , . . .) associated with them. If V instead had values (v1', v2', . . .), and all other variables had re-mained constant, score p' would have been returned."* Duncan Sinclair [134, p. 509] further states: "*The term "counterfactual", despite technical connotations, simply describes the tendency to think about how things might turn out differently "if only", and to imagine "what if?" It finds voice in philosophy, psychology, history, economics and law".* A counterfactual explanation has three main purposes, to enable the data subject to: (i) understand what led to the resulting individual decision [133, p. 863]; (ii) contest an "adverse" or "undesired" individual decision resulting from an automated decision-making system [133, p. 872]; and (iii) understand what changes need to be made in order to produce a desired outcome [133, p. 878].

26]. For instance, where an automated decision making process for motor insurance premiums that monitors driver behaviour determines the level of insurance payments, a possible effect would be higher insurance payments for dangerous drivers [13, p. 26].

(4) Surpass legal minimum standards: data ethics

Where the law stipulates the minimum standards required for data sharing, management and re-usage, data ethics[77] sets out the expected behaviours that are essential to drive forward best practice for open innovation. Compliance with data ethics is therefore complementary to the rest of the toolkit, and embedded into the Data Pitch open innovation programme from the outset. Requirements for ethical conduct[78] are enshrined in both: an ethical statement that Participating SMEs were required to sign in advance of taking part, and a declaration of honour requiring self-disclosure of matters such as past professional misconduct, and a guarantee of future compliance with ethical and legal principles under the Data Pitch programme. Copies of these documents are available in **Appendix C** to this report.

During the course of the Data Pitch programme, the field of data ethics continues to develop. For instance, the OECD Artificial Intelligence Principles that have been adopted by OECD member countries (in May 2019), non-member countries, and the G20 (in June 2019) [51]. It is important to highlight that algorithmic explanations is only one component of a wider-move towards enriched algorithmic accountability. For instance, the Principles for Accountable Algorithms and a Social Impact Statement for Algorithms [52] focus on: (i) “responsibility”, (ii) “explainability”; (iii) “accuracy”, (iv) “auditability”, and (v) “fairness”.

The Ethics Statement used by Data Pitch programme has been adapted (see table below) to include an additional principle that specifically-focuses on the development and/or application of (big data) analytical techniques as part of open innovation activities, especially the use of AI systems (including machine learning) for automated decision-making and profiling – and is derived from [51] and [53].

When taking part in an open innovation programme, Participating SMEs shall:

Figure 7 Table 2. Essential data ethics principles for SMEs participating in open innovation programmes

Table 2. Essential data ethics principles for SMEs participating in open innovation programmes*	
1.	Act in good faith.
2.	Respect human rights.
3.	Ensure research quality and integrity.
4.	Able to show that their findings are independent and non-discriminatory to any groups of individuals.
5.	Not misrepresent credentials.
6.	Demonstrate authenticity and validity of authorship.
7.	Respect confidential information.
8.	Secure any confidential information provided to prevent its misuse or unauthorised access.
9.	Only share confidential information where necessary and only where the prior informed consent of anyone potentially affected by the disclosure of such information has been received.
10.	Respect the privacy of any people identified from the findings of the open innovation programme as far as possible.
11.	Avoid any conduct that may cause anyone harm, and seek relevant individuals’ informed consent for any activities that might affect them directly.

[77] Floridi and Taddeo [100] define “data ethics” as: *“a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values).”*

16 For more background information on data ethics see, e.g. [201] and [200].

[78] The Data Pitch Ethical Statement defines ethical conduct as: *“acting consistently in a way that is ethical and fair and encouraging others to do likewise”*.

12. Determine the applicable laws that apply to their activities under the open innovation programme and plan your activities in accordance with such laws as early as possible.

13. Not collect or otherwise process any personal or sensitive data not essential for their open innovation programme activities.

14. Be fully transparent to the open innovation programme consortium about the purpose, methods and intended possible uses of their open innovation programme activities, and what risks, if any, are involved.

15. Seek advice promptly from the open innovation programme consortium where they believe ethical and/or legal risks may be raised by their activities.

16. Ensure that special attention is given to the development and/or application of (big data) analytical techniques as part of their open innovation activities, especially the use of AI systems (including machine learning) for automated decision making and profiling. The development and/or application of the (big data) analytical techniques should:

   a. Comply with the prior fifteen data ethics principles.
   b. Have legitimate benefits for business and wider society.
   c. Ensure meaningful human involvement.
   d. Be trustworthy-by-design.
   e. Provide "responsible disclosure" [51] to ensure that data subjects understand and can therefore challenge any outcomes.
   f. Be continually reviewed throughout the open innovation programme to identify and mitigate any risks.

(5) Review privacy information

Before you release any privacy information to data subjects, it is vital that you assess whether such information: (i) includes at least the information prescribed by Articles 13, 14 and 15 of the GDPR, (ii) will be easily understood and accessed by data subjects. It is further important that such information is kept up-to-date.

In terms of privacy information specifically about profiling and automated decision-making – it is important that you ensure such information is "meaningful". While there is no explicit definition of the term *meaningful* provided by the GDPR, Andrew D. Selbst and Julia Powles [54] contend that information is meaningful where:

- A person without *"particular technical expertise"* understands it.[79] [54]
- The information serves a "functional" purpose – in particular, where this has "instrumental value" such as the facilitation of a data subject's right to contest a solely automated decision that produces legal or similar effects (Article 22(3)).[80] [54]
- The information meets *"a minimum threshold of functionality"* – e.g. the controller therefore needs to provide enough information to enable a data subject to exercise their rights under the GDPR and human rights law.[81] [54]
- The information is tailored to the data processing in question – i.e. there should not be a strict rules or methodology for creating this meaningful information, as this can stifle the uses of more complex forms of data-driven processing (e.g. neural networks).[82] [54]

[79] The need to provide clear information is re-iterated by authoritative guidance provided by ICO and the Article 29 Working Party.

[80] For instance, Andrew Burt [126] maintains that meaningful information should provide a data subject with enough information to decide to opt-out: *"Taken together, Articles 21 and 22 suggest that the right to understand "meaningful information" about and the "significance" of, automated processing is related to an individual's ability to opt out of such processing. In plain English, the text suggests that a data subject is entitled to enough information about the automated system that she or he could make an informed decision to opt out."* Furthermore, according to Ajay Chander and Ramya Srinivasan [138], explanations may imbue various cognitive values, including: **(i) "trust"** i.e. *"the cognitive value of the explanation is to engender trust in the user"* – e.g. the explanation accounts for "personal values" such as "privacy" and "safety" [138]; **(ii) "troubleshooting"** and **(iii) "design"** i.e. the cognitive value of the explanation is to elucidate any issues pertaining to the "functional aspects of an AI model" – e.g. "accuracy, speed and robustness" – to aid with troubleshooting and/or its design [138]; and **(d) "education"** & **(e) "action"** – i.e. the cognitive value of the explanation is to help the *user "understand the AI's recommendation and aid them in analysis"* in order to *"help them take an appropriate action"* [138].

[81] Note: the obligation to provide meaningful information should not be considered in isolation from your other transparency obligations – such as DPIAs and data protection by design.

[82] In other words, different data-driven processing activities may need to provide different suited to their unique context and purpose –

(6) Keep up-to-date with (forthcoming) authoritative guidance and practice

It is further useful to examine examples of meaningful information to gain further insight. For instance, in 2018, the New Zealand Government: Stats NZ published its open access Algorithmic Assessment Report [55]. The purpose of this report is to the review types of algorithms used by government departments in New Zealand and their safeguards.[83] This report was written for the layperson when providing information about an algorithm it normally uses three categories: (i) the challenge; (ii) the solution; and (iii) the outcome. Furthermore, algorithmic explainability is the subject of other disciplines, e.g. Explainable AI [56], and areas of law, such as the French Digital Republic Act 2016:

Meaningful information – Existing Example 2: Loi pour une République numérique 2016-1321 (French Digital Republic Act 2016: Decree of March 2017(R311-3-1-2))

Lilian Edwards and Michael Veale [57] provide an English translation of the decree in March 2017 (R311-3-1-2):

"The new law provides that in the case of "a decision taken on the basis of an algorithmic treatment" (author translation), the rules that define that treatment and its "principal characteristics" must be communicated upon request. Further details were added by decree in March 2017 (R311-3-1-2) elaborating that the administration shall provide information about:
1. the degree and the mode of contribution of the algorithmic processing to the decision making;
2. the data processed and its source;
3. the treatment parameters and, where appropriate, their weighting, applied to the situation of the person concerned; and
4. the operations carried out by the treatment."

Source: Lilian Edwards and Michael Veale [57, p. 48]

It is crucial to note that further authoritative guidance on algorithmic explanations is forthcoming. For instance, the ICO and the Alan Turing Institute have produced an interim report as part of Project ExplAIN [58]. The purpose of Project ExplAIn is to develop practical guidance to help with organisations to explain artificial intelligence (AI) decisions to data subjects [58, p. 3]. It is therefore important that you keep up-to-date with any upcoming developments in authoritative guidance. The ICO intends to publish a consultation on its AI Auditing Framework by the end of 2019.

Exercise 4: Test your knowledge on automated decision-making and profiling

Instructions

Use the following decision-tree (on the next page) to test your knowledge on automated decision-making and profiling.

You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

(a) A retailer generates partially synthetic customer profiles for use within its customer insights analysis.

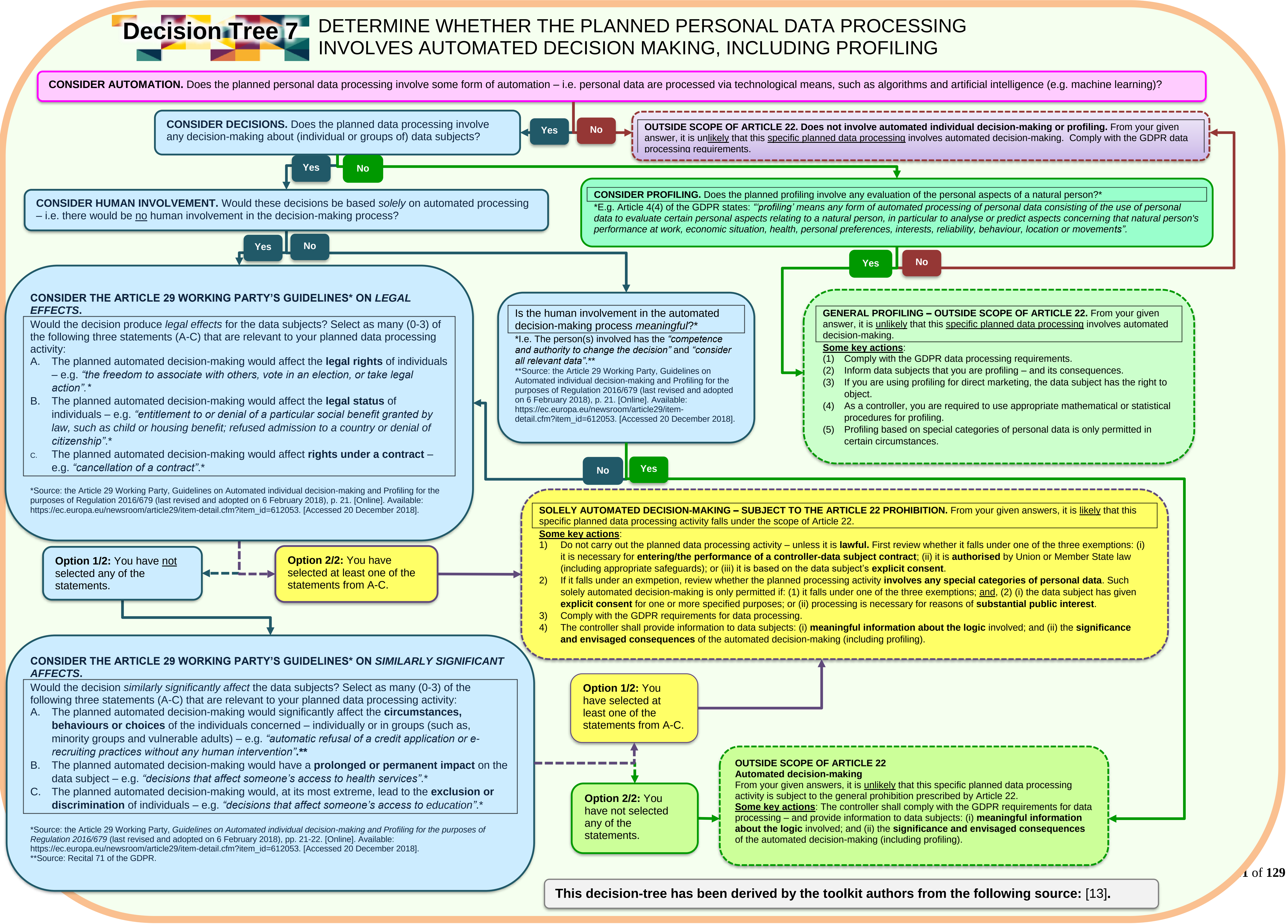
(b) A start-up uses pseudonymous data to train a machine learning algorithm.

(c) A further education college uses an automated system to mark multiple-choice exam papers.

(d) A recruitment agency rates the employability of its applicants through professional profiles and CV information. The recruitment agency uses these ratings to assign relevant job adverts and employers automatically.

there is no one-size fits all approach.

[83] Note this report does not provide an exhaustive review of all algorithms utilised by the New Zealand government departments and agencies.



Decision Tree 7 DETERMINE WHETHER THE PLANNED PERSONAL DATA PROCESSING INVOLVES AUTOMATED DECISION MAKING, INCLUDING PROFILING

CONSIDER AUTOMATION. Does the planned personal data processing involve some form of automation – i.e. personal data are processed via technological means, such as algorithms and artificial intelligence (e.g. machine learning)?

OUTSIDE SCOPE OF ARTICLE 22. Does not involve automated individual decision-making or profiling. From your given answer, it is un<u>likely</u> that this <u>specific planned data processing</u> involves automated decision-making. Comply with the GDPR data processing requirements.

CONSIDER DECISIONS. Does the planned data processing involve any decision-making about (individual or groups of) data subjects?

CONSIDER PROFILING. Does the planned profiling involve any evaluation of the personal aspects of a natural person?*

*E.g. Article 4(4) of the GDPR states: *"'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".*

GENERAL PROFILING – OUTSIDE SCOPE OF ARTICLE 22. From your given answer, it is <u>unlikely</u> that this <u>specific planned data processing</u> involves automated decision-making.

<u>**Some key actions**</u>:

(1) Comply with the GDPR data processing requirements.
(2) Inform data subjects that you are profiling – and its consequences.
(3) If you are using profiling for direct marketing, the data subject has the right to object.
(4) As a controller, you are required to use appropriate mathematical or statistical procedures for profiling.
(5) Profiling based on special categories of personal data is only permitted in certain circumstances.

CONSIDER HUMAN INVOLVEMENT. Would these decisions be based *solely* on automated processing – i.e. there would be <u>no</u> human involvement in the decision-making process?

Is the human involvement in the automated decision-making process *meaningful*?*

*I.e. The person(s) involved has the *"competence and authority to change the decision"* and *"consider all relevant data"*.**

**Source: the Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (last revised and adopted on 6 February 2018), p. 21. [Online]. Available: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053. [Accessed 20 December 2018].

CONSIDER THE ARTICLE 29 WORKING PARTY'S GUIDELINES* ON *LEGAL EFFECTS*.

Would the decision produce *legal effects* for the data subjects? Select as many (0-3) of the following three statements (A-C) that are relevant to your planned data processing activity:

A. The planned automated decision-making would affect the **legal rights** of individuals – e.g. *"the freedom to associate with others, vote in an election, or take legal action"*.*
B. The planned automated decision-making would affect the **legal status** of individuals – e.g. *"entitlement to or denial of a particular social benefit granted by law, such as child or housing benefit; refused admission to a country or denial of citizenship"*.*
C. The planned automated decision-making would affect **rights under a contract** – e.g. *"cancellation of a contract"*.*

*Source: the Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (last revised and adopted on 6 February 2018), p. 21. [Online]. Available: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053. [Accessed 20 December 2018].

Option 1/2: You have <u>not</u> selected any of the statements.

Option 2/2: You have selected at least one of the statements from A-C.

SOLELY AUTOMATED DECISION-MAKING – SUBJECT TO THE ARTICLE 22 PROHIBITION. From your given answers, it is <u>likely</u> that this specific planned data processing activity falls under the scope of Article 22.

<u>**Some key actions**</u>:

1) Do not carry out the planned data processing activity – unless it is **lawful.** First review whether it falls under one of the three exemptions: (i) it is necessary for **entering/the performance of a controller-data subject contract**; (ii) it is **authorised** by Union or Member State law (including appropriate safeguards); or (iii) it is based on the data subject's **explicit consent**.
2) If it falls under an exmpetion, review whether the planned processing activity **involves any special categories of personal data**. Such solely automated decision-making is only permitted if: (1) it falls under one of the three exemptions; <u>and</u>, (2) (i) the data subject has given **explicit consent** for one or more specified purposes; or (ii) processing is necessary for reasons of **substantial public interest**.
3) Comply with the GDPR requirements for data processing.
4) The controller shall provide information to data subjects: (i) **meaningful information about the logic** involved; and (ii) the **significance and envisaged consequences** of the automated decision-making (including profiling).

CONSIDER THE ARTICLE 29 WORKING PARTY'S GUIDELINES* ON *SIMILARLY SIGNIFICANT AFFECTS*.

Would the decision *similarly significantly affect* the data subjects? Select as many (0-3) of the following three statements (A-C) that are relevant to your planned data processing activity:

A. The planned automated decision-making would significantly affect the **circumstances, behaviours or choices** of the individuals concerned – individually or in groups (such as, minority groups and vulnerable adults) – e.g. *"automatic refusal of a credit application or e-recruiting practices without any human intervention"*.**
B. The planned automated decision-making would have a **prolonged or permanent impact** on the data subject – e.g. *"decisions that affect someone's access to health services"*.*
C. The planned automated decision-making would, at its most extreme, lead to the **exclusion or discrimination** of individuals – e.g. *"decisions that affect someone's access to education"*.*

*Source: the Article 29 Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (last revised and adopted on 6 February 2018), pp. 21-22. [Online]. Available: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053. [Accessed 20 December 2018].
**Source: Recital 71 of the GDPR.

Option 1/2: You have selected at least one of the statements from A-C.

Option 2/2: You have not selected any of the statements.

OUTSIDE SCOPE OF ARTICLE 22
Automated decision-making
From your given answers, it is <u>unlikely</u> that this specific planned data processing activity is subject to the general prohibition prescribed by Article 22.
<u>**Some key actions**</u>: The controller shall comply with the GDPR requirements for data processing – and provide information to data subjects: (i) **meaningful information about the logic** involved; and (ii) the **significance and envisaged consequences** of the automated decision-making (including profiling).

This decision-tree has been derived by the toolkit authors from the following source: [13]**.**

Figure 8 Legal Decision-Tree 7: Determine whether the planned data processing involves automated decision-making, including profiling

3.6 Summary – key data protection aspects of transnational, cross-sector data sharing: from an EU perspective

3.6.1 Box A.2 – Some key points on the regulation of transnational, cross-sector data sharing (Part 2 of 2)

Box A.2

- **The GDPR restricts the transnational transfer of personal data** to third countries and international organisations outside the European Economic Area (EEA) – unless controllers and processors are able to comply with (Step 1.) EU data processing requirements and (Step 2.) one of the three following routes in order to guarantee an adequate level of data protection: (i) adequacy decisions, (ii) appropriate safeguards, or (iii) derogations.
- **“Mere transits” of personal data should be distinguished from transfers of personal data** i.e. personal data simply passing through a non-EEA country is not restricted by the GDPR.
- **The territorial scope of the GDPR is broad in that it applies to EEA and (in certain circumstances) non-EEA controllers and processors** i.e. the entities legally responsible and liable for personal data processing.
- **Many other jurisdictions outside the EEA also restrict transnational transfers of personal data.**
- **In terms of cross-border person data sharing within the EEA, the GDPR does not provide for complete legal uniformity across member states** i.e. several national data protection variations exist within the EEA.
- **Controllers and processors have different levels of legal responsibilities and liabilities under the GDPR.**
- **The GDPR places an obligation on controllers to inform data subjects about their processing activities**, including information about any transnational transfers of personal data and automated decision-making, including profiling.
- **The GDPR places a general prohibition on solely automated decision-making that has legal or similarly significant effects (unless an exemption applies).**
- **The GDPR mandates that controllers tell data subjects they are undertaking activities that involve solely automated decision-making, including profiling that produces legal or similarly significant effects. Controllers should further provide meaningful information about the logic involved and the significance and envisaged consequences of the profiling.**
- **Mandatory privacy information about solely automated decision-making, including profiling that produces legal or similarly significant effects.**

3.6.2 Box B.2 – Quick checklist on transnational, cross-sector data sharing in open innovation (Part 2 of 2)

Box B.2

- Identify and ensure that **all personal data transfers that take place to third countries and international organisations** in the course of an open innovation programme (and elsewhere) **are lawful**.
- Identify and ensure that **all cross-border data sharing within the EEA** in the course of an open innovation programme (and elsewhere) **are lawful –** i.e. comply with any applicable national data protection variations and/or pertinent sectoral laws.

- Identify and ensure that **all other transnational data sharing activities are lawful.**
- **Determine which parties are the controllers and processors for a particular data processing activity** in order to assess the possible extent of your legal responsibilities and liabilities for the planned data processing in question**.**
- Make sure you are **familiar with the minimum legal standards for privacy information** outlined by Articles 13, 14 and 15 of the GDPR.
- Identify and ensure that **all profiling and automated decision-making activities** in the course of an open innovation programme (and elsewhere) **are lawful**.
- **Do not engage in any activities that involve solely automated decision-making, including profiling that produces legal or similarly significant effects** unless a legal exemption applies.
- **Open innovation activities that involve automated decision making and/or profiling should be explained to the Consortium, data subjects and where applicable the Data Provider involved**. At minimum, you should provide information about: (i) purpose and legal basis for your activities; including, the use or derivation of sensitive data, characteristics and/or preferences; (ii) the meaningful logic involved, including its rationale; and, (iii) the significance and envisaged consequences of the profiling, including any intended future processing and ways in which automated decision-making may affect the data subject. Note – there is no need for complex mathematical explanations, and subject-centric explanations should be used where possible.
- **Adhere to data ethics principles** specified by the open innovation programme and from authoritative sources.
- **Inform data subjects about your processing activities**, including any transnational transfers of personal data.
- **Keep up-to-date with authoritative guidance on transnational, cross-sector data sharing, including automated-decision making and profiling.**

4. Part C – The development of training tools: update

The first version of a prototype e-learning tool on data protection and the basics of mapping data flows ("the prototype e-learning tool") was created as part of version 2 of the toolkit. See Appendix C to this report for more background information about the prototype e-learning tool.

The second version of the prototype e-learning tool incorporates the following main updates:

- The addition of further descriptive content from the D3.5 report.
- The incorporation of four further legal decision-trees from this (D3.9) report.
- The inclusion of a privacy notice.
- Plans for the tool to be hosted by the University of Southampton.

5. Conclusions

5.1 Key points

This report (D3.9) is the third and final toolkit update that extends the legal guidance provided in the first and second versions of the toolkit. Given open innovation has the potential to transcend both sectors and national borders (within the EU and beyond), this report ultimately aims to show that transnational, cross-border assessment is a crucial part of any planned data processing activity. It addresses the D3.9 objective outlined by the Grant Agreement by providing (i) an overview of the legal aspects of transnational, cross-sector data sharing (explored in Part A) with (ii) particular focus on its privacy and data protection aspects from an EU perspective (outlined in Part B), including the development or application of (big data) analytics techniques as part of open innovation programmes.

5.1.1 Outline of regulation of transnational, cross-sector data sharing

Some key points to consider when you are sharing and re-using data in the course of an open innovation programme (and elsewhere):

General legal aspects

Box A.1

- **The legal framework for data sharing and (re)usage is complex and multi-layered** in that: (a) different types of law act concurrently in relation to a data sharing arrangement – e.g. from IPR and contractual rights and duties to data protection; and, (b) legal rights and duties can differ between sectors and countries (including within the EU bloc).
- **Legislators may decide to constrain certain types of transnational, cross-sector data sharing**, in particular to: (a) safeguard privacy and protect personally identifiable information; (b) meet regulatory objectives; (c) uphold national security; and, (d) support domestic innovation.
- Such regulatory constraints on transnational, cross-sector data sharing may: (i) take the form of **blanket bans**; (ii) target data flows that take place within **specific sectors**; and/or (iii) focus on data flows that occur as part of **processes and/or services**.
- **Other legal factors can constrain transnational, cross-sector data sharing**, such as intellectual property law (e.g. trade secrets) and contract law.
- **Legislators may also decide to support transnational, cross-sector data sharing through regulatory instruments,** such as the Regulation on a Framework for the Free Flow of Non-Personal Data in the European Union.

Key privacy and data protection aspects

Box A.2

- **The GDPR restricts the transnational transfer of personal data** to third countries and international organisations outside the European Economic Area (EEA) – unless controllers and processors are able to comply with (Step 1.) EU data processing requirements and (Step 2.) one of the three following routes in order to guarantee an adequate level of data protection: (i) adequacy decisions, (ii) appropriate safeguards, or (iii) derogations.
- **"Mere transits" of personal data should be distinguished from transfers of personal data** i.e. personal data simply passing through a non-EEA country is not restricted by the GDPR.
- **The territorial scope of the GDPR is broad in that it applies to EEA and (in certain circumstances) non-EEA controllers and processors** i.e. the entities legally responsible and liable for personal data processing.
- **Many other jurisdictions outside the EEA also restrict transnational transfers of personal data.**
- **In terms of cross-border person data sharing within the EEA, the GDPR does not provide for complete legal uniformity across member states** i.e. several national data protection variations exist within the EEA.
- **The GDPR places an obligation on controllers to inform data subjects about their processing activities**, including information about any transnational transfers of personal data and automated decision-making, including profiling.
- **Controllers and processors have different levels of legal responsibilities and liabilities under the GDPR.**
- **The GDPR places a general prohibition on solely automated decision-making that has legal or similarly significant effects.**

5.1.2 Quick checklist on transnational, cross-sector data sharing in open innovation

General legal aspects

Box B.1

- **Map your data flows** to help determine whether your planned data processing involves any transnational and/or cross-sector data sharing. See Appendix B to this report for more information.
- Assess what **regulatory constraints apply** to your planned data processing – such as applicable laws and jurisdictions. Consider any blanket bans and sector/process/service-specific constraints and requirements (e.g. sector-specific codes of best practice, industry standards, and other legislative requirements).
- Ensure you have the **authority to share data** and/or **rights to re-use data** for your planned activity. Appropriate and effective rights management and clearance is imperative.
- Check how any **legal agreements and/or licensing arrangements** (that you plan to enter into) may constrain (transnational, cross-sector) data sharing and re-usage – and how this could potentially limit your planned data processing activity.

Key privacy and data protection aspects

Box B.2

- Identify and ensure that **all personal data transfers that take place to third countries and international organisations** in the course of an open innovation programme (and elsewhere) **are lawful**.
- Identify and ensure that **all cross-border data sharing within the EEA** in the course of an open innovation programme (and elsewhere) **are lawful –** i.e. comply with any applicable national data protection variations and/or pertinent sectoral laws.
- Identify and ensure that **all other transnational data sharing activities are lawful**.
- **Determine which parties are the controllers and processors for a particular data processing activity** in order to assess the possible extent of your legal responsibilities and liabilities for the planned data processing in question**.**
- Make sure you are **familiar with the minimum legal standards for privacy information** outlined by Articles 13, 14 and 15 of the GDPR.
- Identify and ensure that **all profiling and automated decision-making activities** in the course of an open innovation programme (and elsewhere) **are lawful**.
- **Do not engage in any activities that involve solely automated decision-making, including profiling that produces legal or similarly significant effects** unless a legal exemption applies.
- **Open innovation activities that involve automated decision making and/or profiling should be explained to the Consortium, data subjects and where applicable the Data Provider involved**. At minimum, you should provide information about: (i) purpose and legal basis for your activities; including, the use or derivation of sensitive data, characteristics and/or preferences; (ii) the meaningful logic involved, including its rationale; and, (iii) the significance and envisaged consequences of the profiling, including any intended future processing and ways in which automated decision-making may affect the data subject. Note – there is no need for complex mathematical explanations, and subject-centric explanations should be used where possible.
- **Adhere to data ethics principles** specified by the open innovation programme and from authoritative sources.

- **Inform data subjects about your processing activities**, including any transnational transfers of personal data.
- **Keep up-to-date with authoritative guidance on transnational, cross-sector data sharing, including automated-decision making and profiling.**

5.1.3 Suggested further reading on transnational, cross-sector data sharing

- Article 29 Data Protection Working Party, “Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01),” 6 February 2018. [Online]. Available: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053. [Accessed 11 December 2019]
- C. Kuner, Transborder Data Flows and Data Privacy Law, Oxford: Oxford University Press, 2013.
- European Data Protection Board (EDPB), “Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) - Version for public consultation,” 16 November 2018 (adopted). [Online]. Available: https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_3_2018_territorial_scope_en.pdf. [Accessed 11 December 2019]
- European Data Protection Supervisor (EDPS), “The transfer of personal data to third countries and international organisations by EU institutions and bodies (Position Paper),” 14 July 2014. [Online]. Available: https://edps.europa.eu/sites/edp/files/publication/14-07-14_transfer_third_countries_en.pdf. [Accessed 11 December 2019]
- F. Casalini and J. López González, “Trade and Cross-Border Data Flows,” Organisation for Economic Co-operation and Development (OECD) Trade Policy Papers, No. 220, OECD Publishing, Paris., 23 January 2019. [Online]. Available: https://doi.org/10.1787/b2023a47-en. [Accessed 11 December 2019]
- Information Commissioner's Office (ICO), “For organisations - Guide to the General Data Protection Regulation (GDPR): International transfers,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/. [Accessed 11 December 2019]
- Information Commissioner's Office (ICO), “Data Sharing Code of Practice,” May 2011. [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf. [Accessed 11 December 2019] (note that this guidance is currently being updated)
- Information Commissioner's Office (ICO) and The Alan Turing Institute, “Project ExplAIn: Interim Report,” 2019. [Online]. Available: https://ico.org.uk/media/about-the-ico/documents/2615039/project-explain-20190603.pdf. [Accessed 11 December 2019]
- N. Corey, “Cross-Border Data Flows: Where Are the Barriers, and What Do They Cost?,” Information Technology & Innovation Foundation, 1 May 2017. [Online]. Available: https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost. [Accessed 11 December 2019]
- European Union Agency for Fundamental Rights, European Court of Human Rights, and Council of Europe, “Handbook on European data protection law,” June 2014. [Online]. In particular, Chapter 7. Available: http://fra.europa.eu/en/publication/2014/handbook-european-data-protection-law. [Accessed 11 December 2019]

6. References

[1] European Data Protection Supervisor (EDPS), "The transfer of personal data to third countries and international organisations by EU institutions and bodies (Position Paper)," 14 July 2014. [Online]. Available: https://edps.europa.eu/sites/edp/files/publication/14-07-14_transfer_third_countries_en.pdf. [Accessed 12 December 2019].

[2] Information Commissioner's Office (ICO), "Data Sharing Code of Practice," May 2011. [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf. [Accessed 12 December 2019].

[3] S. Stalla-Bourdillon and A. Knight, "D3.1 – Legal and Privacy Toolkit v1.0," Data Pitch H2020-ICT-2016-1; Project Number: 732506 , 30 June 2017. [Online]. Available: http://www.datapitch.eu/wp-content/uploads/2017/06/PUBLIC-LEGAL-AND-PRIVACY-TOOLKIT-VERSION-1.0-DELIVERABLE-8.1-FINAL-30-JUNE-2017.pdf. [Accessed 22 February 2018].

[4] S. Stalla-Bourdillon and L. Carmichael, "D3.5 - Legal and Privacy Toolkit v2," Contributor: Pei Zhang. Data Pitch H2020-ICT-2016-1; Project Number: 732506. , 30 June 2018. [Online]. Available: https://datapitch.eu/datapitch-d3-5-legal-and-privacy-toolkit-v2/. [Accessed 28 August 2018].

[5] F. Casalini and J. López González, "Trade and Cross-Border Data Flows," Organisation for Economic Co-operation and Development (OECD) Trade Policy Papers, No. 220, OECD Publishing, Paris., 23 January 2019. [Online]. Available: https://doi.org/10.1787/b2023a47-en. [Accessed 12 December 2019].

[6] N. Corey, "Cross-Border Data Flows: Where Are the Barriers, and What Do They Cost?," Information Technology & Innovation Foundation (ITIF), 1 May 2017. [Online]. Available: https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost. [Accessed 12 December 2019].

[7] J. L. Koffel, "GDPR adequacy decisions vs GATS: how may the EU's privacy and digital trade commitments be conciliated within a GDPR adequacy decision on cross-border personal data flows?," *International Trade Law & Regulation,* vol. 24, no. 3, pp. 122-143, 2018.

[8] J. Manyika, S. Lund, J. Bughin, J. Woetzel, K. Stamenov and D. Dhingra , "Digital globalization: The new era of global flows," March 2016. [Online]. Available: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/digital-globalization-the-new-era-of-global-flows. [Accessed 12 December 2019].

[9] C. Kuner, Transborder Data Flows and Data Privacy Law, Oxford: Oxford University Press, 2013.

[10] European Commission, "Building a European data economy," [Online]. Available: https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy. [Accessed 12 December 2019].

[11] "REGULATION (EU) 2018/1807 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 November 2018 on a framework for the free flow of non-personal data in the European Union," [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1807. [Accessed 12 December 2019].

[12] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, "General Data Protection Regulation (GDPR)," Official Journal of the European Union, 27 April 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN. [Accessed 12 December 2019].

[13] Article 29 Data Protection Working Party, "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)," 6 February 2018. [Online]. Available: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053. [Accessed 12 December 2019].

[14] Information Commisioner's Office (ICO), "For organisations - Guide to the General Data Protection Regulation (GDPR): International transfers," [Online]. Available: https://ico.org.uk/for-

organisations/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/. [Accessed 12 December 2019].

[15] BPE, “Transferring Data outside the UK,” [Online]. Available: https://www.bpe.co.uk/services/need/data-protection-the-gdpr/brilliantly-simple-guide-to-the-gdpr/transferring-data-outside-the-uk/. [Accessed 18 October 2018].

[16] P. Voigt and A. von dem Bussche, “Material Requirements,” in *The EU General Data Protection Regulation (GDPR),* Springer, Cham, 2017, pp. 87-140.

[17] European Commission (EC), “European Commission - Fact Sheet: Questions & Answers on the Japan adequacy decision,” 17 July 2018. [Online]. Available: http://europa.eu/rapid/press-release_MEMO-18-4503_en.htm. [Accessed 12 December 2019].

[18] European Commission (EC), “Adequacy of the protection of personal data in non-EU countries: How the EU determines if a non-EU country has an adequate level of data protection,” [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en. [Accessed 17 November 2019].

[19] H. McCarthy, “IDTs - transparency, adequacy and GDPR derogations,” *Privacy & Data Protection,* vol. 18, no. 2, pp. 8-9, 2017.

[20] S. Bhaimia, “Legislative Comment - The General Data Protection Regulation: the next generation of EU data protection,” *Legal Information Management,* vol. 18, no. 1, pp. 21-28, 2018.

[21] E. Ustaran, “GDPR series: international data transfers 2.0,” *Privacy & Data Protection,* vol. 17, no. 5, pp. 6-8, 2017.

[22] M. Burri, “The Governance of Data and Data Flows in Trade Agreements: The Pitfalls of Legal Adaptation,” *University of California Davis Law Review,* no. 1, pp. 65-132, 2017.

[23] A. B. Nougrères, “Data Protection and Enforcement in Latin America and in Uruguay (Law, Governance and Technology Series, vol 25),” in *Enforcing Privacy,* Springer, Cham, 2016, pp. 145-180.

[24] Office of the Privacy Commissioner for Personal Data, Hong Kong, “Guidance Note: Guidance on Personal Data Protection,” [Online]. Available: https://www.pcpd.org.hk/english/resources_centre/publications/guidance/files/GN_crossborder_e.pdf. [Accessed 12 December 2019].

[25] L. Fitzsimons and J. Rayner, “The GDPR in the EEA—greater harmony or increased permitted divergence?,” *The New Law Journal* , vol. 168, no. 7781, pp. 11-12, 2018.

[26] K. Nolan, “GDPR: Harmonization or Fragmentation? Applicable Law Problems in EU Data Protection Law,” Berkeley Technology Law Journal Blog, 20 January 2018. [Online]. Available: http://btlj.org/2018/01/gdpr-harmonization-or-fragmentation-applicable-law-problems-in-eu-data-protection-law/. [Accessed 12 December 2019].

[27] P. de Hert and M. Czerniawski , “Expanding the European data protection scope beyond territory: Article 3 of the General Data Protection Regulation in its wider context,” *International Data Privacy Law,* vol. 6, no. 3, p. 230–243, 2016.

[28] European Data Protection Board (EDPB), “Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) - Version for public consultation,” 16 November 2018 (adopted). [Online]. Available: https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_3_2018_territorial_scope_en.pdf. [Accessed 12 December 2019].

[29] Article 29 Data Protection Working Party, “Opinion 1/2010 on the concepts of "controller" and "processor",” 16 February 2010. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf. [Accessed 12 December 2019].

[30] Information Commissioner's Office (ICO), “Data controllers and data processors: what the difference is and what the governance implications are (20140506, v1.0),” [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1546/data-controllers-and-data-processors-dp-guidance.pdf. [Accessed 12 December 2019].

[31] Information Commissioner's Office (ICO), “Key Definitions of the Data Protection Act,” National Archives , 19 April 2012 (Archived). [Online]. Available:

http://webarchive.nationalarchives.gov.uk/20120419092552/http://www.ico.gov.uk/for_organisations/data_protection/the_guide/key_definitions.aspx. [Accessed 6 November 2018].

[32] T. Van Overstraeten and R. Cumbley, “Controllers and processors: it's all about essential means,” Linklaters, 26 February 2010. [Online]. Available: https://www.linklaters.com/en/insights/publications/2010/controllers-and-processors-its-all-about-essential-means. [Accessed 12 December 2019].

[33] T. Olsen and T. Mahler, “Identity management and data protection law: Risk, responsibility and compliance in ‘Circles of Trust’ – Part II,” *Computer Law & Security Review,* vol. 23, no. 5, pp. 415-426, 2007.

[34] R. Cregan and T. Sekou, “How much control makes a controller?,” *Privacy & Data Protection,* vol. 18, no. 6, pp. 10-12, 2018.

[35] Commission Nationale de l'Informatique et des Libertés (CNIL), “General Data Protection Regulation: Guide for Processors - September 2017 Edition,” September 2017. [Online]. Available: https://www.cnil.fr/sites/default/files/atoms/files/rgpd-guide_sous-traitant-cnil_en.pdf. [Accessed 12 December 2019].

[36] B. Treacy, “Working Party confirms 'controller' and 'processor' distinction,” *Privacy and Data Protection,* vol. 10, no. 5, pp. 3-5, 2010.

[37] C. Sullivan, “EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era,” *Computer Law & Security Review,* vol. 35, no. 4, pp. 380-397, 2019.

[38] R. Marchini , “Does the EDPB answer frequently asked questions on territorial scope?,” fieldfisher, 28 November 2018. [Online]. Available: https://privacylawblog.fieldfisher.com/2018/does-the-edpb-answer-frequently-asked-questions-on-territorial-scope. [Accessed 12 December 2019].

[39] R. Blamires, F. M. Maclean and D. van der Merwe , “EDPB Publishes Regulatory Guidance on Territorial Scope of GDPR,” Global Privacy and Security Compliance Law Blog, 3 January 2019. [Online]. Available: https://www.lexology.com/library/detail.aspx?g=52e9200d-ed53-4981-b815-4669532f3a16. [Accessed 12 December 2019].

[40] Information Commissioner's Office (ICO), “Guide to the General Data Protection Regulation (GDPR): Automated Decision Making and Profiling - What is automated decision making and profiling?,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/. [Accessed 12 December 2019].

[41] The Information Commissioner's Office (ICO), “Guide to the General Data Protection Regulation (GDPR): Individual Rights - Rights Related to Automated Decision Making Including Profiling,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/. [Accessed 24 January 2019].

[42] Information Commissioner's Office (ICO), “Guide to the General Data Protection Regulation (GDPR): Automated Decision Making and Profiling - About this Detailed Guidance,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/. [Accessed 24 January 2019].

[43] European Union Agency for Fundamental Rights (FRA), “Preventing unlawful profiling today and in the future: a guide,” 2018. [Online]. Available: https://fra.europa.eu/en/publication/2018/prevent-unlawful-profiling. [Accessed 12 December 2019].

[44] Information Commissioner's Office (ICO), “What is automated individual decision-making and profiling?,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/. [Accessed 12 February 2019].

[45] F. Bouchet , J. M. Harley, G. J. Trevors and R. Azevedo, “Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning,” *Journal of Educational Data Mining,* vol. 5, no. 1, pp. 104-146, 2013.

[46] M. Altman , A. Wood and E. Vayena , “A Harm-Reduction Framework for Algorithmic Fairness,” *IEEE Security & Privacy,* vol. 16, no. 3, 2018.

[47] Kemp Little, “A guide to GDPR profiling and automated decision-making,” 17 November 2017. [Online]. Available: https://www.kemplittle.com/blog/a-guide-to-gdpr-profiling-and-automated-decision-making/. [Accessed 25 October 2019].

[48] Information Commissioner's Office (ICO), “Right to be informed in practice,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/the-right-to-be-informed-in-practice/. [Accessed 21 February 2019].

[49] M. Veale and L. Edwards, “Comment: Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling,” *Computer Law & Security Review,* vol. 34, no. 2, pp. 398-404, 2018.

[50] L. Edwards and M. Veale, “Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For,” *Duke Law & Technology Review,* vol. 16, no. 18, 2017.

[51] “OECD Principles on AI,” [Online]. Available: https://www.oecd.org/going-digital/ai/principles/. [Accessed 12 December 2019].

[52] Fairness, Accountability, and Transparency in Machine Learning (FAT-ML), “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” [Online]. Available: https://www.fatml.org/resources/principles-for-accountable-algorithms. [Accessed 12 December 2019].

[53] [Online]. Available: Nicholas Diakopoulos et al., Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.

[54] A. D. Selbst and J. Powles, “Meaningful information and the right to explanation,” *International Data Privacy Law,* vol. 7, no. 4, p. 233–242, 2017.

[55] New Zealand Government: Stats NZ, “Algorithm Assessment Report,” October 2018. [Online]. Available: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf. [Accessed 12 December 2019].

[56] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access,* vol. 6, 2018.

[57] L. Edwards and M. Veale, “Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?,” *IEEE Security & Privacy,* vol. 16, no. 3, pp. 46-54 , 2018.

[58] Information Commissioner's Office (ICO) and The Alan Turing Institute, “Project ExplAIn: Interim Report,” 2019. [Online]. Available: https://ico.org.uk/media/about-the-ico/documents/2615039/project-explain-20190603.pdf. [Accessed 12 December 2019].

[59] S. Stalla-Bourdillon and L. Carmichael, “Data Pitch D3.7. Data Legality Report v2,” Contributions from Jérémy Decis, August 2018. [Online]. Available: https://datapitch.eu/deliverables/. [Accessed 25 June 2019].

[60] H. Lindvall, “D3.3 Data Legality Report v1,” With contributions from: Knight A. & Stalla-Bourdillon, S. Data Pitch, 31 August 2017. [Online]. Available: https://datapitch.eu/deliverables/. [Accessed 28 October 2019].

[61] G. Thuermer, J. Walker and E. Simperl, “Data Sharing Toolkit,” Data Pitch, [Online]. Available: https://datapitch.eu/datasharingtoolkit/. [Accessed 28 October 2019].

[62] M. Elliot, E. Mackey , K. O’Hara and C. Tudor , “UK Anonymisation Network (UKAN): The Anonymisation Decision-Making Framework,” 2016. [Online]. Available: http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf. [Accessed 20 February 2018].

[63] Article 29 Data Protection Working Party, “Opinion 05/2014 on Anonymisation Techniques (0829/14/EN; WP216),” 10 April 2014. [Online]. Available: http://www.pdpjournals.com/docs/88197.pdf. [Accessed 3 April 2018].

[64] N. Fulford and K. Oastler , “People, processes, technology - a how to guide to data mapping,” *Privacy & Data Protection,* vol. 16, no. 8, pp. 6-8, 2016.

[65] K. Knight, “IAPP Global Privacy Summit 2013 Presentation - Data Flow Mapping: The Good, the Bad, and the Ugly,” 7 March 2013. [Online]. Available: https://iapp.org/media/presentations/13Summit/S13_Good_Bad_Ugly_PPT.pdf. [Accessed 31 May

2018].

[66] J. Clark, “Legislative Comment - GDPR series: building a compliance programme,” *Privacy & Data Protection,* vol. 17, no. 3, pp. 7-9, 2017.

[67] IT Governance , “Data flow mapping under the EU GDPR,” [Online]. Available: https://www.itgovernance.co.uk/gdpr-data-mapping. [Accessed 31 May 2018].

[68] S. Stalla-Bourdillon and A. Knight, “Anonymous Data v. Personal Data - A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data,” *Wisconsin International Law Journal,* 2017.

[69] Information Commissioner's Office (ICO), “Determining what is personal data (v1.1),” [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf.

[70] Information Commissioner’s Office (ICO) , “What is personal data? – A quick reference guide (V1.1),” [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1549/determining_what_is_personal_data_quick_reference_guide.pdf. [Accessed 28 March 2018].

[71] Article 29 Data Protection Working Party, “Opinion 04/2007 on the concept of personal data (01248/07/EN WP 136),” 20 June 2007. [Online]. Available: http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf. [Accessed 31 May 2018].

[72] S. Wood, “ICO blog: Anonymisation – opportunities and risks,” Information Commissioner's Office (ICO) Blog, 16 November 2012. [Online]. Available: https://iconewsblog.org.uk/2012/11/16/ico-blog-anonymisation-opportunities-and-risks/. [Accessed 16 March 2018].

[73] Information Commissioner's Office (ICO), “Anonymisation: managing data protection risk code of practice,” November 2012. [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf. [Accessed 3 April 2018].

[74] Z. Alexin, “Does fair anonymization exist?,” *International Review of Law, Computers & Technology,* vol. 28, no. 1, pp. 21-44, 2013.

[75] M. Oswald , “Something Bad Might Happen: Lawyers, Anonymization, and Risk,” *XRDS: Crossroads, The ACM Magazine for Students - The Complexities of Privacy and Anonymity,* vol. 20, no. 1, pp. 22-26, 2013.

[76] Information and Privacy Commissioner, Ontario, Canada, “Looking Forward: De-identification Developments – New Tools, New Challenges,” May 2013. [Online]. Available: https://www.ipc.on.ca/wp-content/uploads/2013/05/pbd-de-identification_developments.pdf. [Accessed 13 March 2018].

[77] K. Brimsted, “Anonymisation - a process living on borrowed time?,” *Privacy & Data Protection,* vol. 14, no. 7, pp. 3-5, 2014.

[78] J. Clark, “Legislative Comment - GDPR series: anonymisation and pseudonymisation,” *Privacy & Data Protection,* vol. 18, no. 1, pp. 10-12, 2017.

[79] “Data Protection Bill [HL] 2017-19: Progress of the Bill,” [Online]. Available: https://services.parliament.uk/bills/2017-19/dataprotection.html. [Accessed 20 June 2018].

[80] “Data Protection Act 2018,” [Online]. Available: http://www.legislation.gov.uk/ukpga/2018/12/part/6/crossheading/offences-relating-to-personal-data/enacted. [Accessed 20 June 2018].

[81] S. Joyee De and D. Le Métayer, “Privacy Risk Analysis,” in *Synthesis Lectures on Information Security, Privacy, &Trust (eBook),* Morgan & Claypool Publishers, 2016, pp. 1-117.

[82] Article 29 Data Protection Working Party, “Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679 (17/EN WP 248),” 4 April 2017 (Adopted). [Online]. Available: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236. [Accessed 27 May 2018].

[83] Agencia Española de Protección de Datos (AEPD) , “GUIA PRÁCTICA DE Análisis de riesgos en los tratamientos de datos personales sujetos al RGPD,” [Online]. Available: https://iapp.org/media/pdf/resource_center/AnalisisDeRiesgosRGPD.pdf. [Accessed 15 March 2018].

[84] Information Commissioner's Office (ICO), “Examples of processing ‘likely to result in high risk’,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/examples-of-processing-likely-to-result-in-high-risk/. [Accessed 26 May 2018].

[85] R. Clark, “Quality Assurance for Security Applications of Big Data,” in *European Intelligence and Security Informatics Conference (EISIC'16)*, Uppsala, Sweden, 2016.

[86] R. D. Riley, P. C. Lambert and G. Abo-Zaid, “Meta-analysis of individual participant data: rationale, conduct, and reporting,” *The British Medical Journal,* vol. 340, p. c221, 2010.

[87] Norwegian Centre for Research Data (NSD), “Individual Level Data,” [Online]. Available: http://www.nsd.uib.no/nsd/english/individualdata.html. [Accessed 14 March 2018].

[88] Information Commissioner's Office (ICO) , “Guide to the General Data Protection Regulation (GDPR): Lawful basis for processing,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/#ib3. [Accessed 30 May 2018].

[89] Information Commissioner's Office (ICO), “ICO: Lawful Basis Interactive Guidance Tool,” [Online]. Available: https://ico.org.uk/for-organisations/resources-and-support/lawful-basis-interactive-guidance-tool/. [Accessed 30 May 2018].

[90] PricewaterhouseCoopers (PwC), “General Data Protection Regulation: Anonymisation and pseudonymisation,” 2017. [Online]. Available: https://www.pwc.com.cy/en/publications/assets/general-data-protection-regulation-anonymisation-and-pseudonymisation-january-2017.pdf. [Accessed 3 April 2018].

[91] European Union Agency for Network and Information Security (ENISA), “Handbook on Security of Personal Data Processing,” December 2017. [Online]. Available: https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing. [Accessed 12 March 2018].

[92] Information Commissioner's Office (ICO), “Data protection impact assessments,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/. [Accessed 30 May 2018].

[93] Commission nationale de l'informatique et des libertés (CNIL), “The open source PIA software helps to carry out data protection impact assesment,” 29 January 2018. [Online]. Available: https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment. [Accessed 30 May 2018].

[94] Information Commissioner's Office (ICO), “How do we document our processing activities?,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/documentation/how-do-we-document-our-processing-activities/. [Accessed 31 May 2018].

[95] Open Data Institute (ODI), “Mapping data ecosystems,” 23 March 2018. [Online]. Available: https://docs.google.com/document/d/1vSqoHOYT5u6vrCHIebCS0rze0gWwXOspeEowWzwake8/edit. [Accessed 20 June 2018].

[96] J. Graves, “Data flow management: why and how,” *Network Security,* no. 1, pp. 5-6, 2017.

[97] LINDDUN Privacy Threat Modelling, [Online]. Available: https://linddun.org/index.php. [Accessed 31 May 2018].

[98] IT Governance, “Data Flow Mapping Tool,” [Online]. Available: https://www.itgovernance.co.uk/shop/Product/data-flow-mapping-tool. [Accessed 31 May 2018].

[99] HM Government, “Open Data White Paper: Unleashing the Potential,” June 2012. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf. [Accessed 17 November 2019].

[100] L. Floridi and M. Taddeo, “What is data ethics?,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 374, no. 2083, 2016.

[101] S. Yakovleva, “Should fundamental rights to privacy and data protection be a part of the EU's international trade "deals"?,” *World Trade Review,* vol. 17, no. 3, pp. 477-508, 2018.

[102] P. Van den Bulck, “Transfers of personal data to third countries,” *ERA Forum,* vol. 18, no. 2, p. 229–247, 2017.

[103] N. Sen, “Understanding the Role of the WTO in International Data Flows: Taking the Liberalization or the Regulatory Autonomy Path?,” *Journal of International Economic Law,* vol. 21, no. 2, p. 323, 2018.

[104] J. Selby, “Data localization laws: trade barriers or legitimate responses to cybersecurity risks, or both?,” *International Journal of Law and Information Technology,* vol. 25, no. 3, p. 213–232, 2017.

[105] H. Rowe, “Examining The Current Position On Cross-Border Information Transfers – In Theory And In Practice, Part 2,” *Journal of International Banking & Financial Law,* vol. 17, no. 5, pp. 206-212, 2002.

[106] B. Maksó, “Exporting the Policy - International Data Transfer and the Role of Binding Corporate Rules for Ensuring Adequate Safeguards,” *Pecs Journal of International and European Law,* no. 2, pp. 79-86, 2016.

[107] L. Kong, “Data Protection and Transborder Data Flow in the European and Global Context,” *European Journal of International Law,* vol. 21, no. 2, p. 441–456, 2010.

[108] S. Khan, “Invalidity of EU-US Safe Harbor: Part 2: practical implications and the new Privacy Shield,” *Compliance & Risk,* vol. 5, no. 3, pp. 10-12, 2016.

[109] F. Good, S. Sayers and O. Wint, “GDPR series: how to legitimise your profiling activities,” *Privacy & Data Protection,* vol. 18, no. 3, pp. 7-10, 2018.

[110] C. S. Dempwolf, J. Auer and M. D’Ippolito , “Innovation Accelerators: Defining Characteristics Among Startup Assistance Organizations (Small Business Administration, Office of Advocacy under contract number SBAHQ-13-M-0197),” October 2014. [Online]. Available: https://www.sba.gov/sites/default/files/rs425-Innovation-Accelerators-Report-FINAL.pdf. [Accessed 22 March 2018].

[111] B. Casey , A. Farhangi and R. Vogl, “Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise,” *Berkeley Technology Law Journal, Forthcoming. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3143325,* 19 February 2018.

[112] E. Brownsdon, “The impact of the Data Protection Act 1998 on e-business,” *New Law Journal,* vol. 151, no. 7000, pp. 1360-1363, 2001.

[113] M. Bogers, H. Chesbrough and C. Moedas, “Open Innovation: Research, Practices, and Policies,” *California Management Review,* vol. 60, no. 2, pp. 5-16, 2018.

[114] M. F. Badran, “Economic impact of data localization in five selected African countries,” *Digital Policy, Regulation & Governance,* vol. 20, no. 4, pp. 337-357, 2018.

[115] R. J. Allio, “Interview with Henry Chesbrough: innovating innovation,” *Strategy & Leadership,* vol. 33, no. 1, pp. 19-24, 2005.

[116] K. Albrecht and K. L. Lust, “GDPR series: international data transfers - a high level review (legislative comment),” *Privacy & Data Protection,* vol. 18, no. 2, pp. 14-16, 2017.

[117] Open Data Institute (ODI), “The Data Spectrum: The Data Spectrum helps you understand the language of data,” [Online]. Available: https://theodi.org/about-the-odi/the-data-spectrum/. [Accessed 22 March 2018].

[118] European Commission (EC) , “Open Innovation Resources: Policy initiatives, funding schemes and support services related to open innovation.,” [Online]. Available: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-innovation-resources_en. [Accessed 26 May 2018].

[119] CNIL, “How can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence,” 26 December 2017. [Online]. Available: https://www.cnil.fr/en/how-can-humans-keep-upper-hand-report-ethical-matters-raised-algorithms-and-artificial-intelligence. [Accessed 14 August 2018].

[120] “Eight months until Brexit and still no adequacy deal (editorial),” *Privacy & Data Protection,* vol. 18, no. 7, pp. 15-16, 2018.

[121] European Commission (EC) , “Adequacy of the protection of personal data in non-EU countries:

How the EU determines if a non-EU country has an adequate level of data protection," [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en. [Accessed 16 October 2018].

[122] Data Pitch - Innovation Programme, "About Data Pitch," [Online]. Available: https://datapitch.eu/about-us/start-up/. [Accessed 22 March 2018].

[123] G. González Fuster, "Un-mapping Personal Data Transfers," *European Data Protection Law Review,* no. 2, pp. 160-168, 2016.

[124] S. Wachter, B. Mittelstadt and L. Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," *International Data Privacy Law,* vol. 7, no. 2, p. 76–99, 2017.

[125] S. Sayers and J. Drury-Smith, "GDPR series: How to operationalise profiling for your organisation," *Privacy and Data Protection,* vol. 17, no. 1, pp. 3-6, 2017.

[126] A. Burt, "Is there a 'right to explanation' for machine learning in the GDPR?," International Association for Privacy Professionals (iapp), 1 June 2017. [Online]. Available: https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/. [Accessed 22 January 2019].

[127] R. Massey and E. Machin, "European Data Protection Board publishes draft guidelines on the territorial scope of GDPR," *Computer and Telecommunications Law Review,* vol. 25, no. 2, pp. 35-40, 2019.

[128] G. Malgieri and G. Comandé, "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation," *International Data Privacy Law,* vol. 7, no. 4, pp. 243-265, 2017.

[129] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society,* 2016.

[130] P. B. de Laat, "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?," *Philosophy & Technology,* vol. 31, no. 4, p. 525–541, 2018.

[131] H. Nissenbaum, "Accountability in a computerized society," *Science and Engineering Ethics,* vol. 2, no. 1, p. 25–42, 1996.

[132] F. Kaltheuner and E. Bietti, "Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR," *Journal of Information Rights, Policy and Practice,* vol. 2, no. 2, pp. 1-17, 2018.

[133] S. Wachter, B. Mittelstadt and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology,* vol. 31, no. 2, pp. 841-887, 2018.

[134] D. Sinclair, "Case Comment: Counterfactuals in anti-competitive contracts and abuse of dominance cases under articles 101 and 102 TFEU," *European Competition Law Review,* vol. 31, no. 12, pp. 509-513, 2010.

[135] B. Goodman and S. Flaxman, "European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"," *AI Magazine,* vol. 38, no. 3, pp. 50-57, 2017.

[136] S. Olhede and R. Rodrigues, "Fairness and transparency in the age of the algorithm," *Significance,* vol. 14, no. 2, pp. 8-9, 2017.

[137] M. E. Kaminski, "The Right to Explanation, Explained.," University of Colorado Law Legal Studies Research Paper No. 18-24., 15 June 2018. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3196985. [Accessed 14 August 2018].

[138] A. Chander and R. Srinivasan, "Evaluating Explanations by Cognitive Value," in *Machine Learning and Knowledge Extraction. CD-MAKE 2018. Lecture Notes in Computer Science, vol 11015,* Springer, Cham, 2018.

[139] M. Brkan, "Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond," *International Journal of Law and Information Technology,* vol. eay017, 2019.

[140] F. Doshi-Velez, M. Kortz, R. Budish, C. Bavitz, S. J. Gershman, D. O'Brien, S. Shieber, J. Waldo,

D. Weinberger and A. Wood, “Accountability of AI Under the Law: The Role of Explanation,” Berkman Center Research Publication Forthcoming - Harvard Public Law Working Paper No. 18-07 (last revised 13 April 2018), 3 November 2017. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064761. [Accessed 19 February 2019].

[141] N. Wallace, “EU's Right to Explanation: A Harmful Restriction on Artificial Intelligence,” TECHZONE360 (Online), 25 January 2017. [Online]. Available: https://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#. [Accessed 22 February 2019].

[142] C. Kuang, “Can A.I. Be Taught to Explain Itself?,” New York Times Magazine (Online), 21 November 2017. [Online]. Available: https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html. [Accessed 22 February 2019].

[143] [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-else-do-we-need-to-consider-if-article-22-applies/.

[144] “Definition of rationale,” Oxford English Dictionaries (OED) Online, [Online]. Available: https://en.oxforddictionaries.com/definition/rationale. [Accessed 26 February 2019].

[145] M. E. Maisog, “Making the Case Against Data Localization in China,” International Association of Privacy Professionals (iapp), 20 April 2015. [Online]. Available: https://iapp.org/news/a/making-the-case-against-data-localization-in-china/. [Accessed 12 June 2019].

[146] E. van der Marel , H. Lee-Makiyama and M. Bauer , “The Costs of Data Localisation: A Friendly Fire on Economic Recovery,” European Centre for International Political Economy (ECIPE), May 2014. [Online]. Available: https://ecipe.org/publications/dataloc/. [Accessed 11 June 2019].

[147] C. Bennett and S. Oduro-Marfo, “GLOBAL Privacy Protection: Adequate Laws, Accountable Organizations and/or Data Localization?,” in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp'18),* Singapore, 2018.

[148] Lexico (powered by Oxford English Dictionaries), “Definition of 'transnational',” [Online]. Available: https://www.lexico.com/en/definition/transnational. [Accessed 20 June 2019].

[149] Lexico (powered by Oxford English Dictionairies), “Definition of "cross-sectoral",” [Online]. Available: https://www.lexico.com/en/definition/cross-sectoral. [Accessed 20 June 2019].

[150] European Commission, “Guidance on sharing private sector data in the European data economy [COM(2018) 232 final],” COMMISSION STAFF WORKING DOCUMENT. Accompanying the document: Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the Regions "Towards a common European data space" , 25 April 2018. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy. [Accessed 28 May 2019].

[151] E. Scaria, A. Berghmans, M. Pont, C. Arnaut and S. Leconte, “Study on data sharing between companies in Europe,” A study prepared for the European Commission Directorate-General for Communications Networks, Content and Technology by everis Benelux , 24 April 2018. [Online]. Available: https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en. [Accessed 29 May 2019].

[152] Organisation for Economic Co-operation and Development (OECD), “Open Innovation in Global Networks,” 2008. [Online]. Available: https://doi.org/10.1787/9789264047693-4-en. [Accessed 25 June 2019].

[153] [Online]. Available: https://www.lexico.com/en/definition/innovate.

[154] H. Chesbrough, “1. Open Innovation: A New Paradigm for Understanding Industrial Innovation,” in *Open Innovation : Researching a New Paradigm,* Oxford , Oxford University Press, 2006, pp. 1-14.

[155] C. Henry, Open Services Innovation : Rethinking Your Business to Grow and Compete in a New Era, John Wiley & Sons, 2011.

[156] D. Cushman and J. Burke, “Open Innovation,” in *The 10 Principles of Open Business,* London, Palgrave Macmillan, 2014, pp. 106-134.

[157] I. Susha , M. Janssen and S. Verhulst , “Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development,” in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[158] HM Treasury, “Review of quality assurance of Government analytical models: final report,” March 2013. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/206946/review_of_qa_of_govt_analytical_models_final_report_040313.pdf. [Accessed 16 July 2019].

[159] E. Broad, “Closed, shared, open data: what’s in a name?,” Open Data Institute (ODI), 17 September 2015. [Online]. Available: http://oldsite.theodi.org/blog/closed-shared-open-data-whats-in-a-name.

[160] “Council Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector, as amended in 2009 by the Citizens' Rights Directive 2009/136/EC (E-Privacy Directive),” [Online]. Available: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML. [Accessed 22 October 2019].

[161] “Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC,” [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017PC0010. [Accessed 22 October 2019].

[162] E. Duhs, “The future of the ePrivacy Regulation and the impact of Brexit on its application in UK,” Field Fisher Blog, 29 April 2019. [Online]. Available: https://privacylawblog.fieldfisher.com/2019/the-future-of-the-eprivacy-regulation-and-the-impact-of-brexit-on-its-application-in-uk. [Accessed 22 October 2019].

[163] European Commission - Cybersecurity & Digital Privacy Policy (Unit H.2) |, “Digital Single Market - Digital Privacy,” 25 September 2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/online-privacy. [Accessed 6 November 2019].

[164] Article 29 Data Protection Working Party, “Guidelines on Article 49 of Regulation 2016/679,” 6 February 2018. [Online]. Available: https://www.pdpjournals.com/docs/8878884.pdf. [Accessed 16 November 2019].

[165] W. Kerber, “Editorial - Governance of data: exclusive property v access,” *International Review of Intellectual Property and Competition Law,* vol. 47, no. 7, pp. 759-762, 2016.

[166] G. Thomas, “Assigning Data Ownership,” The Data Governance Institute, 28 September 2013. [Online]. Available: http://www.datagovernance.com/assigning-data-ownership/. [Accessed 10 April 2018].

[167] G. Thomas, “Working with Data Stewards,” The Data Governance Institute, 28 September 2013. [Online]. Available: http://www.datagovernance.com/working-with-data-stewards/. [Accessed 10 April 2018].

[168] T. Hoeren, “Big data and the ownership in data: recent developments in Europe,” *European Intellectual Property Review* , vol. 36, no. 12, pp. 751-754 , 2014.

[169] R. Hu, S. Stalla-Bourdillon, M. Yang, V. Schiavo and V. Sassone, “Bridging Policy, Regulation and Practice?A techno-legal Analysis of Three Types of Data in the GDPR,” in *Data Protection and Privacy*, Hart Publishing, 2017.

[170] Oxford English Dictionaries (OED) Online , “Define: anonymise,” [Online]. Available: https://en.oxforddictionaries.com/definition/anonymize. [Accessed 11 April 2018].

[171] C. M. O’Keefe, S. Otorepec, M. Elliot, E. Mackey and K. O’Hara, “The De-Identification DecisionMaking Framework (CSIRO Reports EP173122 and EP175702),” 18 September 2017. [Online]. Available: https://publications.csiro.au/rpr/download?pid=csiro:EP173122&dsid=DS2. [Accessed 12 April 2018].

[172] Australian Government: , “De-Identification Decision-Making Framework: Office of the Australian Information Commissioner,” [Online]. Available: https://www.oaic.gov.au/agencies-and-organisations/guides/de-identification-decision-making-framework. [Accessed 12 April 2018].

[173] Information Commissioner's Office (ICO), [Online]. Available: https://ico.org.uk/. [Accessed 20 6 2018].

[174] European Data Protection Supervisor (EDPS): The EU's independent data protection authority , “Glossary: "Article 29 Working Party",” [Online]. Available: https://edps.europa.eu/data-protection/data-protection/glossary/a_en. [Accessed 13 April 2018].

[175] “DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data,” 24 October 1995. [Online]. Available: https://edps.europa.eu/sites/edp/files/publication/dir_1995_46_en.pdf. [Accessed 13 April 2018].

[176] European Commission Website, “Article 29 Working Party Newsroom,” [Online]. Available: http://ec.europa.eu/newsroom/article29/news.cfm?item_type=1358. [Accessed 13 April 2018].

[177] The International Association of Privacy Professionals (iapp) and OneTrust: Privacy Management Software, “IAPP-OneTrust Research: Bridging ISO 27001 to GDPR,” 27 March 2018. [Online]. Available: https://iapp.org/media/pdf/resource_center/IAPP-OneTrust-Bridging-ISO-GDPR-final.pdf. [Accessed 27 March 2018].

[178] B. Lubarsky, “Re-Identification of “Anonymized” Data,” *Georgetown Law Technology Review,* vol. 202, no. 1, 2017.

[179] J. Polonetsky , O. Tene and K. Finch, “Shades of Gray: Seeing the Full Spectrum of Practical Data De-identification,” *Santa Clara Law Review,* vol. 56, no. 3, 2016.

[180] P. M. Schwartz and D. J. Solove, “The PII Problem: Privacy and a New Concept of Personally Identifiable Information,” *N.Y.U. L.Q. Rev.,* vol. 86, no. 1814, 2011.

[181] McDermott Will & Emery, “Article 29 working party opinion on the definition of consent: an unambiguous view of the future,” Lexology, 28 September 2011. [Online]. Available: https://www.lexology.com/library/detail.aspx?g=01d43c3a-2a36-4d3d-b0d6-cfdb1d3be067. [Accessed 13 April 2018].

[182] OUT-LAW.COM, “ICO: Anonymised data doesn't HAVE to guarantee your privacy,” The Register , 26 November 2012. [Online]. Available: https://www.theregister.co.uk/2012/11/26/anonymising_data_does_not_guarantee_privacy/. [Accessed 13 April 2018].

[183] European Union Agency for Fundamental Rights, European Court of Human Rights, and Council of Europe, “Handbook on European data protection law,” June 2014. [Online]. Available: http://fra.europa.eu/en/publication/2014/handbook-european-data-protection-law. [Accessed 13 April 2018].

[184] A. Hern, “New law could criminalise uncovering personal data abuses, advocate warns,” The Guardian Online, 14 August 2017. [Online]. Available: https://www.theguardian.com/technology/2017/aug/14/data-protection-bill-criminalise-privacy-research-advocate-warns. [Accessed 11 April 2018].

[185] P. Ralph and S. Ng, “Will there be a new data protection offence for the UK beyond GDPR?,” PricewaterhouseCoopers (PwC): Data protection and privacy gloabl insights, 8 September 2017. [Online]. Available: http://pwc.blogs.com/data_protection/2017/09/will-there-be-a-new-data-protection-offence-for-the-uk-beyond-gdpr.html. [Accessed 13 April 2018].

[186] R. Chirgwin, “UK Data Protection Bill tweaked to protect security researchers: Re-identification of data will not be a crime, as long as you warn the authorities,” The Register, 10 January 2018. [Online]. Available: https://www.theregister.co.uk/2018/01/10/uk_data_protection_bill_tweaked_to_protect_security_researchers/. [Accessed 13 April 2018].

[187] M. Phillips, E. S. Dove and B. M. Knoppers, “Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?,” *Journal of Bioethical Inquiry,* vol. 14, no. 4, p. 527–539, 2017.

[188] R. Thomas , “Risk, accountability, and binding corporate codes: a "smarter" approach to data protection,” *Privacy & Data Protection,* vol. 13, no. 7, pp. 3-6, 2013.

[189] Agencia Española de Protección de Datos (AEPD), [Online]. Available: https://www.agpd.es/portalwebAGPD/index-iden-idphp.php. [Accessed 15 March 2018].

[190] The UK Data Service, “Census microdata guide: "Samples of individual person-level records drawn from the 1991, 2001 and 2011 Censuses",” [Online]. Available:

https://census.ukdataservice.ac.uk/use-data/guides/microdata. [Accessed 30 May 2018].

[191] The Organisation for Economic Co-operation and Development (OECD), “Glossary of Statistical Terms: Aggregation,” 10 June 2013 (last updated). [Online]. Available: https://stats.oecd.org/glossary/detail.asp?ID=68. [Accessed 14 March 2018].

[192] M. Rouse, “Definition: data aggregation,” TechTarget, September 2005 (last updated). [Online]. Available: http://searchsqlserver.techtarget.com/definition/data-aggregation. [Accessed 14 March 2018].

[193] UK Data Service, “Census aggregate data guide,” [Online]. Available: https://census.ukdataservice.ac.uk/use-data/guides/aggregate-data. [Accessed 30 May 2018].

[194] UK Data Service, “Manage Data: Document Your Data - "Make data clear to understand and easy to use",” [Online]. Available: https://www.ukdataservice.ac.uk/manage-data/document. [Accessed 4 June 2018].

[195] L. Moreau, “The Foundations for Provenance on the Web,” *Foundations and Trends in Web Science* , vol. 2, no. 2–3, pp. 99-241, 2010.

[196] Executive Office of the President, “Big Data: Seizing Opportunities, Preserving Values,” The White House, Washington D.C. , May 2014. [Online]. Available: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. [Accessed 17 November 2019].

[197] A. Mantelero, “From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era,” in *Group Privacy: New Challenges of Data Technologies*, Springer, Cham, 2017, pp. 139-158.

[198] Information Commissioner's Office (ICO), “GDPR Guidance - Security,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/security/. [Accessed 17 November 2019].

[199] Article 29 Data Protection Working Party , “Opinion 05/2012 on Cloud Computing,” 1 July 2012. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf. [Accessed 17 November 2019].

[200] Department for Digital, Culture, Media & Sport, “Guidance: Data Ethics Framework,” 30 August 2018. [Online]. Available: https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework. [Accessed 17 November 2019].

[201] Open Data Institute (ODI), “Why should we care about data ethics?,” 14 September 2019. [Online]. Available: https://theodi.org/article/why-should-we-care-about-data-ethics/. [Accessed 17 November 2019].

[202] European Free Trade Association (EFTA), “The Basic Features of the EEA Agreement,” [Online]. Available: http://www.efta.int/eea/eea-agreement/eea-basic-features. [Accessed 12 December 2019].

[203] G. Greenleaf, “The Influence of European Data Privacy Standards Outside Europe: Implications for Globalisation of Convention 108,” *International Data Privacy Law,* vol. 2, no. 2, p. 68–92, 2012.

[204] S. A. Aaronson and P. Leblond, “Another Digital Divide: The Rise of Data Realms and its Implications for the WTO,” *Journal of International Economic Law,* vol. 21, no. 2, pp. 245-272, 2018.

[205] Commission nationale de l'informatique et des libertés (CNIL), “Data protection around the world,” 19 November 2019. [Online]. Available: https://www.cnil.fr/en/data-protection-around-the-world. [Accessed 12 December 2019].

[206] DLA Piper, “Data Protection Laws of the World: Full Handbook,” [Online]. Available: https://www.dlapiperdataprotection.com/index.html?t=about&c=AO. [Accessed 12 December 2019].

[207] International Association of Privacy Professionals (IAPP), [Online]. Available: https://iapp.org/. [Accessed 12 December 2019].

[208] Eversheds - Sutherland, “General Data Protection Regulation – International data transfers,” 15 April 2016. [Online]. Available: https://www.eversheds-sutherland.com/global/en/what/articles/index.page?ArticleID=en/Data-

Protection/International_Data_Transfers. [Accessed 12 December 2019].

[209] W. K. Hon, C. Millard, J. Singh, I. Walden and J. Crowcroft, "Policy, legal and regulatory implications of a Europe-only cloud," *International Journal of Law & Information Technology,* vol. 24, no. 3, pp. 251-278, 2016.

[210] Information Commissioner's Office (ICO), "The eighth data protection principle and international data transfers (20170630, Version: 4.1)," [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1566/international_transfers_legal_guidance.pdf. [Accessed 12 December 2019].

[211] Oxford English Dictionaries (OED) Online, "Define: transit," [Online]. Available: https://www.lexico.com/definition/transit. [Accessed 12 December 2019].

[212] P. Hustinx , "Opinion of the European Data Protection Supervisor (EDPS) on the data protection reform package," 7 March 2012. [Online]. Available: https://edps.europa.eu/sites/edp/files/publication/12-03-07_edps_reform_package_en.pdf. [Accessed 12 December 2019].

[213] X. Konarski, D. Karwala, H. Schulte-Nölke and S. Charlton, "Reforming the Data Protection Package Study (IP/A/IMCO/ST/2012-02)," European Parliament - Directorate General for Internal Policies: Policy Department Economic and Scientific Policy A, September 2012. [Online]. Available: http://www.europarl.europa.eu/document/activities/cont/201209/20120928ATT52488/20120928ATT52488EN.pdf. [Accessed 12 December 2019].

[214] Information Commissioner's Office (ICO), "Data Protection in 2010: Workshop F - Introduction to international data transfers," 2010. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20130102222852/http://www.ico.gov.uk/news/current_topics/~/media/documents/dpo_conference_2010/WORKSHOP_F_OUTCOMES.ashx. [Accessed 12 December 2019].

[215] Out-law.com (Pinsent Masons), "Introduction to overseas transfers of personal data," March 2013. [Online]. Available: https://www.pinsentmasons.com/out-law/guides/introduction-to-overseas-transfers-of-personal-data. [Accessed 12 December 2019].

[216] European Commission (EC), "What rules apply if my organisation transfers data outside the EU?," [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-rules-apply-if-my-organisation-transfers-data-outside-eu_en. [Accessed 12 December 2019].

[217] "COMMISSION IMPLEMENTING DECISION (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield (notified under document C(2016) 4176)," Official Journal of the European Union, 12 July 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016D1250&from=EN. [Accessed 12 December 2019].

[218] European Commission, Directorate-General Justice and Consumers, "Notice to stakeholders: withdrawal of the United Kingdom and EU rules in the field of data protection," 9 January 2018. [Online]. Available: http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=611943&utm_source=just_newsroom&utm_medium=Website&utm_campaign=just&utm_content=Notice%20to%20stakeholders%20withdrawal%20of%20the%20United%20Kingdom%20and%20EU%20rules%20in%20th&utm_term=Data%20p. [Accessed 12 December 2019].

[219] HM Government: Department for Exiting the European Union, "The exchange and protection of personal data: a future partnership paper (Policy Paper)," 24 August 2017. [Online]. Available: https://www.gov.uk/government/publications/the-exchange-and-protection-of-personal-data-a-future-partnership-paper. [Accessed 12 December 2019].

[220] European Commission, "Binding Corporate Rules (BCR): corporate rules for data transfers within multinational companies," [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/binding-corporate-rules-bcr_en. [Accessed 12 December 2019].

[221] European Commission, “Standard Contractual Clauses (SCC): Standard contractual clauses for data transfers between EU and non-EU countries,” [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en. [Accessed 12 December 2019].

[222] The In-house Lawyer, “Are there restrictions on the transfer of personal data overseas?,” 10 October 2017. [Online]. Available: http://www.inhouselawyer.co.uk/wgd_question/are-there-restrictions-on-the-transfer-of-personal-data-overseas/. [Accessed 12 December 2019].

[223] Amberhawk Training, “How 'flexible' can the UK actually be on EU data protection law?,” The Register, 4 May 2016. [Online]. Available: https://www.theregister.co.uk/2016/05/04/will_the_uks_approach_to_the_gdpr_be_harmonised/. [Accessed 12 December 2019].

[224] Hunton Andrews Kurth, “EDPB Publishes Guidelines on Extraterritorial Application of the GDPR,” Privacy & Information Security Law Blog: Global Privacy and Cybersecurity Law Updates and Analysis, 27 November 2018. [Online]. Available: https://www.huntonprivacyblog.com/2018/11/27/edpb-publishes-guidelines-on-extraterritorial-application-of-the-gdpr/. [Accessed 12 December 2019].

[225] Privacy Matters: DLA Piper's Global Privacy & Data Protection Resource, “EU: New EDPB Guidelines on the territorial scope of the GDPR,” 28 November 2018. [Online]. Available: https://blogs.dlapiper.com/privacymatters/eu-new-edpb-guidelines-on-the-territorial-scope-of-the-gdpr/. [Accessed 12 December 2019].

[226] N. Wallace and D. Castro, “The Impact of the EU’s New Data Protection Regulation on AI,” Center for Data Innovation, 27 March 2018. [Online]. Available: http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf. [Accessed 12 December 2019].

7. Appendices: Final Legal and Privacy Toolkit

7.1 Appendix A. Toolkit overview

Given the D3.9 report comprises the last update to the Legal and Privacy Toolkit (“the toolkit”), this toolkit overview presents an opportunity to summarise the final configuration of the toolkit as well as highlight key areas covered by earlier versions.[84] This toolkit overview begins by outlining the aims of the EU Horizon 2020 (H2020) funded Data Pitch open innovation programme (which commenced in January 2017 and concluded in December 2019) and the principal underlying factors that necessitated the development of the toolkit for the programme. In essence, the need to support those involved in the Data Pitch programme – and similar data sharing schemes – to share and re-use data lawfully and ethically. It then provides a brief outline of the toolkit, including its purpose, intended audience, focus, configuration, and relation to other parts of the Data Pitch Consortium strategy for handling data processing issues under the programme.

7.1.1 Aims of the Data Pitch open innovation programme

The overall objective of the 3-year Data Pitch programme was to increase the value of data and to enable small European companies to reap the full benefits of secure data sharing and re-use. This toolkit is aimed at giving those organisations who took part in the programme – and who participate in similar data sharing schemes – confidence to innovate with data in legally compliant and ethical ways. In brief, Participating SMEs[85] were able to securely access and process data shared by Data Providers[86] and/or (re)use self-sourced data in response to a particular challenge – after they qualified to take part in the main acceleration (experimental) stage of the programme:

[84] Note: this preamble is an extended and modified version of sections 1 to 4 of the Legal and Privacy Toolkit v1 [3, pp. 11-17] that was delivered in June 2017.

[85] Those organisations that were selected to process data under the acceleration stage of the open innovation programme. Such data was collected externally by the Data Providers for purposes other than the Data Pitch programme.

[86] Those organisations that agreed to supply their data under the Data Pitch programme.

Three principal types of data shared and (re-)used for innovation purposes as part of the Data Pitch Programme:

- **“(Data Provider) Data.** These data are: (a) shared by a Data Provider under a signed Data Provider (Data Sharing) Agreement; and subsequently (b) re-used for innovation purposes by successful SMEs – as part of the data provider challenges track* – pursuant to a signed SME Contract.
- **SME Data.** These data are: (a) gathered and/or collected by a successful SME; and/or (b) obtained from a third party by an SME. These data can be (re-)used for innovation purposes alongside (Data Provider) Data - as part of the data provider challenges track by successful SMEs (subject to conditions) under a signed SME Contract.
- **SME Self-Sourced Data.** These data are: (a) gathered and/or collected by a successful SME; and/or (b) obtained from a third party by an SME. These data are (re-)used for innovation purposes – as part of the sectoral or open innovation challenges tracks* by successful SMEs (subject to conditions) under a signed SME Self-Sourced Data Contract.”

Three Data Pitch challenge tracks:

“*Note that in the first call of the Data Pitch programme (2017-2018), there were twelve challenges and in the second call of the Data Pitch programme (2018-2019), there were sixteen challenges. Data Pitch sets its challenges across the following three challenge tracks:

- **Track 1: data provider challenges.** The data provider determines the challenge and provides certain data (i.e. data provider data) for the successful applicant to reuse as part of their solution to the specific problem raised. Furthermore, the successful applicant is able to include other datasets (i.e. SME Data) if required.
- **Track 2: sectoral challenges.** The Data Pitch consortium sets the challenge that requires the successful applicant to provide their own data (i.e. SME Self-Sourced Data) in order to solve the problem raised.
- **Track 3: open innovation challenge.** Applicants to the programme are able to propose an innovative solution that (re)uses data that are self-sourced (i.e. SME Self-Sourced Data).”

Source: Extract taken from the Data Legality Report v2 [59, pp. 10-11]

Incentivising the sharing and re-usage of closed data

While Participating SMEs were able to self-source data as part of their open innovation activities,[87] a central focus for the Data Pitch programme was to incentivise Data Providers to share closed datasets via the data provider challenges track. For the purposes of the toolkit, we define closed data as: datasets collected externally by Data Providers for purposes other than the Data Pitch programme, which would otherwise remain inaccessible to Participating SMEs.[88] Participating SMEs processed these data in a secure environment, for multiple defined challenge purposes, and provided the results from the analysed data to Data Providers. This exercise creates new opportunities, but also legal and ethical challenges.

7.1.2 Potential legal risks involved in data sharing and re-use

Data sharing and its re-use for new purposes is crucial for the economy and society to work smarter. At the same time, appropriate safeguards must be put in place to protect privacy. This is fundamental to any data sharing and data (re)use regime. Organisations involved in the processing of data must comply with the rules set out in any legislation that might apply to such processing, which in turn requires that they be aware of what rules are relevant and what exactly they require.

[87] Where this is lawful, ethical and considered appropriate for the specific challenge.

[88] Also note the ODI [159] definition of closed data: *"Data that can only be accessed by its subject, owner or holder."*

Unfortunately, the legal rules surrounding data sharing and re-use can be complex.[89] In particular, while data cannot be owned in the same way that, say, a house or car is owned, extensive rights and obligations can arise in relation to data. For example, there are important legal principles about how data can be shared by those, who have vested rights, with third parties who may in turn re-purpose that data for different uses. These complexities can arise from different areas of law, and with respect to different regulatory policies, apply to various categories of data/information.

At the same time, some areas of law that map such rights and obligations are developing rapidly, and are likely to develop even more quickly - as big data analytical techniques become more prevalent. These legislative and regulatory developments also point towards greater legal analysis and risk management being required by organisations handling data in the future – including in respect of analysing what rights subsist in relation to data in any given case and managing associated risks. This has led to a cautious approach to data sharing, which may go beyond what is required by the law in any one situation. Indeed, confusion around the legalities and risks of data sharing can lead to data simply not being made available to third parties, even though in reality there may be lawful solutions to data sharing. These solutions need to be explored and then an appropriate strategy can be identified.

When considering whether to share data for a specific analytical purpose or purposes, organisations should first consider the rights that arise in relation to such data, as well as the obligations upon them in relation to such data, including any constraints that flow from these obligations. For example, some data (such as highly confidential data) may only be legitimate to share in very narrowly-drawn scenarios. Alternatively, other powers may allow for greater flexibility as long as broad principles are complied with. However, that flexibility will rarely be unlimited where rights in relation to data exist. As a general rule, the assessment of whether an individual act of data-sharing is permissible should be considered on a case-by-case basis.

Consequences of non-compliance with relevant laws

Compliance with relevant laws is very important. For instance, a breach of data protection laws can result in enforcement action against the organisations involved by national data protection authorities which in the case of the UK, would be the Information Commissioner's Office (ICO). The consequences of non-compliance can also be severe, over and above the reputational damage that might arise from involvement in, for example, a data security breach. They can give rise to the imposition of fines and other sanctions, as well as 'damages' (financial settlement) being awarded by a court against the parties involved following a court case. For instance, the General Data Protection Regulation (GDPR) sees the introduction of a new regime of significant fines that can be imposed by data protection authorities where data protection laws are broken. These fines, in the cases of the most severe data protection law infringements can be as high as €20 million, or 4% of global annual turnover for the preceding financial year, whichever is greater.

The role of the Data Pitch Consortium

A core duty of the Data Pitch Consortium was to ensure that Data Providers and Participating SMEs met the same high standards and effective level of legal compliance respected by the organisations that made up the Consortium, i.e. the University of Southampton, Beta-I, Dawex, and the UK Open Data Institute (ODI). To this extent, the Consortium acted as a facilitator so that Data Providers and Participating SMEs were able to access and process data shared under the programme lawfully and ethically. At the same time, the Consortium recognised that a balance needed to be struck between the use of new data and techniques, and associated risks – such as re-identification risk[90] arising from the processing of data relating to persons shared under the programme. At the planning stage of the programme, it was therefore necessary for the Consortium to adopt strategies applicable to all types of data shared and re-used under the programme – including strategies for anonymisation,

[89] For more information see: 'Mapping the legal framework' in Appendix B.

[90] The probability in which an individual/individuals could be identified from specific data [74, p. 26], [75, p. 24], and the likely resultant harm or impact if re-identification were to occur [75, p. 24]. The Article 29 Working Party [63, pp. 11-12] identifies three principal ways in which individuals may be re-identified, by: (1) singling-out an individual; (2) linking records relating to an individual; and (3) inferring information concerning an individual.

pseudonymisation and re-identification that are outlined in **Appendix B** to this report.

Proposed solutions to these data processing risks need to be pragmatic, in particular to mitigate the likelihood of any harm befalling stakeholders to a safe level. We therefore accepted that – while legal risk cannot be excluded totally – a Consortium-wide strategy for dealing with EU and EU Member State law was appropriate. The UK was used as an exemplar for this strategy, as it was the residing place of two members of the Consortium (i.e. the University of Southampton and the ODI) and the authors of this toolkit. This strategy was therefore designed to cover key areas of concern that might arise around the sharing and re-use of data.

The importance of demonstrating legal compliance

It is therefore essential that organisations involved with an open innovation programme – including the Consortium Members, Data Providers and Participating SMEs – retain an audit trail of steps taken to ensure legal and ethical compliance. This audit trail should demonstrate that:

- The different areas of the law relevant to the open innovation programme and the data sharing, management and re-usage activities have been considered and properly addressed across the lifetime of the programme.
- Such legal analyses are periodically reviewed to ensure they remain accurate, and are revised where there has been a material change in facts.

This audit trail is important for both internal and external accountability. First, an audit trail can be used to show that all those involved with an open innovation programme are practising high standards for data governance. Second, this audit trail is important for external accountability, as regulatory authorities will give due weight to any recorded compliance by organisations, including authoritative guidance embedded within the organisation, when deciding whether there has been a breach of the law. This audit trail can also have an impact on what level of sanctions to impose as a result of a breach of data protection laws. Organisations therefore need accurate record-keeping to prove that legal compliance was at least addressed notwithstanding that a breach still occurred. This is also particularly relevant as many data protection laws require risk assessments to be carried out and any identified risks managed as part of legal/regulatory compliance duties that each organisation bears when it processes data (the so-called ‘accountability’ principle).

7.1.3 Brief outline of the Legal and Privacy Toolkit (“the toolkit”)

The principal purpose of the toolkit is to offer guidance on the key legal and privacy aspects of data sharing, management and usage of (closed) data for a variety of innovative purposes that can be understood by non-legal specialists. This guidance is given in order to ensure that all data-related activities as part of the Data Pitch programme remained lawful and ethical. Beyond the Data Pitch programme, it is anticipated that the toolkit acts as a springboard to push forward similar data sharing schemes in Europe and elsewhere.

Audience for the toolkit

The toolkit was designed to provide practical advice for all non-Consortium organisations involved in the Data Pitch programme – i.e. Data Providers and Participating SMEs – as well as other external persons and organisations interested in guidance on how to share and re-use (closed) data lawfully and ethically.

Specific-focus on privacy and data protection by the toolkit

Although this report applies generally to the legal and regulatory issues raised by data sharing and re-use, it focuses on the legal issues raised under the Data Pitch programme. It emphasises some of the particularities of the challenges associated with the programme and considers the legal issues with these in mind. Privacy and data protection is a key focus for the toolkit – e.g. (i) it sets out the legal framework, and reflects on best practice guidance for anonymisation or pseudonymisation of personal data - before such data are shared by Data Providers or re-used by Participating SMEs; and (ii) ensuring data are re-used appropriately and lawfully as part of open innovation activities carried out by Participating SMEs.

Building on existing good governance practices

It is recognised that the legal and privacy issues that emerged under the Data Pitch programme – and which arise in similar data sharing schemes – do not occur for organisations in a vacuum. It is commonplace for larger organisations – as well as small organisations, such as start-ups – to have pre-existing risk assessment and management strategies for their data-related activities. Such strategies may range from data protection and privacy governance frameworks to more detailed governance and management structures, e.g. which focus on information architecture, data accuracy, security, and regulatory compliance. It is therefore crucial to note that the intention of this toolkit is to build on such existing good data governance practices, rather than replace them. The guidance provided by the toolkit therefore will help organisations sharing and re-using data to: (i) think through a methodology of best practice data handling; and (ii) ask appropriate questions at each stage of their data sharing scheme as relevant to their role.

Notwithstanding the focus on compliance for the Data Pitch programme, the purpose of this toolkit is also a generic one to provide guidelines on data sharing that raise similar sorts of legal challenges in practice. However, it is not intended to be a substitute for obtaining formal legal advice. Case-by-case assessments of the general principles of law and regulation set out in this toolkit will need to be assessed by the Data Providers and Participating SMEs with the Consortium providing support.

Relation of the toolkit to other parts of the Data Pitch Consortium strategy for handling data processing issues under the programme

The toolkit complemented other parts of the Data Pitch Consortium strategy for handling data processing issues under the programme. Principally, these issues were addressed by the following:

- **Signed contracts.** All organisations formally taking part in the Data Pitch programme – i.e. Data Providers and Participating SMEs – signed a contract. For further information on the Data Pitch contractual portfolio (including its three key contractual templates) see: **Data Legality Report v1** [60] **and Data Legality Report v2** [59] available via the Data Pitch website at: https://datapitch.eu/deliverables/.
- **Ethical declarations.** Participating SMEs signed a **Declaration of Honour** and an **Ethics Statement**. Copies of these documents are located in **Appendix C** to this report.
- **Oversight provided by the Consortium.** The Data Pitch Consortium supported all organisations formally taking part in the programme to interpret and instil best practices throughout their involvement. Measures carried out under contractual obligations were intended to be performed by organisations in a spirit of commitment to strong internal checks, to prevent shared data from being used in inappropriate ways. These included the imposition of best practice measures as well as the imposition of strong protection against privacy and other legal concerns. Where Data Providers proposed to supply data that had been subject to pseudonymisation processes ahead of re-use by the Participating SMEs, the Data Pitch Consortium were able to oversee and recommend the implementation of best practice safeguards on a case-by-case basis. Furthermore, the Data Pitch Consortium required Data Providers to complete a **Data Provider Questionnaire**, and Participating SMEs to complete a **Record of Information about Self-Supplied Data** in order to identify any potential risks from the outset. Copies of these documents are located in **Appendix C** to this report.
- **Training provided by the Consortium.** In order to promote legal and ethical awareness, training related to the toolkit was provided to Participating SMEs via workshops and webinar. Feedback from this training has helped to enrich the development of the toolkit by revealing any practical gaps that require further attention as well as ensuring that the toolkit is user-friendly. The provision other more interactive forms of dissemination is important for further engagement with toolkit users. Therefore, we developed a prototype e-learning tool on data protection and the basics of mapping data flows that comprises of seven interactive legal decision-trees trees[91] in order to better-communicate some of the key aspects of the GDPR in a simple way to Participating SMEs. See **Appendix B** of this report for further information on the prototype e-

[91] Note: Legal Decision Trees 1, 2 and 3 are located in Appendix B to this report, and Legal Decision Trees 4, 5, 6 and 7 are found in the main body of the D3.9 report. Refer to list of figures and tables on p. 8 of this report.

learning tool.

For a summary of the *"lessons learned, resources and recommendations for sharing data"* [61, p. 1] as part of the Data Pitch programme – taken from a wide-range of views (e.g. financial, technical, business) – see the **Data Sharing Toolkit** available via the Data Pitch website at: https://datapitch.eu/datasharingtoolkit/.

Configuration of the toolkit

Given the nature of the toolkit and the fact that the legal landscape is constantly evolving, the toolkit was updated periodically over the course of the Data Pitch programme. Each deliverable report was published incrementally on the Data Pitch website: http://datapitch.eu/. This interactive format reflected the anticipated emerging demands for guidance in specific areas in association with the key features of this toolkit, especially from Participating SMEs.

In total, a series of three toolkit reports were delivered during the course of the Data Pitch programme:

(1) Deliverable 3.1. Legal and Privacy Toolkit v1 (delivered in June 2017).
(2) Deliverable 3.5. Legal and Privacy Toolkit v2 (delivered in June 2018).
(3) Deliverable 3.9. Transnational, Cross-Sector Data Sharing in Open Innovation (delivered in December 2019).

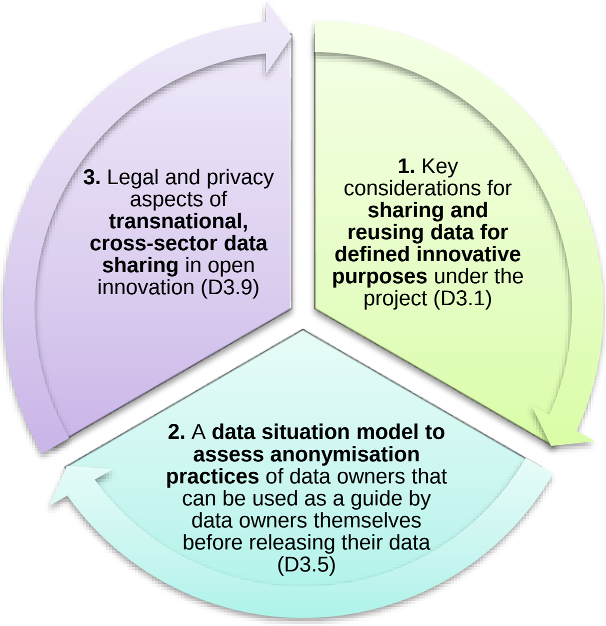

1. Key considerations for **sharing and reusing data for defined innovative purposes** under the project (D3.1)
2. A **data situation model to assess anonymisation practices** of data owners that can be used as a guide by data owners themselves before releasing their data (D3.5)
3. Legal and privacy aspects of **transnational, cross-sector data sharing** in open innovation (D3.9)

Figure 9. The three focus areas for the toolkit outlined by the Data Pitch Grant Agreement

Seven legal decision-trees

A core component of the toolkit is seven legal decision-trees, which aim to communicate various key aspects of the GDPR in a simple way to non-data protection specialists (interactive versions are provided by the prototype e-learning tool):

Seven legal decision-trees created as part of the toolkit:

- **Legal Decision-Tree 1.** Determine whether the planned processing involves personal data. (D3.5)
- **Legal Decision-Tree 2.** Assess whether the planned processing is likely to be a high risk to individuals. (D3.5)
- **Legal Decision-Tree 3.** Determine the status of control of your current and planned data situations. (D3.5)
- **Legal Decision-Tree 4.** Determine whether the data sharing activity involves a transit or transit

of personal data. (D3.9)

- **Legal Decision-Tree 5.** Determine legal responsibilities and liabilities under the GDPR: controllers and processors. (D3.9)
- **Legal Decision-Tree 6.** Determine who is subject to the GDPR. (D3.9)
- **Legal Decision-Tree 7.** Determine whether the planned data processing involves automated decision-making, including profiling. (D3.9)

Final Legal and Privacy Toolkit

As aforementioned, this last toolkit report provides an opportunity to re-visit and highlight key areas covered by earlier versions, including the seven legal decision-trees. Given those involved with the Data Pitch programme should be familiar with earlier versions of the toolkit, it is important that lengthy repetitions within the main body of the D3.9 report be prevented. For ease of reading, we therefore attach the final toolkit in the form of an appendices to this report rather than integrate the text from the D3.1 and D3.5 reports within the main body. Furthermore, it is important that the D3.9 report is able to standalone as a new contribution to the toolkit, as each toolkit deliverable (i.e. D3.1, D3.5 and D3.9) is tasked with addressing a specific objective under the Grant Agreement.

In addition to **Appendix A**, the final toolkit comprises the following two parts:

- **Appendix B** – presents a summary of the key legal and privacy aspects of data sharing and (re)usage for anyone involved in a data sharing scheme to consider, first, by offering an overview of the legal framework for data sharing and (re)usage. It then offers further detail on the following focus areas: (i) the data spectrum and raising-awareness of data flow mapping as method to create data situation models; (ii) the key legal and privacy aspects of transnational, cross-sector innovation; and (iii) the Data Pitch strategies for anonymisation, pseudonymisation and re-identification risk. It concludes by offering key recommendations for the development of a legal and ethical compliance strategy for data sharing and (re)usage.
- **Appendix C** – provides copies of the following supporting documents used by the Data Pitch programme as part of its compliance strategy: (i) Data Provider Questionnaire; (ii) Record of Information about Self-Supplied Data; (iii) Declaration of Honour; and (iv) Ethics Statement. It further offers a note on toolkit dissemination.

Furthermore, a glossary of key terms used by the toolkit is available in section 8 of this report.

Disclaimer

The content of the Legal and Privacy Toolkit (including any related training resources, such as the prototype e-learning tool) does not constitute legal advice. If in doubt, you should always contact a lawyer.

7.2 Appendix B. Key legal and privacy considerations

7.2.1 Introduction

As aforementioned, the final toolkit is an opportunity to bring together the key elements of the three previous toolkit reports – and provide a concise summary of the key legal and privacy aspects of data sharing and (re)usage (raised in the course of the Data Pitch programme) for anyone involved in a data sharing scheme to consider. As well as providing an overview of the key legal and privacy aspects, the toolkit also identifies three crucial focus areas that require specific attention. In particular, that data shared as part of an open data innovation programme (i) are *or* could likely become personal data, (ii) are likely to traverse national borders and/or sectors, and (iii) are likely to be used in the development or application of (big data) analytics techniques. Furthermore, the final toolkit gathers the seven legal decision-trees[92] created during the course of the Data Pitch

[92] I.e. Legal Decision-Tree 1. Determine whether the planned processing involves personal data; Legal Decision-Tree 2. Assess whether the planned processing is likely to be a high risk to individuals; Legal Decision-Tree 3. Determine the status of control of your current and planned data situations; Legal Decision-Tree 4. Determine whether the data sharing activity involves a transit or transit of personal data;

programme into a single report.

Appendix B is divided into the following four sections: (1) overview of the legal framework for data sharing and (re)usage; (2) understanding the data spectrum, data situation models and data flow mapping; (3) transnational, cross-sector data sharing; and (4) Data Pitch strategies for anonymisation, pseudonymisation and re-identification risk.

7.2.2 Overview of the legal framework for data sharing and (re)usage

Mapping the legal framework. In order to effectively identify the different types of law that relate to a planned data sharing and/or (re)usage activity, it is crucial to possess an adequate understanding of the legal framework for data sharing and (re)usage.

Data can be a complex subject matter in legal and ethical terms. In particular, while data cannot be owned in the same way that, say, a house or car is owned, extensive rights and obligations can arise in relation to data. For example, there are important legal principles about how data can be shared by those with rights in it with third parties who may in turn repurpose that data for different uses. At the same time, some areas of law which map such rights and obligations are developing rapidly, and are likely to develop even more quickly as big data analytical techniques become more prevalent.

The laws applicable to data are numerous. They are also scattered across a very large number of legislative instruments, and may be set out expressly or implied. In addition, there is so-called 'common law' developed by judges in courts in countries such as England through case law, contrasting with a system of 'civil law' commonly found in other European countries derived from the interpretation of codified statutes. EU laws, by comparison, come in many different forms. For example, they can be divided into 'primary' and 'secondary' legislation. The treaties (primary legislation) are the basis or ground rules for all EU action. Secondary legislation – which includes regulations, directives and decisions – are derived from the principles and objectives set out in the treaties.

The legal framework for data sharing, management and (re)usage is complex and multi-layered, principally, in that:

- **Different types of law act concurrently in relation to a data sharing arrangement**, such as:
    - **Data protection laws** – set the rules for processing personal data.
    - **Electronic privacy laws** – govern the protection of privacy in the electronic communications sector.[93]
    - **Intellectual property laws –** encompass a number of different rights that may be asserted of a more proprietary type, including in association with the use of data and its expression, such as copyright, trade marks and database rights.[94]

Legal Decision-Tree 5. Determine legal responsibilities and liabilities under the GDPR: controllers and processors; Legal Decision-Tree 6. Determine who is subject to the GDPR; and Legal Decision-Tree 7. Determine whether the planned data processing involves automated decision-making, including profiling.

[93] Privacy laws independent to the GDPR also exist under EU law, notably the E-Privacy Directive [160].[93] Electronic privacy laws govern the protection of privacy in the electronic communications sector. The principal purpose of the E-Privacy Directive therefore is *"to ensure that all communications over public networks maintain respect for fundamental rights, in particular a high level of data protection and of privacy, regardless of the technology used"* [163]. It is important to note that the E-Privacy Directive is currently under review as part of the European Commission's agenda to reinforce trust and security in the Digital Single Market. On 17 January 2017, the European Commission published a draft E-Privacy Regulation (COM(2017) 10 final) [161], which is yet to be adopted [162].

[94] Note that analysing these rights in data often requires a multi-layered legal assessment as requirements for different types of IPR to exist – and related rules – can vary widely. For example, they may provide copyright protection protecting a particular expression of information (such as a literary work); have the attributes of a database to attract database rights protection (and, also, database copyright); and, in some circumstances, could give rise to trade mark or so-called 'passing off' rights. Furthermore, as well as different types of rights, another distinction across the spectrum of IPR types relates to legal enforceability of the IPR (including potentially powerful infringement remedies, from temporary and permanent injunctions to damages and account of profits). Some IPR are in theory enforceable against the world; however, others are primarily national rights that operate different in different countries, and which are enforceable in one country (though the courts of that country) but not others. This includes disparities within the EU bloc, as well as outside it. For these reasons, contractual agreement is often relied upon to clarify and confer strong, enforceable rights at least between the contracting parties, including by assigning rights of use from one party to another via licence.

   - **Competition laws** – aim to prevent anti-competitive harm that might result from commercial activities, including from the sharing of information.[95]
   - **Laws of confidentiality** – protect confidential information.[96]
   - **Contract laws –** govern the ways in which private parties can agree to work together, including in respect of data sharing agreements[97] that include certain rights and obligations regarding data usage and access. Ultimately, if terms in agreed contracts are broken, contracting parties could try to enforce such terms in a court of law.[98]

- **Legal rights and duties can differ between sectors and countries, including within the EU bloc**, such as:
   - **Horizontal legal rights and duties – EU regulations do not always guarantee absolute legal uniformity between member states.** Any national variations therefore must be considered alongside such regulations. For instance, Recital 10 of the General Data Protection Regulation (GDPR) *"provides a margin of manoeuvre for Member States to specify its rules, including for the processing of special categories of personal data ('sensitive data'). To that extent, this Regulation does not exclude Member State law that sets out the circumstances for specific processing situations, including determining more precisely the conditions under which the processing of personal data is lawful."*
   - **Vertical legal rights and duties – further sector-specific legal rights and duties may arise where data sharing, management and usage takes place within or across certain industries.** Data regulation is also deepening in many vertical industry sectors. This is not necessarily a novel development; for example, the rules on the confidentiality of client information and privilege have been cornerstones of the legal profession for generations. However, the digitisation of data is changing the regulatory landscape fundamentally in some sectors. Examples include the financial sector, insurance, air travel (specifically, rules on passenger name record ('PNR') data about an airline customer's itinerary), and healthcare (including rules about aggregating anonymised clinical outcome patient data). These sector-specific requirements are tending to become more intrusive as regulatory authorities obtain wider supervisory powers to obtain information, investigate business practices and conduct, and audit organisations under their charge.
   - **Regions and/or countries may implement different types of restrictions or prohibitions on transnational, cross-sector data sharing.** Such regulatory constraints on transnational, cross-sector data sharing may take the form of blanket bans, target data flows that take place within specific sectors, and/or focus on data

[95] National and EU competition authorities have shown increasing interest this decade in analysing data-centric business practices – including mergers and licenses – through the lens of competition law. Sectors investigated include electronic communications and particularly financial markets. Furthermore, EU competition law – implemented into EU Member States' domestic laws – include certain restrictions on information sharing implemented through agreements, or so-called 'concerted practices', between at least two organisations that have the object or effect of restricting competition in horizontal/vertical markets.

[96] Laws of confidentiality protect confidential information. In some cases, data with a quality of confidentiality may attract legal protection and remedies that can be enforced in national courts where such confidential information is shared without authority. In other words, rules governing the confidentiality of information exist in some countries that protect the substance of data that is not generally known-publicly. For example, there is UK case law* that suggests a right to confidentiality can exist in respect of dataset aggregation despite some of the data not being confidential, alongside legal protection extending to second and subsequent generation data derived from initially confidential data. Nevertheless, consideration should be given to whether contracts – and websites and other notices – state expressly that data should be considered confidential, and that it is not freely publicly available. Furthermore, historic data inevitably holds less value and is likely to be more widely disseminated than real-time data. [*Albert (Prince) v Strange, ([1849] 1 M&G 25); Exchange Telegraph Co. Ltd v Gregory & Co., ([1896] 1 QB 147); Exchange Telegraph Co. Ltd v Central News Ltd ([1897] 2 Ch 48); Weatherby & Sons v International Horse Agency and Exchange Ltd, ([1910] 2 Ch 297).]

[97] For more information on data sharing agreements see: Data Legality Report v1 [60] and Data Legality Report v2 [59] that provide an overview of the creation of the Data Pitch contractual portfolio, including the three key contractual templates utilised by the programme.

[98] Data suppliers therefore should ensure that contracts contain express acknowledgements to the effect that relevant rights subsist in the data and are owned by them. Whilst, data recipients should agree to take, allow, or suffer no action that is inconsistent with the rights of the data supplier under the agreement. Some common issues in data sharing and re-use legal agreements include: scope of rights in data being agreed; warranties of compliance with laws and regulation, and indemnities in the case of later non-compliance; treatment of derived and commingled data; and post-term use of data.

flows that occur as part of processes and/or services [5, p. 17], [6], [7, p. 125].

- **The legal framework is constantly changing.** Organisations sharing, managing and using data must keep up-to-date with all relevant (non-)sectoral legal amendments, repeals and enactments. For instance, in the case of Data Pitch, the GDPR – an extremely significant development in EU data protection law – entered into force midway during the programme. Given that authoritative guidance often follows on from new legal developments, such inevitable delays may lead to some uncertainty over new legal rights and duties. For instance, authoritative guidance on providing explanations on artificial intelligence (AI) decisions to data subjects from ICO and the Alan Turing Institute is anticipated at the end of 2019, as part of Project ExplAIN [58].

Understanding the data spectrum. To add further complication, the same dataset can attract different legal responsibilities, rights, liabilities and levels of risk under different sets of circumstances. For instance, personal and non-personal data are not binary concepts. The nature of data is changeable – and ultimately predicated on the specific context and purpose of a (planned) data processing activity. For instance, anonymised data could be re-identified – and the same dataset can be considered as non-personal and personal under different sets of circumstances. In consequence, an effective identification of the different types of law that relate to a planned data sharing and/or (re)usage activity necessitates not only an adequate understanding of the legal framework, but its application to the specific context and purpose of the planned data activity in question. The data spectrum is now explored in more detail in the following section.

7.2.3 Understanding the data spectrum, data situation models and data flow mapping

Note: The following D3.5 Legal and Privacy Toolkit v2 report was delivered in June 2018. Co-ordinators: Professor Sophie Stalla-Bourdillon and Dr Laura Carmichael. With contributions from: Dr Pei Zhang (Developer). Quality reviewer: Open Data Institute (ODI). Appendices of original report omitted – see original document for full details.

Abstract

The Legal and Privacy Toolkit v2 is conceived as a supplement to the Legal and Privacy Toolkit v1 (available at: https://datapitch.eu/privacytoolkit/), which extends the data protection guidance provided in the first version of the toolkit. The objective is to provide data owners with guidance on the creation of data situation models to be used as part of anonymisation assessment. In particular, it aims to raise-awareness of data flow mapping as an effective and pragmatic approach to the creation of data situation models – and compliance with the General Data Protection Regulation (GDPR). It offers practical guidance that can be understood by non-data protection specialists through devised training materials: (i) three legal decision-trees – an interactive version of these is provided through a prototype e-learning tool; and (ii) a workshop. Part A focuses on some of the key data protection considerations that underpin the robust assessment of anonymisation practices. Part B focuses on the basic components of mapping data flows.

Disclaimer: The content of the Legal and Privacy Toolkit does not constitute legal advice. If in doubt, you should always contact a lawyer.

Executive summary

D3.5 toolkit update: The Legal and Privacy Toolkit ("the toolkit") is a crucial component of the Data Pitch programme. In June 2017, the first version of the toolkit was published on the Data Pitch website: https://datapitch.eu/privacytoolkit/. The toolkit covers a wide-range of key legal considerations that are likely to occur in the course of an open innovation programme – from contractual obligations to intellectual property rights. The Legal and Privacy Toolkit v2 (D3.5) is a toolkit update that extends the data protection guidance provided in the first version of the toolkit.

Objective: In accordance with the Grant Agreement, the central focus of Data Pitch Deliverable 3.5 is as follows:

"A data situation model to assess anonymisation practices of data owners that can be used as a

guide by data owners themselves before releasing their data."

Definitions: For the purpose of this deliverable, the terms used by the Grant Agreement objective are defined as follows:

A data owner: An individual and/or organisation who is involved with, and therefore able to make decisions about: the collection, management, release and/or (re)usage of a dataset. E.g. a data provider who releases data within an open innovation programme or an SME who re-uses these data.

Note: this term is widely used in ICT and business sectors – it is not a legal term.

A data situation model: A representation of a (particular version of a) dataset and the relationships with its various environments – past, current and planned – that can be used as a basis for anonymisation assessment.

Note: "data situation" is a term utilised by the UK Anonymisation Network (UKAN) see: [62, p. 130].

An anonymisation practice: A technical and/or organisational method employed to de-identify a dataset (i.e. remove personally identifiable information from a dataset) [63, p. 11]. There are two main approaches to anonymisation [63, pp. 11-19]: (i) *"randomization techniques"* – e.g. *"noise addition"*, *"permutation"* and *"differential privacy"* and (ii) "*generalization techniques*" – e.g. *"aggregation and K-anonymity"* and *"L-diversity and T-Closeness"*.

Note: this term is not to be confused with the higher benchmark set by the legal standard for anonymisation (see definition below).

Assessment of anonymisation practices: A process of evaluation undertaken by data owners to ensure their past, current and planned data practices and activities are legally and ethically compliant.

Legal standard for anonymisation: *"[…] personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable […]"* Source: Recital 26 of the General Data Protection Regulation [12].

The Legal and Privacy Toolkit v2 focuses on data mapping: It aims to equip data owners with guidance on the basics of mapping data flows. This is because data flow mapping is an effective and practical approach to the creation of data situation models. For the purpose of this deliverable, data flow mapping is defined as follows:

Data flow mapping: The creation of a graphical representation that charts the actual and potential movement of a (particular version of) a dataset as it is collected, managed, shared and (re)used across various data environments. An approach employed for the creation of data situation models.

The Legal and Privacy Toolkit v2 is therefore conceived as a supplement to the Legal and Privacy Toolkit v1, which aims to raise-awareness of the important role that data flow mapping can play in responsible data sharing and re-usage within open innovation programmes. Furthermore, it offers practical guidance that can be understood by non-data protection specialists through devised training materials. In particular, data flow mapping is presented as a way in which data owners can demonstrate compliance with the General Data Protection Regulation (GDPR). The Legal and Privacy Toolkit v2 is divided into three parts:

Part A – Understanding the data spectrum: A data flow map is useless for data protection compliance without prior understanding of how personal data is defined by the GDPR. Therefore, effective data flow mapping depends on a solid understanding of (at minimum) the following key data protection considerations:

1. What types of data are likely to fall within and outside the scope of the GDPR.
2. What types of data processing activities are considered as high-risk under the GDPR.
3. What types and levels of measures are required to control the flow of data.

Guidance is therefore given on these three areas of key data protection considerations by: (i) an overview of existing, authoritative guidance; and, (ii) the creation of three legal decision-trees that aim to help raise-awareness.

Part B – The basics of mapping data flows: The Legal and Privacy Toolkit v2 introduces the following key elements of mapping data flows:

- Data flow mapping can be used at enterprise-level and/or dataset-level.
- Data flow mapping has numerous benefits, including gap-identification and risk mitigation.
- The content of a data flow map is most crucial – not the format it takes.
- An effective data flow map will take into consideration both the technical and organisational aspects pertaining to a particular data situation.
- Access to robust provenance information is an advantage for mapping data flow activities.

It further provides three fictional scenarios that data owners can utilise to create data situation models through mapping data flows.

Part C – The development of training materials: Work carried out in Parts A and B – in particular the three legal decision-trees and three fictional scenarios for data flow mapping – led to: (i) a legal training workshop held in May 2018; and (ii) the development of a prototype e-learning tool on data protection and the basics of mapping data flows. The current version of this e-learning tool comprises a series of three interactive legal decision trees (produced in Part A). The workshop hand-out and screen-shots from the interactive e-learning tool are available as annexes to this supplementary report.

Conclusions: The Legal and Privacy Toolkit v2 concludes with two key points for data owners to take forward when (i) assessing their anonymisation practices and (ii) building data situation models to guide these practices before data are released and/or re-used. An effective assessment of anonymisation practices requires understanding of the following:

1. **Context and purpose.** In order to determine whether a particular planned data processing activity is personal or non-personal, this decision will heavily rely on the context and purpose of the specific activity under consideration. Therefore, just because a dataset was used for non-personal purposes in the past does not mean that it cannot be utilised for personal purposes in the future.
2. **An understanding of the basics of mapping data flows.** While data flow mapping is not a panacea for GDPR-compliance, it is a useful tool to employ so that: data owners demonstrate compliance with the GDPR; gaps between the regulatory framework and how data are processed in practice are revealed; (potential) high-risk data processing activities are identified and risks can be mitigated; individuals are well-positioned to make good decisions about data processing; and, the areas that require further legal training and guidance are exposed.

Further toolkit updates: Given the nature of this toolkit and the fact that the legal landscape is constantly evolving, this toolkit (and its final version due for publication in 2019) will be updated periodically at https://datapitch.eu/privacytoolkit/. This interactive format reflects the anticipated emerging demands for guidance in specific areas in association with the key features of this toolkit.

Introduction

D3.5: Legal and Privacy Toolkit Update

The Legal and Privacy Toolkit ("the toolkit")[99] [3] is a crucial component of the Data Pitch

[99] The principal focus of this toolkit [3] is to ensure that all those involved with Data Pitch: (a) are made aware of the key legal rights that arise in relation to data sharing as part of the programme; and, therefore (b) adhere to any legal obligations that concern these data sharing activities. This legal guidance is achieved through the provision of an overview concerning the legal and regulatory framework that applies to data sharing and data reuse. This overview: (i) sets out the key considerations that govern the data sharing arrangements between the parties involved in the programme; (ii) maps the relevant legal issues that arise in the context of data sharing; and (iii) outlines a methodology for handling these legal issues in a suitably risk-averse manner. This framework aims to treats data ethically and responsibly, with comprehensive, yet pragmatic guidance on data disclosure and its handling.

programme.[100] In June 2017, the first version of the toolkit was published on the Data Pitch website: https://datapitch.eu/privacytoolkit/. The toolkit covers a wide-range of key legal considerations that are likely to occur in the course of an open innovation programme[101] – from contractual obligations to intellectual property rights. The Legal and Privacy Toolkit v2 is a toolkit update that extends the data protection guidance provided in the first version of the toolkit (at the half-way point – M18 – of the programme).

Objective

In accordance with the Grant Agreement, the central focus of Data Pitch Deliverable 3.5 is as follows:

"A data situation model to assess anonymisation practices of data owners that can be used as a guide by data owners themselves before releasing their data." [Underlining added for emphasis.]

Objective definitions

For the purposes of the deliverable, the terms used by the Grant Agreement objective are defined as follows:

A data owner: An individual and/or organisation who is involved with, and therefore able to make decisions about: the collection, management, release and/or (re)usage of a dataset. E.g. a data provider who releases data within an open innovation programme or an SME who re-uses these data.

Note: this term is widely used in ICT and business sectors – it is not a legal term.[102]

A data situation model: A representation of a (particular version of a) dataset and the relationships with its various environments – past, current and planned – that can be used as a basis for anonymisation assessment.

Note: "data situation" is a term utilised by the UK Anonymisation Network (UKAN) see: Mark Elliot et al. The Anonymisation Decision-Making Framework (2016), p. 130 <http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf> [last accessed 24 May 2018].

A pseudonymisation practice: A technical and/or organisational method employed to de-identify a dataset by replacing an attribute with another attribute [63, p. 20]. Examples of pseudonymous practices include [63, pp. 20-21]: (i) *"encryption with secret key"*; (ii) *"hash function"*; (iii) *"keyed-hash function with stored key"*; (iv) *"deterministic encryption of keyed-hash function with deletion of the key"*; and (v) *"tokenization"*.[103]

Note: pseudonymisation practices are distinct from anonymisation practices (see definition below).

An anonymisation practice: A technical and/or organisational method employed to de-identify a dataset (i.e. remove personally identifiable information from a dataset) [63, p. 11]. There are two main approaches to anonymisation [63, pp. 11-19]: (i) *"randomization techniques"* – e.g. *"noise addition"*, *"permutation"* and *"differential privacy"* and (ii) "*generalization techniques*" – e.g. *"aggregation and K-anonymity"* and *"L-diversity and T-Closeness"*.[104]

[100] The principal motivation for the Data Pitch programme is to support successful applicants (i.e. start-ups) with their high-impact, innovative and data-centric business ideas, products and services that directly respond to the specific challenges defined by the programme [122]. This support is given in numerous forms, mentoring and training services to financial assistance and rights to re-use valuable data that would otherwise remain inaccessible i.e. closed data. For more information about closed, shared and open datasets – the data spectrum – see [117].

[101] Open innovation acceleration programmes strive for the development of high impact, cutting-edge products and services. In order to bring these innovative ideas to fruition, participants are often required to share and re-use data. For further information on innovation accelerators see [110] – and the European Commission (EC) online resources on open innovation [118].

[102] The term "data ownership" is commonly used by those in ICT and business sectors to refer to: *"the de facto holder of data, and [who] can therefore decide on the use and trade of these data"* [165, p. 760]. Note that the terms "data holder" or "data steward" may also be used instead of data owner. Data Pitch does not therefore utilise the phrase "data owners" in *"the sense of legal property"* [165, p. 760], but to cover the range of roles involved with the sharing and (re)usage of data in the course of the programme. For further background information on the role of data ownership within the business sector – see the following articles by the Data Governance Institute: [166] and [167]. For legal analysis on whether "big data" requires data ownership rules as a new data property right – see [168].

[103] See [63, pp. 20-21] for more information on these examples of these pseudonymisation practices.

[104] See the Article 29 Working Party's Opinion 05/2014 on Anonymisation Techniques [63, pp. 11-19] for more information on these examples of anonymisation practices. For analysis of how this Opinion on Anonymisation Techniques relates to the GDPR see: [169].

Note: this term is <u>*not*</u> *to be confused with the higher benchmark set by the legal standard for anonymisation (see definition below).*[105]

Assessment of anonymisation practices: A process of evaluation undertaken by data owners to ensure their past, current and planned data practices and activities are legally and ethically compliant.

Legal standard for anonymisation: *"[…] personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable […]"* Source: Recital 26 of the General Data Protection Regulation [12].

Overview: data situation models and data flow mapping

It is anticipated that a robust data situation model will better-position data owners to make effective decisions about their planned data processing activities – a notion that is at the heart of the UKAN Anonymisation Decision-Making Framework. The Anonymisation Decision-making Framework (ADF) is a "practical guide to anonymisation" [62, p. xii] that is principally intended for use by those involved with the sharing and (re)usage of personal data and anonymised data.[106] Through the process of creating a data situation model, data owners should have an improved understanding of (i) the overall context and purpose of the planned data processing activity under consideration and therefore (ii) be better-placed to establish a specially-devised plan for appropriate anonymisation that ensures any planned data processing activity is both legally and ethically compliant [62, pp. 68-69].

The Legal and Privacy Toolkit v2 focuses **on data mapping as an effective and practical approach to the creation of data situation models**, as Mark Elliot et al. [62, p. 69] state in the UKAN Anonymisation Decision-Making Framework: *"by mapping the data flow from the point at which data is collected to the point after which it is shared or released you will be able to define the parameters of your data situation."*

Define: data flow mapping

For the purpose of this deliverable, data flow mapping is defined as follows:

Data flow mapping: The creation of a graphical representation that charts the actual and potential movement of a (particular version of) a dataset as it is collected, managed, shared and (re)used across various data environments. An approach employed for the creation of data situation models.

Some key benefits of data flow mapping:

Demonstrate compliance [64, p. 6] – Data flow mapping can help data owners to adhere to the obligation under Recital 82 of the GDPR that requires a controller or processor to *"maintain records of processing activities under its responsibility"*.[107]

Reveal gaps [65, p. 8], [66, p. 7] – Data flow mapping can help to highlight any gaps between the regulatory framework with how data are collected, managed, shared and (re)used in practice.

Risk mitigation [65, p. 8] – It can draw attention to (potential) high-risk data processing activities before data are shared and re-used within an open innovation environment. It can further help to identify the appropriate technical and organisational measures required to assist with

[105] In common parlance, the term "anonymise" means: *"[r]emove identifying particulars or details from (something, especially medical test results) for statistical or other purposes"* [170]. However, as James Clark [78, p. 10] states: "*For many organisations which aren't deeply versed in these issues, there is often a gap between on the one hand, the popular notion of what constitutes anonymisation, and its legal meaning in the context of data protection on the other. Inevitably, the latter represents a higher bar."*

[106] Note that the original UK version of the Anonymisation Decision-Making Framework (ADF) has been adapted for the Australian context*: "The De-identification Decision-Making Framework"* [171], [172].

[107] Furthermore, note that under certain circumstances, Article 30 of the GDPR identifies the types of information that are required in a record of processing activities for both controllers (see Article 31(1)) and processors (see Article 32(2)). Article 30(5) of the GDPR states: *"The obligations referred to in paragraphs 1 and 2 shall not apply to an enterprise or an organisation employing fewer than 250 persons unless the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects, the processing is not occasional, or the processing includes special categories of data as referred to in Article 9(1) or personal data relating to criminal convictions and offences referred to in Article 10."*

the desired level of control over a dataset.

Robust decision-making. It can provide a knowledge-base for robust decision-making about if and how best to share and re-use data.

Legal training [65, p. 8] – By understanding where the gaps between practice and the regulatory framework lie, open acceleration programmes can better-target the areas that require further legal training.

Enhanced guidance

The Anonymisation Decision-making Framework covers a broad-range of *"core anonymisation activities"* that all relate to the assessment of anonymisation practices: (1) *"a data situation audit"*; (2) *"risk analysis and control"*; and, (3) *"impact management"* [62, p. 67]. The objective of this report falls under the first type of core anonymisation activities – (1) *"a data situation audit"*. Furthermore, a data situation audit is comprised of five components: (a) *"describe your data situation"*; (b) *"understand your legal responsibilities"*; (c) *"know your data"*; (d) *"understand the use case"*; and, (e) *"meet your ethical obligations"* [62, p. 67].

In the opinion of IT Governance [67], the three most challenging issues that arise for those who are mapping data flows all relate to legal understanding: (i) *"identifying personal data"*; (ii) *"identifying appropriate technical and organisational safeguards"*; and (iii) *"understanding legal and regulatory obligations".* The Legal and Privacy Toolkit v2 therefore has a much narrower focus than the UKAN Anonymisation Decision-Making Framework, and as such aims to provide further detail on the creation of data situation models for anonymisation assessment by: (i) enhanced focused on key data protection considerations that arise under the GDPR – and are most challenging for data flow mapping; and, (ii) enriched guidance on the basics of mapping data flows.

Report overview

The Legal and Privacy Toolkit v2 is therefore conceived as a supplement to the Legal and Privacy Toolkit v1, which aims to raise-awareness of the important role that data flow mapping can play in responsible data sharing and re-usage within open innovation programmes. The Legal and Privacy Toolkit v2 aims to equip data owners with guidance on the basics of this approach.

Furthermore, it offers practical guidance that can be understood by non-data protection specialists through devised training materials. In particular, data flow mapping is presented as a way in which data owners can demonstrate compliance with the General Data Protection Regulation (GDPR) 2016/679 [12].[108]

The Legal and Privacy Toolkit v2 is divided into the following parts:

- **Part A – Understanding the data spectrum.** A data owner will only be able to assess their anonymisation practices effectively when they have sufficient knowledge of their legal obligations under the GDPR. Therefore, Part A therefore focuses on some of the key data protection considerations that underpin useful assessment of anonymisation practices:

1. What types of data are likely to fall within and outside the scope of the GDPR.
2. What types of data processing activities are considered as high-risk under the GDPR.
3. What types and levels of measures are required to control the flow of data.

Guidance is therefore given on these three areas of key data protection considerations by: (i) an overview of existing, authoritative guidance; and, (ii) the creation of three legal decision-trees that aim to help raise-awareness by communicating these three key aspects of the GDPR in a simple way to data owners.

[108] This guidance is very timely as the GDPR entered into force on 25 May 2018 – during the deliverable D3.5.

- **Part B – The basics of mapping data flows.** Part B outlines the basics of mapping data flows – as a useful approach to the creation of data situation models for GDPR-compliance – which data owners can take forward as part of their anonymisation assessment practices before data are released and/or re-used.

- **Part C – The development of training materials.** Part C explains how Parts A and B led to the development of further legal training materials: (i) a prototype e-learning tool on data protection and the basics of mapping data flows; and, (ii) a workshop – delivered to invited SMEs in May 2018.

This supplementary report then concludes by: (i) summarising the key points from Parts A-C; and (ii) outlining areas for future work.

Part A: Understanding the data spectrum

Brief overview

It is crucial that all those involved with data sharing and re-usage within open acceleration programmes act responsibly by remaining compliant with all applicable legal and ethical obligations (from contractual obligations to intellectual property rights). Non-compliance can have severe consequences – such as litigation and reputational damage.

In order to make sure that data flow mapping is effective, it is important to first outline some of the key data protection considerations that must be considered as part of an anonymisation assessment. Guidance is given to data owners on the following three areas in Part A:

1. What types of data are likely to fall within and outside the scope of the GDPR.
2. What types of data processing activities are considered as high-risk under the GDPR.
3. What types and levels of measures are required to control the flow of data.

This guidance is achieved through review of existing and authoritative guidance, in particular that of the Information Commissioner's Office (ICO)[109] and the Article 29 Working Party.[110] Furthermore, Part A follows the development of three paper-based legal decision-trees (produced by the report authors based on authoritative guidance) that cover these three main areas for consideration.[111] The decision-tree technique is employed as a learning device to further communicate some of the key aspects of the GDPR in a simple way to data owners by representing a key series of concepts and their outcomes.

Disclaimer: important notice to readers

Disclaimer: The content of the Legal and Privacy Toolkit (including any updates) does not constitute legal advice. If in doubt, you should always contact a lawyer.

Define: personal and non-personal data

Context and purpose

Personal data and non-personal data are not binary concepts – data exist on a spectrum.[112] For

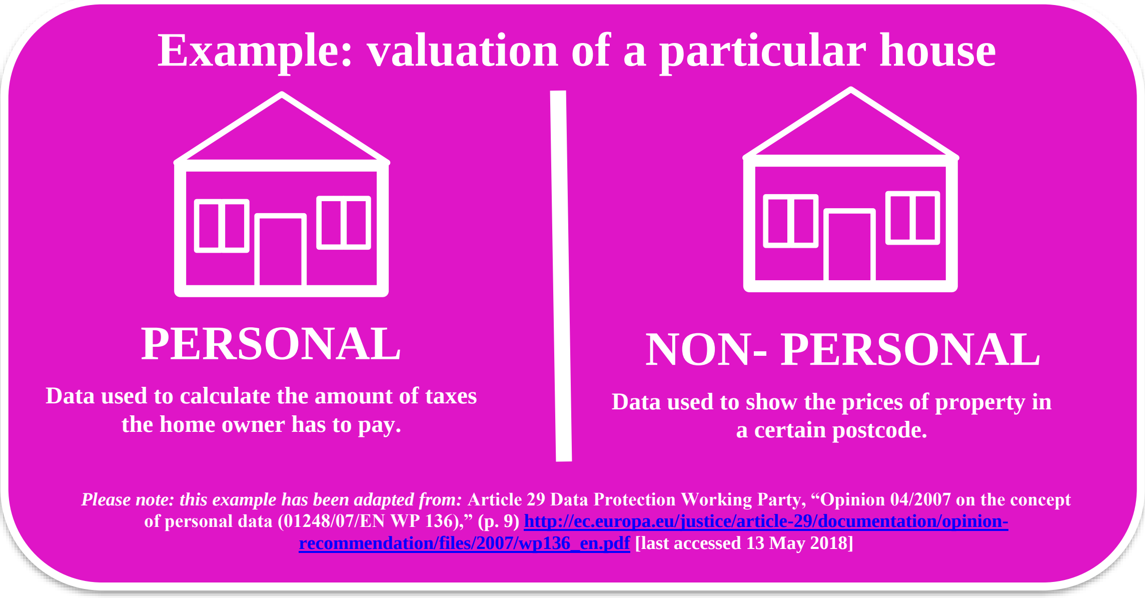
[109] ICO is *"the UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals"* [173].

[110] The Article 29 Working Party is an independent body that provides impartial advice to the European Commission on data protection and aims to facilitate policy harmonisation across the EU member states [174]. For further background information about the Article 29 Working Party see: Article 29 of the Data Protection Directive [175] and the Article 29 Working Party Newsroom [176].

[111] An interactive version of this series of three legal decision-trees has also been created, see Part C of this report for further information.

[112] *"Privacy has traditionally worked along a spectrum that's context dependent. Is it personal data? Well, that depends."* [177, p. 2]. For instance, Boris Lubarsky [178] represents data identifiability as a staircase. For more information about the identifiabillity data spectrum see: [68], [179] and [180]. Note: this report utilises the term "data spectrum" in the context of data protection, this usage should not be confused with other uses of the term, e.g. the ODI's data spectrum [117] focuses on the level of access to data, i.e. closed, shared and open.

instance, one use of a particular dataset could be personal, but another use of the exact same dataset could be non-personal.



Example: valuation of a particular house

PERSONAL

Data used to calculate the amount of taxes the home owner has to pay.

NON- PERSONAL

Data used to show the prices of property in a certain postcode.

Please note: this example has been adapted from: Article 29 Data Protection Working Party, "Opinion 04/2007 on the concept of personal data (01248/07/EN WP 136)," (p. 9) http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf [last accessed 13 May 2018]

Figure 10 Data spectrum diagram - example: valuation of a particular house

Furthermore, there is a *"fluid line"* [68, p. 38] between properly anonymised data (for further information see section 2.2.3 of this report) and personal data as: *"anonymized data can always become personal data again depending upon the evolution of the data environment"* [68, p. 38].

Therefore, the context and purpose of each planned data processing activity (e.g. sharing data with a third party) determines whether data fall under or outside the scope of the GDPR. In the words of ICO guidance [69, p. 14]: *"It is important to remember that the same piece of data may be personal data in one party's hands while it may not be personal data in another party's hands."* Furthermore, Mark Elliot et al. [62, p. 75] state: *"A person is identifiable where the conditions exist to identify them. […] we consider whether a person is identifiable or not to be heavily contextualised."*

Does this dataset fall inside the scope of the GDPR?

Legal definition of personal data

In order to assess whether a planned data processing activity (e.g. a data release or re-usage) of a specific dataset falls under the GDPR, the data owner must first refer to the legal definition of personal data:

Legal definition of personal data:
"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"
Source: Article 4(1) of the General Data Protection Regulation (GDPR) [12]

A fundamental aspect of this legal definition is that information relates to an identified or identifiable person i.e. a data subject.

How can data relate to individuals?

Data can relate to an individual in a number of ways. In most instances, a name combined with further information (e.g. address or telephone number) will be an adequate amount of information to identify a person [69, p. 7]. Nonetheless, it is important to note the following: (1) a name will not always be sufficient to distinguish a person from other members of a group, e.g. if two members of a group have exactly the same name [69, p. 7]; and (2) a person can be identified without reference to a name [69, p. 8].

The Information Commissioner's Office (ICO) [70, pp. 4-6] outlines the six most common ways in which data relate to an individual as follows (see original document for full information):

ICO six commons ways data relate to individuals:

1. Data are *"obviously about a particular individual"*.
2. Data are *"linked to an individual"*.
3. Data are *"used […] to inform and/or influence actions and decisions affecting an identifiable individual".*
4. Data have *"biographical significance"*.
5. Data *"focus or concentrate on the individual as its central theme".*
6. Data *"have the potential to impact on an individual".*

Source: ICO guidance [70, pp. 4-6]

It is obvious

In some cases, it is extremely obvious that a dataset relates to a data subject, because that specific dataset is about individuals [71, p. 9], [70, p. 4]. In other words, individuals are the unmistakable *"'focus' of the information"* [70, p. 5] as *"the data units are people"* [62, p. 9]. The Article 29 Working Party [71, p. 9] offers three illustrative examples of where data are clearly about individuals: *"the data registered in one's individual file in the personnel office […] the data on the results of a patient's medical test contained in his medical records, or the image of a person filmed on a video interview of that person."* The Information Commissioner's Office (ICO) [70, p. 4] provides four further examples of data that obviously relate to a data subject: *"medical history, criminal record, record of […] work or […] achievements in a sporting activity."*

In consequence, the first step towards an assessment over whether a planned data processing involves personal data is the consideration of the following question:

Legal decision-tree 1: determine whether the planned processing involves personal data
Step 1 of 4: Consider the focus of the data.
Q: Is it obvious that the data you intend to process are about individuals (i.e. the data units are people)?
A: Yes/No/I do not know

Identifiability and identified persons

The following legal decision-tree question has been adapted (by the report authors) from the following definition of anonymous data: *"information that does not identify any individuals, either in isolation or when cross referenced with other data already in the public domain"* [72].

Legal decision-tree 1: determine whether the planned processing involves personal data
Step 2 of 4: Assess whether individuals are identifiable from the data you intend to process.
(i) Identifiability in the immediate processing environment:
Q: Can you identify individual(s) from the data you intend to process in isolation (i.e. without reference to any other data and/or information)?
A: Yes/No/I do not know

The next legal decision-tree question has been adapted (by the report authors) from a question posed by ICO [69, p. 7] in its guidance on determining what is personal data: *"Can a living individual be identified from the data, or, from the data and other information in the possession of, or likely to come into the possession of, the data controller?"*

Legal decision-tree 1: determine whether the planned processing involves personal data

Step 2 of 4: Assess whether individuals are identifiable from the data you intend to process.
(i) Identifiability in the immediate processing environment:
Q: Can you identify individual(s) from the data you intend to process in isolation (i.e. without reference to any other data and/or information)?
A: Yes/No/I do not know

Again, the following legal decision-tree question has been adapted (by the report authors) from the following definition of anonymous data: *"information that does not identify any individuals, either in isolation or when cross referenced with other data already in the public domain"* [72].

Legal decision-tree 1: determine whether the planned processing involves personal data
Step 2 of 4: Assess whether individuals are identifiable from the data you intend to process.
(iii) Identifiability in the public domain:
Q: Is there more than a remote possibility that you can identify individuals(s) from these data when cross referenced with other data and/or information which is: (a) already in the public domain; and (b) likely to come into the public domain?
A: Yes/No/I do not know

It is not so obvious: content, purpose and result

In other cases, it is less obvious whether a dataset relates to a data subject, where at first glance it may appear that a particular dataset is not about individuals i.e. where the data units are objects, processes, events or other non-people entities [71, p. 9]. In other words, a dataset does not have to relate to individuals through content, but by the purpose or the result of a particular processing activity.

The Article 29 Working Party [71, p. 11] offers the following example of personal data by purpose (paraphrased from original document): a telephone call log could be used by a company to provide information about the number of callers or to learn something about the employee responsible for that phone line. The following legal decision-tree question therefore covers the purpose element of how data may relate to an individual. It has been adapted from the description of the purpose element provided by the Article 29 Data Protection Working Party Opinion (04/2007) on the concept of personal data [71, p. 10]:

> *"[…] a **"purpose"** element can be responsible for the fact that information "relates" to a certain person. That "purpose" element can be considered to exist when the data are likely to be used, taking into account all the circumstances surrounding the precise case, with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual."*

Legal decision-tree 1: determine whether the planned processing involves personal data
Step 3 of 4: Consider whether the reasons behind your planned data processing relates to individuals.
Q: Does the reason for carrying out the planned data process relate to (at least) one of the following: (a) to learn about individuals; (b) to evaluate individuals; (c) to make a decision that affects individuals; (d) to treat individuals in a certain manner; and/or (e) to influence the status or behaviour of an individual?
A: Yes/No/I do not know

The Article 29 Working Party [71, p. 11] also provides the following example of personal data by result (paraphrased from original document): a taxi company could use location-monitoring data to make their service more efficient which would in turn impact on the taxi drivers. The following legal decision-tree question therefore covers the result element of how data may relate to an individual. It has been adapted from the description of the purpose element provided by the Article 29 Data Protection Working Party Opinion (04/2007) on the concept of personal data [71, p. 11]:

> *"[…] data can be considered to "relate" to an individual because their use is likely to have an*

> *impact on a certain person's rights and interests, taking into account all the circumstances surrounding the precise case. It should be noted that it is not necessary that the potential result be a major impact."*

Legal decision-tree 1: determine whether the planned processing involves personal data
Step 4 of 4: Evaluate whether the consequences of your planned data processing are likely to impact on the rights and freedoms of individuals.
Q: The outcome of this planned data processing is likely to have an impact on the rights and freedoms of individuals?
A: Yes/No/I do not know

Does this dataset fall outside the scope of the GDPR?

There are two main types of non-personal data processing activities that fall outside the scope of the GDPR: (1) data that are properly anonymised; and (2) data that are apersonal.

Data that are properly anonymised

The legal standard for anonymisation

In order to comply with data protection law and wider ethical obligations,[113] it is of paramount importance that all those involved with the sharing and (re)usage of data are fully cognisant of the legal standard for anonymisation. Data are properly anonymised when the legal standard of anonymisation is reached.

Legal standard of anonymisation:
"[…] personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable […]"
Source: Recital 26 of the General Data Protection Regulation

According to guidance provided by ICO [73], while absolute anonymity[114] is not required, the risk of re-identification must be mitigated *"until it is remote"* [73, p. 6].[115] As part of this authoritative guidance, ICO [73, p. 48] further defines anonymised data as: *"[d]ata in a form that does not identify individuals and where identification through its combination with other data is not likely to take place".* Moreover, Steve Wood [72] defines anonymous data as: *"information that does not identify any individuals, either in isolation or when cross referenced with other data already in the public domain."*

[113] For instance, Luciano Floridi and Mariarosaria Taddeo [100] consider issues relating to anonymisation, such as the sharing and (re)use of large-scale data and the potential for re-identification of individuals to be a key part of data ethics. Floridi and Taddeo [100] define "data ethics" as: *"a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)."*

[114] The position of the Article 29 Working Party is that the de-identification of personal data must be "irreversible" for data to reach the legal standard of anonymity [63, pp. 5, 7]. In other words, this view appears to support absolute anonymity, i.e. personal data are only considered as rendered anonymous where there is zero risk of data subjects being re-identified from an anonymised dataset. In practice, absolute anonymity (i.e. where there is zero risk that individuals can be identified from an anonymised dataset) is extremely difficult to achieve when you take into consideration: the utility of data, technological advancements and new data releases (e.g. open data – mosaic effect) – seen and unforeseen. Note that it appears that the Article 29 Working Party has faced some criticism in the past *"for being too conservative on data protection issues and for setting out positions that are commercially impractical"* [181].

[115] Given that absolute anonymity is impracticable, it is unsurprising that various authoritative data protection bodies disagree with the Article 29 Working Party's strict interpretation of legal standard for anonymity (see previous footnote) by favouring a more pragmatic approach. For instance, the German Data Protection Authority in Hamburg – that is *"known for its strong stance on privacy issues"* [182] – also acknowledges that absolute anonymity cannot be realised in all cases: *""German privacy law defines 'rendering anonymous' as 'the alteration of personal data so that information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense and effort'"* [182]. Furthermore, the Handbook on European Data Protection Law [183, p. 44] considers data to be anonymised as follows: *"Data are anonymised if all identifying elements have been eliminated from a set of personal data. No element may be left in the information which could, by exercising reasonable effort, serve to re-identify the person(s) concerned."* Outside the EU, the Information and Privacy Commissioner of Ontario Canada [76, p. 3] accepts that de-identification methods are unable to completely mitigate the risks of re-identification: "*While it is not possible to guarantee that de-identification will work 100 per cent of the time, it is still an essential tool that drastically reduces the risk that personal information will be used or disclosed for unauthorized or malicious purposes."*

The risk of re-identification

The risk of re-identification is defined as: (1) the probability in which an individual/individuals could be identified from specific data [74, p. 26], [75, p. 24]; and, (2) the likely resultant harm or impact if re-identification were to occur [75, p. 24].[116] The Article 29 Working Party [63, pp. 11-12] identifies three principal ways in which individuals may be re-identified, by: (1) singling-out an individual; (2) linking records relating to an individual; and (3) inferring information concerning an individual.[117]

It is asserted that where personal data are *"properly de-identified"* [76, p. 4], the risk of re-identification will be *"extremely low"* [76, p. 4]. However, you cannot *"leave and forget"* – with the technological advances need to ensure anonymity is sustainable [77, p. 3].[118] Given these difficulties, it is unsurprising that the legal standard for anonymisation is described as *"very high"* [63, p. 6] Furthermore, in view of this uncertainty, it is again unsurprising that James Clark [78, p. 11] describes the *"required standard for anonymisation as a moving target"*. It will be of considerable interest to observe how the required standard for anonymisation develops over the coming years.

It is important to note that in May 2018 [79], the re-identification of de-identified personal data (without an appropriate defence) has become a criminal offence under Section 171(1) of the UK Data Protection Act [80]:*"It is an offence for a person knowingly or recklessly to re-identify information that is de-identified personal data without the consent of the controller responsible for de-identifying the personal data".*[119]

Data that are apersonal

Apersonal data are the second type of non-personal data – these data do not relate to individuals through content, purpose or result. Elliot et al. [62, p. 10] provide the following examples of apersonal data: *"Astronomical data, meteorological data, food nutrition data, bus timetables, seismological data, stress readings for the Humber Bridge and lists of endangered species".*

Legal decision-tree 1: determine whether the planned processing involves personal data

Instructions

Use the following decision-tree (on the next page) to determine whether a planned data processing activity is likely to involve data that relates to individuals.

You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

- A company wants to use data collected from its loyalty card scheme to target new products at specific customers.
- Researchers want to re-use patient-monitoring data for a specific research study.
- An organisation wants to publish a dataset about levels of traffic in a particular city on its website.
- A dataset about levels of traffic is re-used by a taxi firm to monitor its employees.
- The locations and exact numbers of endangered species in a named location.
- The publication of census data.

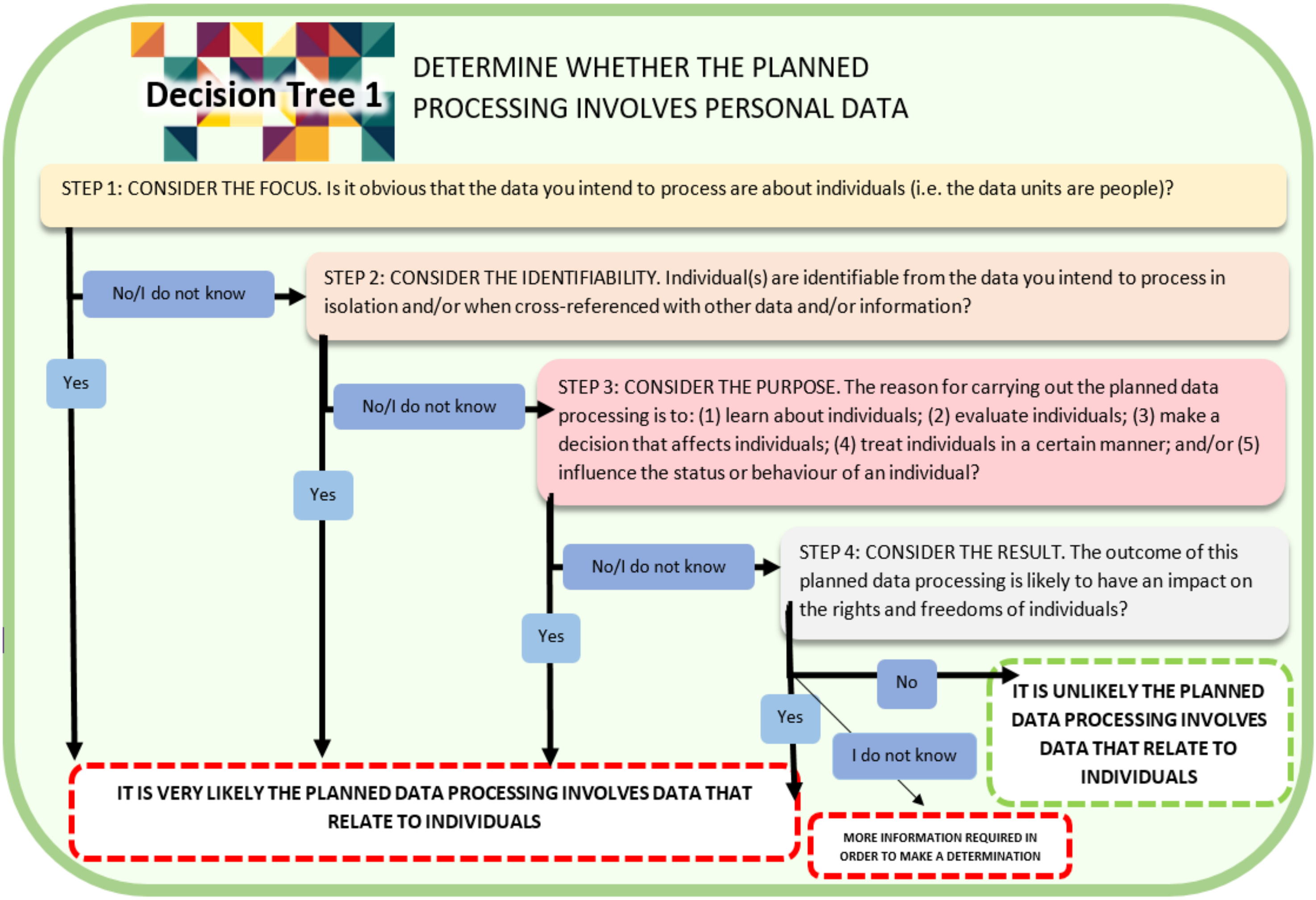
[116] In the words of Marion Oswald [75, p. 24], it is an assessment about *"the possibility of something bad happening"*.

[117] Moreover, Boris Lubarsky [178] outlines the three principal methods of re-identification: (a) *"insufficient de-identification"*; (b) *"pseudonym reversal"*; and, (c) *"combining datasets".*

[118] Kate Brimsted [77, p. 3] states: *"sever the link between the individuals and their data, and you can side step the Gordian complexity of EU data protection law. […] [/] However, […] anonymised data are starting to exhibit troubling signs of becoming 're-identified' or at least 're-identifiable'. […] it is clear that -- due to the speed of technological advances in analytics -- the data reversibility clock is ticking. Anonymisation -- the solution and the price to pay to unlock the broader utility present in massive data sets -- is getting more and more difficult to carry out with confidence."*

[119] For further background information on this offence under Section 171(1) of the Data Protection Act 2018 see: [184], [185], [186]. For an extensive overview on the history and debates surrounding the *"criminal prohibition of wrongful re-identification"* of anonymised data see the following article [187] published by Mark Philips et al.

Figure 11 Legal Decision-Tree 1: Determine whether the planned processing involves personal data



Decision Tree 1

DETERMINE WHETHER THE PLANNED PROCESSING INVOLVES PERSONAL DATA

STEP 1: CONSIDER THE FOCUS. Is it obvious that the data you intend to process are about individuals (i.e. the data units are people)?

STEP 2: CONSIDER THE IDENTIFIABILITY. Individual(s) are identifiable from the data you intend to process in isolation and/or when cross-referenced with other data and/or information?

STEP 3: CONSIDER THE PURPOSE. The reason for carrying out the planned data processing is to: (1) learn about individuals; (2) evaluate individuals; (3) make a decision that affects individuals; (4) treat individuals in a certain manner; and/or (5) influence the status or behaviour of an individual?

STEP 4: CONSIDER THE RESULT. The outcome of this planned data processing is likely to have an impact on the rights and freedoms of individuals?

IT IS UNLIKELY THE PLANNED DATA PROCESSING INVOLVES DATA THAT RELATE TO INDIVIDUALS

IT IS VERY LIKELY THE PLANNED DATA PROCESSING INVOLVES DATA THAT RELATE TO INDIVIDUALS

MORE INFORMATION REQUIRED IN ORDER TO MAKE A DETERMINATION

NOTE –this decision-tree has been derived by the toolkit authors from the following sources: [62], [69], [70], [71] and [72]. Refer to sections 2.2.2 and 2.2.3 for further information.

High-risk data processing

GDPR: a risk-based approach

The General Data Protection Regulation (GDPR) 2016/679 [12] marks an important change in the overall legal approach to EU data protection law – from *"an administrative process based on a priori controls to a risk-based accountability"* [81, p. xiii]. Article 35(3) of the GDPR provides three examples of where data processing is likely to result in a high risk [82, p. 8] (see section 2.3.2 below), and therefore requires a mandatory data protection impact assessment (DPIA) to assess *"the impact of the envisaged processing operations on the protection of personal data"* (Article 35(1) of the GDPR):

Security of processing:
"Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk […]"
Source: Article 32(1) of the General Data Protection Regulation (GDPR) [12]

Once again, the likelihood and severity of such risks to the rights and freedoms of natural persons are dependent on the specific context and purpose of the planned data processing activity under scrutiny.[120] An advantage of using data flow mapping is that it can draw attention to (potential) high-risk data processing activities before data are shared and re-used within an open innovation environment. It can further help to identify the appropriate technical and organisational measures required to assist with the desired level of control over a dataset.

GDPR: types of high-risk processing

Therefore, a crucial part of anonymisation assessment is for data owners to determine whether the planned data processing is likely to constitute a high risk to the rights and freedoms of data subjects [82], [83, p. 13]. Despite no *"definitive"* [84] list for high-risk data processing activities, three examples of high-risk data processing activities (see below) are outlined by Article 35(3) of the GDPR. Furthermore, under Article 35(4) of the GDPR the supervisory authority can release a public list of high-risk processing examples that require a mandatory DPIA.[121] For instance, the Article 29 Data Protection Working Party [82] and ICO [84] have both released such lists of examples.

Article 35 of the GDPR

Article 35(3) provides examples of where data processing is likely to result in a high risk [82, p. 8]:

Examples of high-risk data processing activities:
"A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of: [/] (a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person; [/] (b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or [/] (c) a systematic monitoring of a publicly accessible area on a large scale."

[120] Richard Thomas [188, p. 4] provides three types of harm that may result from the processing of personal data: "*This is potentially controversial territory, but I suggest that three main types of harm can be identified and should be set out in suitable terms in the [GDPR] legislation: [/] Material/tangible harm to individuals: For example, damage to health, financial interests, liberty or freedom of movement. [/] Moral/non-tangible harm to individuals: For example, damage to reputation or to expectations of privacy and family life. [/] Societal harm: For example, threats to the democratic values of a free society or the prospect of excessive State power."*

[121] Article 35(4) of the GDPR [12] states: *"The supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment pursuant to paragraph 1. The supervisory authority shall communicate those lists to the Board referred to in Article 68."*

Source: Article 35(3) of the General Data Protection Regulation (GDPR) [12]

In consequence, the first step towards an assessment over whether a planned data processing is high risk requires consideration of the following question:

Legal Decision-Tree 2: Determine whether the planned processing is likely to be a high risk to individuals?
Step 1 of 3: Assess whether the planned processing is one of the three types of processing operations that are considered as high risk under Article 35(5) of the GDPR.
Q: Does the planned processing involve (at least one of the following):
(i) Systematic and extensive profiling of individuals (e.g. profiling and prediction) with significant effects?
(ii) Large scale use of sensitive data.
(iii) Public monitoring.
A: Yes/No/I do not know

Information Commissioner's Office (ICO) guidance

National data protection authorities also have an important role to issue further guidance to data owners on the types of high-risk processing (aside from the three examples explicitly featured in the GDPR) that would require a mandatory DPIA. For instance, in the UK, ICO has released a list of ten high-processing examples (the following is paraphrased – see original webpage [84] for full information – including the entire list of examples): [122]

ICO ten examples of processing likely to result in high risk:

(i) *"New technologies"* – e.g. *"Artificial intelligence, machine learning and deep learning".*
(ii) *"Denial of service"* – e.g. *"Credit checks".*
(iii) *"Large-scale profiling"* – e.g. *"Social media networks"*
(iv) *"Biometrics"* – e.g. *"Facial recognition systems".*
(v) *"Genetic data"* – e.g. "*Medical diagnosis".*
(vi) *"Data matching"* – e.g. *"Fraud prevention".*
(vii) *"Invisible processing"* – e.g. *"Online advertising".*
(viii) *"Tracking"* – e.g. *"Data processing in the context of home and remote working".*
(ix) *"Targeting of children or other vulnerable adults"* – e.g. *"Connected toys".*
(x) *"Risk of physical harm"* – e.g. *"Whistle-blowing/compliant procedures".*

Source: ICO guidance online [84]

Therefore, the second step towards an assessment over whether a planned data processing is high risk requires consideration of further high-risk processing examples provided by national authoritative bodies, such as ICO:

Legal Decision-Tree 2: Determine whether the planned processing is likely to be a high risk to individuals?
Step 2 of 3: Assess whether the planned processing is one of the ten types of processing operations that are considered as high risk by the Information Commissioner's Office (ICO).
Q: Does the planned processing involve (at least one of the following): (i) new technologies; (ii) denial of service; (iii) large-scale profiling; (iv) biometrics; (v) genetic data; (vi) data matching; (vii) invisible processing; (viii) tracking; (ix) targeting of children or other vulnerable adults; and/or (x) risk of

[122] It is important to note that other national data protection authorities within the EU also issue guidance on high-risk processing under the GDPR. For instance, the Agencia Española de Protección de Datos (AEPD) [189] outlines the following four high-risk scenarios [83, p. 13]: (i) decision-making; (ii) profiling; (iii) predictive analysis; and, (iv) health-related services, monitoring, control and observation of persons (monitoring). [This has been translated into English via Google Translate – original text [83, p. 13] in Spanish: *"Finalidades del tratamiento: Se deben identificar cada una de las finalidades del tratamiento y analizar si estas derivan en un alto riesgo. Por ejemplo, si la finalidad incluye: [/] Toma de decisiones [/] Elaboración de perfiles [/] Análisis predictivo [/] Prestación de servicios relacionados con la salud Seguimiento, control y observación de personas (monitorización)".*]

physical harm.
A: Yes/No/I do not know

Article 29 Working Party guidance

The Article 29 Working Party [82, pp. 9-11] provides the following nine point criteria to be taken into account when considering whether a (planned) data processing activity is likely to result in a high risk to the rights and freedoms of individuals (the following is paraphrased – see original guidelines [82, pp. 9-11] for full information – including the entire list of examples):

Article 29 Working Party nine point criteria of processing likely to result in high risk:

(i) *"Evaluation or scoring"* – e.g. *"a company building […] marketing profiles based on usage or navigation of its website".*

(ii) *"Automated-decision making with legal or similar legal effect"* – e.g. *"the processing may lead to the exclusion or discrimination against individuals".*

(iii) *"Systematic monitoring"* – e.g. *"data collected through networks".*

(iv) *"Sensitive data or data of a highly personal nature" – e.g. "a general hospital keeping patients' medical records".*

(v) *"Data processed on a large-scale"* – In order to determine whether data are processed on a large-scale, the following factors must be considered: *"number of subjects", "volume" and/or "range", "duration" and/or "permanence"*, and *"geographical extent"*.

(vi) *"Matching or combining datasets".*

(vii) *"Data concerning vulnerable data subjects"* – e.g. "*children"* and *"employees"*.

(viii) *"Innovative use or applying new technological or organisational solutions"* – E.g. "*combining use of finger print and face recognition for improved physical access control".*

(ix) *"When the processing in itself "prevents data subjects from exercising a right or using a service or contract"* – e.g. *"a bank screens its customers against a credit reference database in order to decide whether to offer them a loan".*

Source: Article 29 Working Party Opinion guidelines on DPIA [82, pp. 9-11]

Following this approach, a DPIA is required when two criteria are present in a planned data processing activity. A DPIA is recommended if only one criterion is existent. Hence, the third step towards an assessment over whether a planned data processing is high risk requires consideration of further high-risk criteria provided by the Article 29 Working Party:

Legal Decision-Tree 2: Determine whether the planned processing is likely to be a high risk to individuals?
Step 3 of 3: Assess whether the planned processing is falls under the nine-point criteria of processing operations that are considered as high risk by the Article 29 Working Party.
Q: Does the planned processing involve (at least one of the following): (i) evaluation or scoring; (ii) automated-decision making with legal or similar legal effect; (iii) systematic monitoring; (iv) sensitive data or data of a highly personal nature; (v) data processed on a large-scale; (vi) matching or combining datasets; (vii) data concerning vulnerable subjects; (viii) innovative use or applying new technological or organisational solutions; and (ix) preventing data subjects from exercising a right or using a service or contract.
A: Yes/No/I do not know

Legal Decision-Tree 2: Is the planned data processing likely to result in a high risk to the rights and freedoms of individuals?

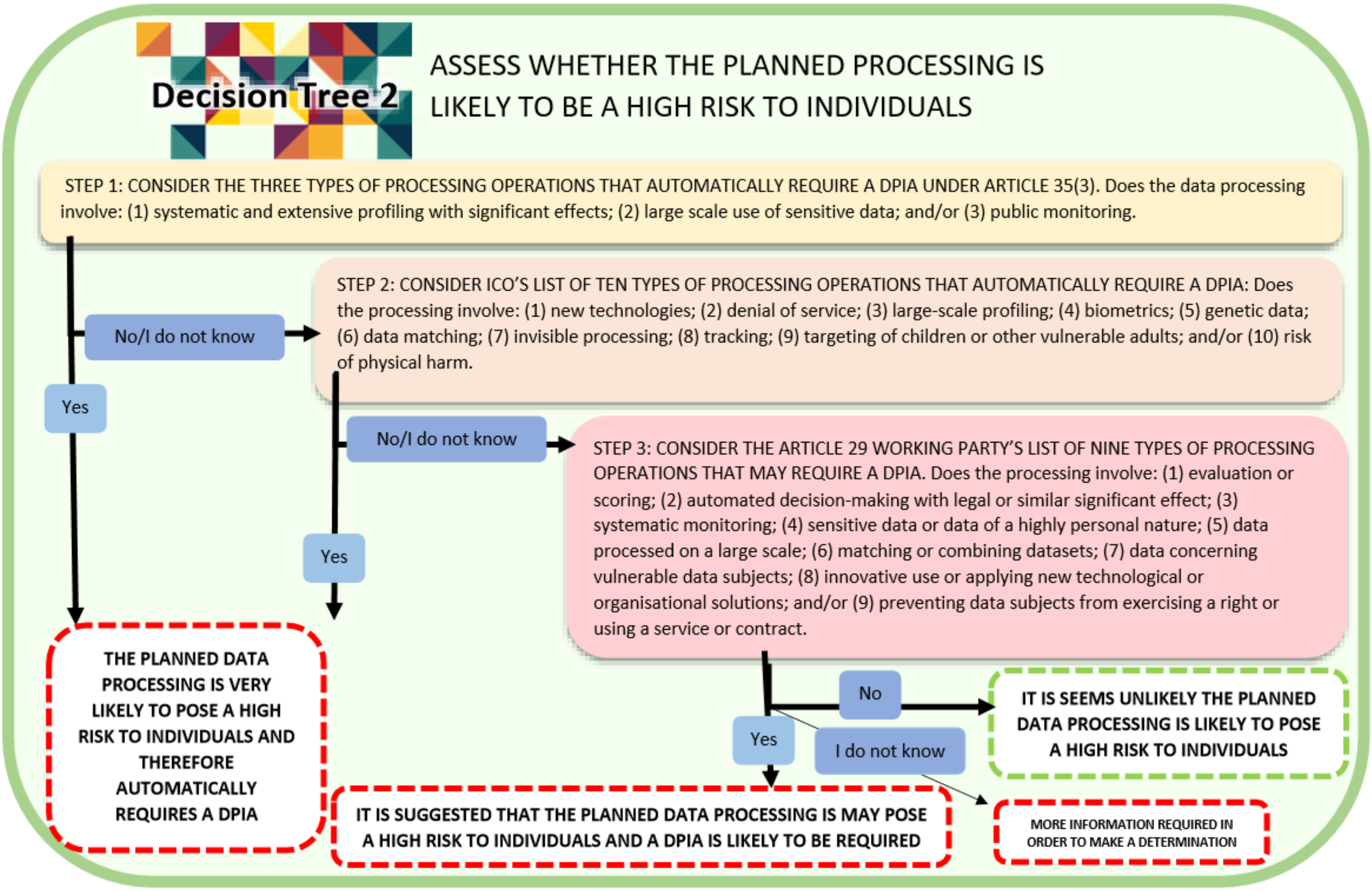
Instructions

Use the following decision-tree (on the next page) to determine whether a planned data processing activity is likely to be high risk based on examples provided by Article 35(3) of the GDPR, ICO [84], and the Article 29 Working Party [82]. It is important to note that this Legal Decision Tree is indicative of high-risk processing – it is not a definitive list of high-risk processing activities.

You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

- A company wants to use data collected from its loyalty card scheme to target new products at specific customers.
- Researchers want to re-use patient-monitoring data for a specific research study.
- An organisation wants to publish a dataset about levels of traffic in a particular city on its website.
- A dataset about levels of traffic is re-used by a taxi firm to monitor its employees.
- The locations and exact numbers of endangered species in a named location.
- The publication of census data.

Figure 12 Legal Decision-Tree 2: Assess whether the planned processing is likely to be a high risk to individuals



ASSESS WHETHER THE PLANNED PROCESSING IS LIKELY TO BE A HIGH RISK TO INDIVIDUALS

STEP 1: CONSIDER THE THREE TYPES OF PROCESSING OPERATIONS THAT AUTOMATICALLY REQUIRE A DPIA UNDER ARTICLE 35(3). Does the data processing involve: (1) systematic and extensive profiling with significant effects; (2) large scale use of sensitive data; and/or (3) public monitoring.

STEP 2: CONSIDER ICO'S LIST OF TEN TYPES OF PROCESSING OPERATIONS THAT AUTOMATICALLY REQUIRE A DPIA: Does the processing involve: (1) new technologies; (2) denial of service; (3) large-scale profiling; (4) biometrics; (5) genetic data; (6) data matching; (7) invisible processing; (8) tracking; (9) targeting of children or other vulnerable adults; and/or (10) risk of physical harm.

STEP 3: CONSIDER THE ARTICLE 29 WORKING PARTY'S LIST OF NINE TYPES OF PROCESSING OPERATIONS THAT MAY REQUIRE A DPIA. Does the processing involve: (1) evaluation or scoring; (2) automated decision-making with legal or similar significant effect; (3) systematic monitoring; (4) sensitive data or data of a highly personal nature; (5) data processed on a large scale; (6) matching or combining datasets; (7) data concerning vulnerable data subjects; (8) innovative use or applying new technological or organisational solutions; and/or (9) preventing data subjects from exercising a right or using a service or contract.

THE PLANNED DATA PROCESSING IS VERY LIKELY TO POSE A HIGH RISK TO INDIVIDUALS AND THEREFORE AUTOMATICALLY REQUIRES A DPIA

IT IS SEEMS UNLIKELY THE PLANNED DATA PROCESSING IS LIKELY TO POSE A HIGH RISK TO INDIVIDUALS

IT IS SUGGESTED THAT THE PLANNED DATA PROCESSING IS MAY POSE A HIGH RISK TO INDIVIDUALS AND A DPIA IS LIKELY TO BE REQUIRED

MORE INFORMATION REQUIRED IN ORDER TO MAKE A DETERMINATION

NOTE – If you answer "I don't know" to any of the questions – overall outcome = more information is required.

This decision-tree has been derived by the Toolkit authors from the following sources: Article 35(3) of the GDPR, [82] and [84]. Refer to sections 2.3.1 and 2.3.2 for further information.

What levels of control are required?

Data protection by default and design

Pursuant to Article 25(1) of the GDPR [12], controllers need to ensure that any planned data processing preserves the rights and freedoms of data subjects by design and default. This protection is achieved by: (i) implementing appropriate legal and technical measures to uphold data protection principles (see Article 5 of the GDPR); and (ii) integrating safeguards. According to Article 25(1), any decision taken on these measures and safeguards must take: *"into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing".*

Data type definitions

According to Roger Clark [85], the function of big data analytics can be split into the following two categories: (1) individual-focused data analytics – *"concerned about individual instances within populations"*; and, (2) population-focused data analytics – *"focus on populations and sub-populations"*. To help data (re)users to better-review the potential risks of the planned data processing, there must therefore be an assessment of whether this processing is individual-focused or population-focused. For the purpose of the Legal and Privacy Toolkit v2, the following data types are defined:

Data at individual-level: Data are recorded for each specific person [85] i.e. the data units are people [62, p. 9]. E.g. data relate to a particular customer, patient or survey participant [85], [86], [87].[123]

Data at aggregate-level: Data are recorded in a summary form (e.g. statistics) about sub-populations and populations i.e. the data units are about groups of people [85].[124] E.g. statistics about customer preferences.[125]

Apersonal data: The data units are non-people entities such as events or processes. E.g. a bus timetable.[126]

Dependant on context and purpose

Different levels of access and control may be applied to different versions of the particular (version of a) dataset that a data owner plans to process based on the specific circumstances of each data processing activity.

Example:
Version of dataset *y* – Raw data in closed environment. A medical professional collects sensitive personal information about a patient in order to identify and action the most effective course of treatment for medical condition *x*.
Version of dataset *y* – Pseudonymised data in restricted environment. A research team is given permission to (re)use a pseudonymous form of this data as part of a large-scale study into the treatments for medical condition *x*.
Version of dataset *y* – Aggregated data in public domain with an open licence. The hospital releases statistics about the number of patients treated for the medical condition *x* over the past

[123] A further example of individual-level data are census microdata, refer to [190] for more information.

[124] For further definitions of data at an aggregated level see the following references. (i) The Organisation for Economic Co-operation and Development (OECD) [191] defines the term aggregation in its glossary of statistical terms as*: "the combination of related categories, usually within a common branch of a hierarchy, to provide information at a broader level to that at which detailed observations are taken."* (ii) Margaret Rouse [192] defines data aggregation as: *"any process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis. A common aggregation purpose is to get more information about particular groups based on specific variables such as age, profession, or income."*

[125] A further example of aggregate-level data are census aggregate data, refer to [193] for more information.

[126] For further examples of apersonal data see [62, p. 10].

year.

For data processing that falls under the scope of the GDPR, data owners will have to take appropriate technical and organisational measures to ensure that they stay in control of the level of agreed access and re-usage. Furthermore, where data are properly anonymised, data owners will also have to take appropriate technical and organisational measures to ensure that they stay in control of the level of agreed access and (re)usage. The GDPR focuses on the roles of controllers and processors:

(i) **Controller.** A controller is defined by Article 4(7) of the GDPR as: *"the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data […]"*. According to Article 24(1) of the GDPR, the controller is responsible for the implementation of *"[…] appropriate technical and organisational measures […]".*

(ii) **Processor.** A processor is defined by Article 4(8) of the GDPR as: *"a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*". Article 28(1) of the GDPR explains the responsibilities of the processor, including the following requirement: *"[…] sufficient guarantees [to the controller] to implement appropriate technical and organisational measures […]".*

Legal Decision-Tree 3: Determine the status of control over current and planned data situations

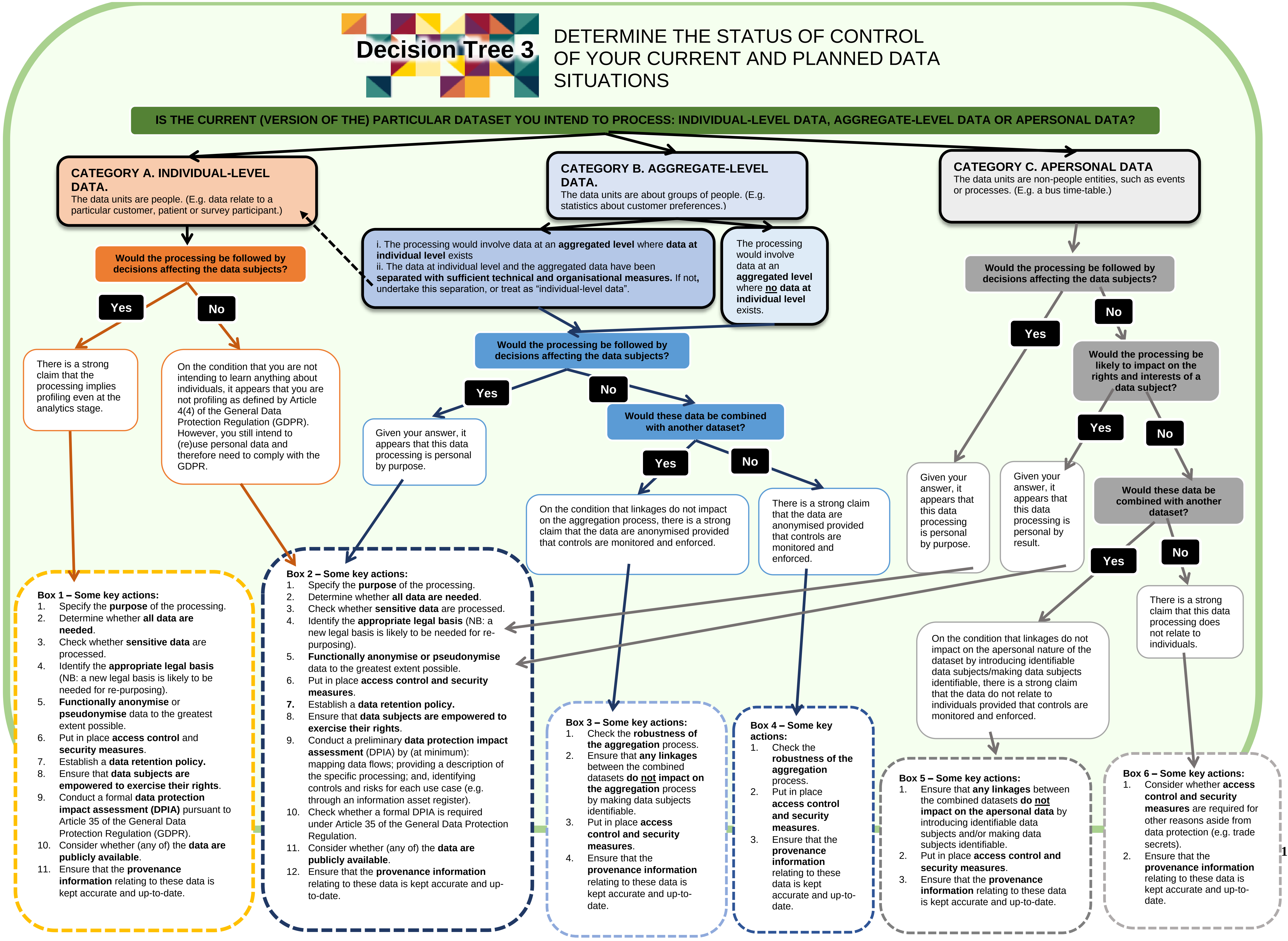
The following legal decision-tree provides an overview of the likely types of controls required for the following three categories of data: (i) data at an individual-level; (ii) data at an aggregate-level; and, (iii) apersonal data. The accompanying table (located directly after this decision-tree) offers further descriptive information about the key actions outlined by this decision-tree. Again, this legal decision-tree is indicative of the types of controls required for different types of data. It is the responsibility of data owners to take into account the individual circumstances of their planned processing activity as part of their anonymisation assessment.

Instructions

Use the following decision-tree to determine the likely level of controls required for a planned data processing activity. You may wish to think about data you intend to process. You may also find it useful to utilise the following fictional scenarios:

- A company wants to use data collected from its loyalty card scheme to target new products at specific customers.
- Researchers want to re-use patient-monitoring data for a specific research study.
- An organisation wants to publish a dataset about levels of traffic in a particular city on its website.
- A dataset about levels of traffic is re-used by a taxi firm to monitor its employees.
- The locations and exact numbers of endangered species in a named location.
- The publication of census data.

Figure 13 Legal Decision-Tree 3: Determine the status of control of your current and planned data situations



Decision Tree 3

DETERMINE THE STATUS OF CONTROL OF YOUR CURRENT AND PLANNED DATA SITUATIONS

IS THE CURRENT (VERSION OF THE) PARTICULAR DATASET YOU INTEND TO PROCESS: INDIVIDUAL-LEVEL DATA, AGGREGATE-LEVEL DATA OR APERSONAL DATA?

CATEGORY A. INDIVIDUAL-LEVEL DATA.
The data units are people. (E.g. data relate to a particular customer, patient or survey participant.)

Would the processing be followed by decisions affecting the data subjects?

Yes

There is a strong claim that the processing implies profiling even at the analytics stage.

No

On the condition that you are not intending to learn anything about individuals, it appears that you are not profiling as defined by Article 4(4) of the General Data Protection Regulation (GDPR). However, you still intend to (re)use personal data and therefore need to comply with the GDPR.

CATEGORY B. AGGREGATE-LEVEL DATA.
The data units are about groups of people. (E.g. statistics about customer preferences.)

i. The processing would involve data at an **aggregated level** where **data at individual level** exists
ii. The data at individual level and the aggregated data have been **separated with sufficient technical and organisational measures.** If not, undertake this separation, or treat as "individual-level data".

The processing would involve data at an **aggregated level** where **no data at individual level** exists.

Would the processing be followed by decisions affecting the data subjects?

Yes

Given your answer, it appears that this data processing is personal by purpose.

No

Would these data be combined with another dataset?

Yes

On the condition that linkages do not impact on the aggregation process, there is a strong claim that the data are anonymised provided that controls are monitored and enforced.

No

There is a strong claim that the data are anonymised provided that controls are monitored and enforced.

CATEGORY C. APERSONAL DATA
The data units are non-people entities, such as events or processes. (E.g. a bus time-table.)

Would the processing be followed by decisions affecting the data subjects?

Yes

Given your answer, it appears that this data processing is personal by purpose.

No

Would the processing be likely to impact on the rights and interests of a data subject?

Yes

Given your answer, it appears that this data processing is personal by result.

No

Would these data be combined with another dataset?

Yes

On the condition that linkages do not impact on the apersonal nature of the dataset by introducing identifiable data subjects/making data subjects identifiable, there is a strong claim that the data do not relate to individuals provided that controls are monitored and enforced.

No

There is a strong claim that this data processing does not relate to individuals.

Box 1 – Some key actions:
1. Specify the **purpose** of the processing.
2. Determine whether **all data are needed**.
3. Check whether **sensitive data** are processed.
4. Identify the **appropriate legal basis** (NB: a new legal basis is likely to be needed for re-purposing).
5. **Functionally anonymise** or **pseudonymise** data to the greatest extent possible.
6. Put in place **access control** and **security measures**.
7. Establish a **data retention policy.**
8. Ensure that **data subjects are empowered to exercise their rights**.
9. Conduct a formal **data protection impact assessment (DPIA)** pursuant to Article 35 of the General Data Protection Regulation (GDPR).
10. Consider whether (any of) the **data are publicly available**.
11. Ensure that the **provenance information** relating to these data is kept accurate and up-to-date.

Box 2 – Some key actions:
1. Specify the **purpose** of the processing.
2. Determine whether **all data are needed**.
3. Check whether **sensitive data** are processed.
4. Identify the **appropriate legal basis** (NB: a new legal basis is likely to be needed for re-purposing).
5. **Functionally anonymise or pseudonymise** data to the greatest extent possible.
6. Put in place **access control and security measures**.
7. Establish a **data retention policy.**
8. Ensure that **data subjects are empowered to exercise their rights**.
9. Conduct a preliminary **data protection impact assessment** (DPIA) by (at minimum): mapping data flows; providing a description of the specific processing; and, identifying controls and risks for each use case (e.g. through an information asset register).
10. Check whether a formal DPIA is required under Article 35 of the General Data Protection Regulation.
11. Consider whether (any of) the **data are publicly available**.
12. Ensure that the **provenance information** relating to these data is kept accurate and up-to-date.

Box 3 – Some key actions:
1. Check the **robustness of the aggregation** process.
2. Ensure that **any linkages** between the combined datasets **do not impact on the aggregation** process by making data subjects identifiable.
3. Put in place **access control and security measures**.
4. Ensure that the **provenance information** relating to these data is kept accurate and up-to-date.

Box 4 – Some key actions:
1. Check the **robustness of the aggregation** process.
2. Put in place **access control and security measures**.
3. Ensure that the **provenance information** relating to these data is kept accurate and up-to-date.

Box 5 – Some key actions:
1. Ensure that **any linkages** between the combined datasets **do not impact on the apersonal data** by introducing identifiable data subjects and/or making data subjects identifiable.
2. Put in place **access control and security measures**.
3. Ensure that the **provenance information** relating to these data is kept accurate and up-to-date.

Box 6 – Some key actions:
1. Consider whether **access control and security measures** are required for other reasons aside from data protection (e.g. trade secrets).
2. Ensure that the **provenance information** relating to these data is kept accurate and up-to-date.

Legal decision-tree 3: key actions in brief

Legal Decision-Tree 3 aims to show data owners how the type of data they intend to process may affect the types of technical and organisational measures and safeguards, which are required to control the flow of data. The following table provides further information about the key actions outlined in Boxes 1-6:

Figure 14 Table 5: Legal Decision-Tree 3 - Further information on the key actions outlined in Boxes 1-6

Key action	Box(es)	Brief description
Specify the purpose of the processing.	1 and 2	The principle of purpose limitation is enshrined by Article 5(1)(b) of the GDPR: *"Personal data shall be […] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation')".* Data owners therefore need to ensure that they specify the purpose of the planned processing.
Determine whether all data are needed.	1 and 2	The principle of data minimisation is enshrined by Article 5(1)(c) of the GDPR: *"Personal data shall be […] adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')".* Data owners therefore need to ensure that all data are necessary for the planned processing activity.
Check whether sensitive data are processed.	1 and 2	It is imperative for the data owner to review whether the planned data processing activity would involve any sensitive data. This is because special categories of data – defined by Article 9 of the GDPR – may only be processed if an exemption applies (see Article 9(2)(a)-(j) for these exemptions): *""Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."*
Identify the appropriate legal basis (NB: a new legal basis is likely to be needed for re-purposing).	1 and 2	It is crucial for the data owner to determine the lawful basis for the planned data processing activity before processing begins. Article 6 of the GDPR provides six lawful bases: (i) consent, (ii) contract, (iii) legal obligation, (iv) vital interests, (v) public task, and (vi) legitimate interests. The context and purpose of specific, planned data processing activity will impact the legal basis/bases that is/are selected [88]. In addition to detailed guidance on lawful basis for processing [88], ICO also provide a lawful basis interactive tool [89] on their website – to offer *"tailored guidance"* [88].
Functionally anonymise or pseudonymise data to the greatest extent possible.	1 and 2	Where possible, personal data should be anonymised [73, p. 13]. Functional anonymisation takes into consideration the data situation of the planned data processing – see [62, pp. 21-22] for further information. In some cases, it may not be possible to fully anonymise data, because it would adversely affect the utility of the specific data in question [73, p. 13], [90]. In this instance, data should be pseudonymised to the greatest extent possible.

Put in place access control and security measures.	1, 2, 3, 4 and 5	The principle of integrity and confidentiality is enshrined by Article 5(1)(f) of the GDPR: *"Personal data shall be […] processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')."* Furthermore, Article 25(2) states: *"The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed."* Data owners therefore need to establish appropriate access and security measures for their planned data processing activities. For instance, the European Union Agency for Network and Information Security (ENISA) provides: a Handbook on Security of Personal Data Processing [91].
Consider whether access control and security measures are required for other reasons aside from data protection (e.g. trade secrets).	6	While this report focuses on control measures for GDPR-compliance, data owners remain cognisant of other legal and ethical obligations. Data owners must therefore assess what other factors may require the use of access and security measures – e.g. commercial confidentiality.
Establish a data retention policy.	1 and 2	The principle of storage limitation is enshrined by Article 5(1)(e) of the GDPR*: "Personal data shall be […] kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed […] ('storage limitation')".*[127] A data retention policy (e.g. a document that specifies how data are stored, managed and deleted) is an essential way of adhering to this principle.
Ensure that data subjects are empowered to exercise their rights.	1 and 2	Chapter III of the GDPR describes the rights of data subjects – these rights cover four main areas: (i) *"transparency and modalities"* (Article 12); (ii) *"information and access to personal data"* (Articles 13-15); (iii) *"rectification and erasure"* (Articles 16-20); and, (iv) *"right to object and automated individual decision-making"* (Articles 21-22). It is critical that data owners are cognisant of the rights of data subjects, and therefore ensure that data subjects are empowered to exercise such rights.
Conduct a formal data protection impact assessment (DPIA) pursuant to Article 35 of the General Data Protection Regulation (GDPR).	1	As aforementioned, a DPIA is mandatory for a (planned) data processing activity that is likely to result in a high-risk to the rights and freedoms of natural persons (see section 2.3 of this report for further information). A number of national data authorities provide guidance information about DPIAs, (including ICO – see [92], and Commission Nationale de l'Informatique et des Libertés (CNIL) which provides an open source PIA software to help users carry out DPIAs – see [93]).
Conduct a preliminary data protection impact assessment (DPIA)	2	While the GDPR has made DPIAs a legal requirement in certain circumstances, DPIAs (also known as privacy impact assessments) are not new and are an established part of best practice. It is therefore important for data owners to conduct a preliminary impact assessment by (at minimum): mapping data flows; providing a description of the specific processing; and,

[127] Note that under Article 5(1)(e) of the GDPR personal data that is *"processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes"* can be stored for longer periods in accordance with the conditions specified by this Article.

		identifying controls and risks for each use case (e.g. through an information asset register).
Check whether a formal DPIA is required under Article 35 of the General Data Protection Regulation.	2	Data owners must make sure that the planned data processing does not require a formal DPIA, i.e. that it is unlikely to result in a high-risk to the rights and freedoms of data subjects.
Consider whether (any of) the data are publicly available.	1 and 2	The principle of data minimisation is enshrined by Article 5(1)(c) of the GDPR – if these data are already available for re-use, the planned data processing activity may contravene this principle.
Ensure that the provenance information relating to these data is kept accurate and up-to-date.	All	Refer to section 3.2.4 of this [original] report for further information.
Check the robustness of the aggregation process.	3 and 4	Data owners need to ensure that the aggregation process meets the legal standard for anonymisation now and in the future [63, p. 24]. In view of the *"residual risk of identification"*, anonymised data cannot be released and forgotten [63, p. 24]. The residual risk of identification should therefore be actively assessed, monitored and controlled [63, p. 24].
Ensure that any linkages between the combined datasets do <u>not</u> impact on the aggregation process by making data subjects identifiable.	3	Anonymised data cannot be released and forgotten [63, p. 24], data owners must actively monitor and manage any risks of re-identification. For instance, there is a possibility for re-identification to occur following a combining of datasets.
Ensure that any linkages between the combined datasets do <u>not</u> impact on the apersonal data by introducing identifiable data subjects and/or making data subjects identifiable.	5	While a dataset may be apersonal in isolation, it may be personal in combination with another dataset. E.g. a bus timetable is apersonal in isolation, but when combined with passenger data – for the purpose of monitoring the route and length of journeys taken over a specific period – it is likely to be personal data. Again, it is critical to take into account the specific context and purpose of the planned data processing.

Summary: data spectrum

After reviewing Part A, it is anticipated that the user of this guidance document should now have a better or re-affirmed understanding of (i) how personal and non-personal data are legally defined, including (ii) what types of data processing are high risk, and (iii) what appropriate control measures should be implemented. This legal understanding is essential for effective data flow mapping as part of an overall approach to GDPR-compliance.

Any decision that focuses on the appropriate level and type of measures to control a data flow (e.g. a specific data sharing activity) will <u>heavily rely</u> on the individual circumstances that surround the particular activity under consideration. Since open innovation programmes are highly likely to traverse a multitude of sectors – e.g. from hospitality to health care – it is not possible to provide a data situation model that covers all these possibilities. In consequence, the next part of this report (Part B) presents the basics of data flow mapping as a useful approach data owners can employ to create such data situation models for assessment of anonymisation practices.

Part B – the basics of mapping data flows

Brief overview

Data flow mapping is one way in which data owners can demonstrate compliance with the GDPR through adherence to Recital 82 of the GDPR [64, p. 6]:

Demonstrating compliance with the GDPR:
"In order to demonstrate compliance with this Regulation, the controller or processor should maintain records of processing activities under its responsibility. Each controller and processor should be obliged to cooperate with the supervisory authority and make those records, on request, available to it, so that it might serve for monitoring those processing operations."
Recital 82 of the General Data Protection Regulation (GDPR) [12]

Data flow mapping can be used at enterprise-level [94], [64], [95, p. 4] (i.e. to chart the movement of all data through an organisation and/or network of organisations) and at dataset-level [62, p. 69] (i.e. to chart the movement of a particular dataset).[128] This guidance focuses on data flow mapping at dataset-level. Data flow mapping can also be undertaken at technical and organisational levels [66, p. 7].

The basics

The key benefits

In order to manage data flows effectively in accordance with the GDPR, those involved with the processing of data need to first understand these data flows. As James Graves [96] states: *"It's much easier to describe how you are protecting the confidentiality, integrity and availability of your data when you can provide details such as where it is going, how it is getting there and who is sending and receiving it."* As aforementioned, data flow mapping has a number of benefits for those involved with data processing.

Some key benefits of data flow mapping:

Demonstrate compliance [64, p. 6] – Data flow mapping can help data owners to adhere to the obligation under Recital 82 of the GDPR that requires a controller or processor to *"maintain records of processing activities under its responsibility"*.

Reveal gaps [65, p. 8], [66, p. 7] – Data flow mapping can help to highlight any gaps between the regulatory framework with how data are collected, managed, shared and (re)used in practice.

Risk mitigation [65, p. 8] – It can draw attention to (potential) high-risk data processing activities before data are shared and re-used within an open innovation environment. It can further help to identify the appropriate technical and organisational measures required to assist with the desired level of control over a dataset.

Robust decision-making. It can provide a knowledge-base for robust decision-making about if and how best to share and re-use data.

Legal training [65, p. 8] – By understanding where the gaps between practice and the regulatory framework lie, open acceleration programmes can better-target the areas that require further legal training.

Approach

The most important part of data flow mapping is the content not the format [65, p. 14]. While some

[128] For instance, the ICO [94] provide documentation templates to document processing activities within an organisation.

approaches adhere to formal standards (e.g. LINDDUN Privacy Threat Modelling [97]) and are software-based (e.g. the Data Flow Mapping Tool [98]), other approaches do not insist on strict formatting rules (e.g. the ODI's Mapping Data Ecosystems Methodology [95]). In the words of Nicola Fulford and Krysia Oastler [64, p. 7]: *"There is no one right way to carry out data mapping".*

Data situation approach

An essential part of effective data flow mapping is an understanding of the data situation(s) that pertain to the particular (version of the) dataset you plan to process. A data situation is defined by UKAN's Anonymisation Decision-Making Framework [62, p. 130] as: *"**Data Situation:** The relationship between some data and their environment"* [Bold emphasis in original]. Each data environment should be considered as a distinct *"configuration"* of *"people, other data, infrastructure and governance structures"* [62, p. 69]. For example, where a data provider shares data with a start-up as part of an innovation acceleration programme – there could be four different data environments linked together in a data flow: (1) the data provider environment; (2) the start-up environment; (3) the intermediary environment (i.e. where data are hosted); and (4) the innovation acceleration programme environment. [See Annex 4 – for further information].[129]

Example of technical content

James Graves [96] outlines three key technical areas to focus on during data flow mapping (the following has been paraphrased from the original – see [96] for full information):

1. *"Policy and standards"* – Consider the rules that govern how and what types of data are permitted to flow: (i) externally: in and out of the organisation; and (ii) internally: via *"various zones"*. Moreover, examine the standards utilised in order to configure data flows. [96]
2. "*Architecture and design"* – Ascertain where data are situated *"at rest"* and *"in motion"* by identifying *"databases", "applications", "users and groups",* and *"current controls"*. [96]
3. *"Technical controls"* – Focus on the technical controls that: (a) monitor data *"at rest"* and *"in motion"*; and, (b) manage the data flows through access restrictions. [96]

Example of organisational content

Nicola Fulford and Krysia Oastler [64, pp. 7-8] highlight six key organisational areas to focus on during data flow mapping (the following has been paraphrased from the original – see [64, pp. 7-8] for full information):

1. *"The reasons/purposes for processing personal data"* – Consider the legal basis for processing personal data. [64, p. 7]
2. *"Key stakeholders"* – Identify the key stakeholders and their roles in a data processing activity. [64, p. 7]
3. *"Types/categories and sub-categories of personal data"* – Review details about datasets including their attributes. [64, p. 7]
4. *"Source and location of personal data"* – For example, examine the *"data entry point"* and data storage arrangements.
5. *"[W]ith whom personal data are shared or disclosed"* – Document the types of data sharing relationships between the organisation and other third parties, e.g. *"group companies, with suppliers and services providers, with public/official authorities (such as law enforcement and tax authorities), regulators and with business partners/sponsors."* [64, p. 8]
6. *"[D]ocument the relevant retention periods or retention criteria for that data".* [64, p. 8]

Key stakeholders

It is crucial to identify the key stakeholders that are involved with a data processing activity [64, p. 7] and involve them in the data mapping process. Where it is possible, the person overseeing data flow

[129] For further examples of data flow maps see [95, pp. 6-8].

mapping should be the designated data protection officer [66, p. 8].

Role of provenance

Access to robust provenance information is an advantage for those who are mapping data flows.[130] Provenance information can be defined as data about data – e.g. information pertaining to the origins, licensing, versioning, (non-)personal nature and quality of a particular dataset. For many datasets, their status as personal and non-personal may change throughout their lifecycle as they enter different environments, undergo various (re)uses and a number of transformations.

Therefore, provenance information should be able to facilitate the mapping of past and current data flows that relate to the particular (version of a) dataset that a data owner plans to share or re-use. By understanding the provenance, data owners will be able to assess factors such as:

- Information about any previous anonymisation assessment(s) and/or data impact assessments conducted internally and/or by third parties.
- What versions of the dataset exist and in what form (e.g. does the raw personal data exist as well as a pseudonymised version of the dataset).
- Whether and how the particular dataset has been shared and (re-)used before and by whom.
- If there are plans to share versions of the current dataset as different products.
- The history of control over the dataset, e.g. who has had access and (re)use of the previous versions.

Mapping data flow exercises

Effective mapping of data flows – by its very nature – require the data owner to assess the individual circumstances that surround the specific, planned data processing activity under consideration together with the regulatory framework. For those who are inexperienced with data flow mapping to those who would like to maintain their skills, it can be useful to practise this approach.

The following three scenarios have been designed (by the report authors) in order to practise mapping data flows for different types of data. These three scenario-based data flow mapping exercises may be used by an individual or as part of a group-based exercise.[131]

Scenario A: Medical Research

Brief overview: You are part of a privately-funded medical research team who are focused on an examination of the (non-) effective treatment of *condition A*. Your research relies on large-scale data analysis of health data from a variety of sources, including private health clinics, pharmaceutical companies, survey companies and personal health-trackers.

You have been given access to *dataset x* to re-use as part of your research into the treatment of *condition A*.

***Dataset x* – some further information:**

- Patient-monitoring data, including the course of treatment of administered to 1278 patients with *condition A.*
- Individual-level data.
- Collected by medical staff working at *Hospital Z*, *Ward G* during 2010-2015.
- Direct identifiers are masked (e.g. no patient names appear in the dataset).
- The data provider is a data market place that claims *dataset x* is anonymised.

[130] For further information on how to "document your data" see: [194]; for further background information about provenance see: [195].

[131] Note that the ODI has also produced workshop material on mapping data ecosystems – see [95, pp. 9-15]. However, this ODI workshop guidance (and the wider report) does not share the same focus as this Data Pitch Deliverable 3.5 (i.e. data flow mapping for GDPR-compliance).

1. Sketch a data situation model that maps the flow of *Dataset x* from *Ward G* of *Hospital Z* to the medical research team.
2. Your medical research team aims to publish a research paper in an open access journal and, where possible, the underlying datasets. Add this activity to your data situation model sketch for *Dataset x*.

Scenario B: Pollution-Reduction Strategy

Brief overview: You are part of policy team tasked with the development of a strategy to reduce pollution in the middle of *City A*. Your analysis relies on data from a variety of sources, including traffic monitoring data, real-time sensor readings monitoring levels of pollution, and information from public transport providers.

A private bus firm – that operates within *City A* and the surrounding areas – has given you access to *dataset y* to re-use as part of your strategy development.

***Dataset y*– some further information:**

- Passenger-monitoring data taken from registered bus cards that details the individual journeys taken by each passenger during 2017 (where this data is available).
- The number of non-registered passengers who use the bus each day.
- Direct identifiers are masked (e.g. no passenger names appear in the dataset.)
- Age ranges are recorded.
- The private bus firm claims that *dataset y* is anonymised.

1. Sketch a data situation model that maps the flow of *Dataset y* from the private bus firm to the policy team.
2. Your policy team aims to publish this strategy on the City Council's website for consultation and, where possible, the underlying datasets. Add this activity to your data situation model sketch for *Dataset y*.

Scenario C: Predicting Future Product Trends

Brief overview: You are part of product strategy and marketing team tasked with the enrichment of your company's product portfolio through future trends predictions. Your analysis relies on data from a variety of sources, including internal customer data, survey data and information from competitors about new product releases.

A digital analytics company that analyses market trends has given you access to *dataset z* to re-use for your future trends predictions.

***Dataset z* – some further information:**

- Large-scale data taken from a social media platform.
- Click-rates for advertisements on the social media platform.
- Online tracking information (e.g. what websites have been visited by social platform users).
- Predicted preferences from social media data.
- The digital analytics company claims that *dataset z* is anonymised.

1. Sketch a data situation model that maps the flow of *Dataset z* from the social media platform to the product strategy and marketing team.
2. A member of your product and strategy team aims to use this analysis as part of a talk at an international corporate event that focuses on future products in your sector. Add this activity to your data situation model sketch for *Dataset z*.

Summary: mapping data flows

- Data flow mapping is one way in which those involved with data processing activities can demonstrate compliance with the GDPR.
- Data flow mapping can be used at enterprise-level and/or dataset-level.
- Data flow mapping has numerous benefits, including gap-identification and risk mitigation.
- The content of a data flow map is most crucial – not the format it takes.
- An effective data flow map will take into consideration both the technical and organisational aspects pertaining to a particular data situation.
- Access to robust provenance information is an advantage for mapping data flow activities.

Part C – The development of training tools

D.3.5: dissemination

As aforementioned, the key rationale for this toolkit update is to raise-awareness of the important role that data flow mapping can play in responsible data sharing and re-usage within open innovation programmes; in particular with GDPR-compliance. This aim is achieved via the guidance provided in (i) this report – and further value-added legal training materials (derived from Parts A and B of this report) in the form of (ii) a prototype e-learning tool on data protection and the basics of mapping data flows ("the prototype e-learning tool") and (iii) a workshop.

The prototype e-learning tool

Purpose

As previously stated, the purpose of the three paper-based legal decision-trees is to help communicate some of the key aspects of the GDPR in a simple way to data owners; by representing a key series of concepts and their outcomes. A decision was taken to make these decision-trees interactive through the development of the prototype e-learning tool (accessible via http://pz-wp-test.synote.io/ [last accessed 4 June 2018]) in order to enable tailored responses to a data owner's planned data processing activity. Note: screen-shots from this prototype e-learning tool are located in Annex 1 of this report.

Technical development

This section provides a brief overview of the technical development of the prototype tool, which was built via the following three stages:

Stage 1: Design.

- **An agile co-design process.** An agile co-design process was adopted to suit the ongoing development of the legal decision-trees.
- **Mock-up screen shots.** From the outset, mock-up screen shots of the tool were shared between the content designers and developer. These mock-up screen shots were used to ensure that the real requirements were addressed in the generation of potential applications.
- **Use of a content management system (CMS).** A CMS system was selected so as to provide the non-developer (i.e. one of the content designers) with a simple and straight-forward means for (later) customisation of the content and workflow. Therefore, it was important for the developer to evaluate several CMSs in order to select the most appropriate CMS that met the specified requirements.

Stage 2: Implementation.

- **WordPress.** WordPress was used and deployed in the implementation phase.
- **Three plug-ins.** The following three WordPress plugins were used during the tool development:
  i) **Elementor Page Builder** – This was applied in order to create the basic webpages.

ii) **Chained Quiz** – This was installed and used in order to: (a) create the decision-tree questions; and (b) manage the decision-tree logic (i.e. where the next question depends on the answer to the previous question).

iii) **Popup Builder** – This was activated in order to control and manage pop-up windows (that are used to display further information about a given term within the tool).

Stage 3: Testing.

- **Three test processes:** The prototype e-learning tool was tested via the following three processes:
  i) **Workflow test.** This test assessed whether the pages of the tool were continuously displayed in the correct order.
  ii) **Logic test.** This test validated that the tool would lead users to different pages depending on their answers to specific-questions.
  iii) **User test.** This test asked end users to utilise the tool and provide feedback. In accordance with this feedback, the developer would then apply any necessary amendments to the tool.

Legal training workshop (May 2018)

As part of the Data Pitch programme, a legal training workshop – on data protection and the basics of mapping data flows – was delivered to invited SMEs in May 2018. This workshop provided an opportunity to test the prototype e-learning tool and receive feedback from data owners (see Annex 3 for further information).

Summary

The guidance given on data protection and the basics of mapping data flows is provided via three modes of communication:

1. Paper-based – this deliverable report.
2. Workshop – the workshop materials in Annex 2 can be re-used.
3. Prototype e-learning tool – an interactive version of the three legal decision-trees.

There is scope to improve the current version of the prototype e-learning tool – see Annex 3 [to original report] for further information.

Conclusion

Key points

The Legal and Privacy Toolkit v2 (D3.5) is a toolkit update that extends the data protection guidance provided in the first version of the toolkit. Ultimately, this report aims to show that anonymisation assessment is a crucial part of any planned data processing activity. It addresses the D3.5. objective outlined by the Grant Agreement by providing: (i) data owners with a basis of legal guidance about GDPR-compliance (explored in Part A) in order to take advantage of (ii) the data flow mapping approach (the basics of which are outlined in Part B) that can be used in the creation of data situation models to further support anonymisation assessment.

Data spectrum, context and purpose

The nature of data is changeable – and ultimately predicated on the specific context and purpose of a (planned) data processing activity. For instance, anonymised data can be re-identified – and the same dataset can be considered as non-personal and personal under different sets of circumstances. Therefore, just because a dataset was used for non-personal purposes in the past does not mean that it cannot be utilised for personal purposes in the future. In consequence, it is crucial for data owners to be cognisant of the significant impact the context and purpose of a planned data situation has on every aspect of an assessment of anonymisation practices.

Data flow mapping

While data flow mapping is not a panacea for GDPR-compliance, it is a useful tool to employ so that: gaps between the regulatory framework and how data are processed in practice are revealed; (potential) high-risk data processing activities are identified and risks can be mitigated; individuals are well-positioned to make good decisions about data processing; and, the areas that require further legal training and guidance are exposed.

After a data owner has mapped their data flows, the next step will be to create (i) a risk register, (ii) a catalogue of risks and (iii) associated control measures in order to demonstrate appropriate mitigation of risks.

Suggested further reading

- Article 29 Data Protection Working Party, “Opinion 04/2007 on the concept of personal data (01248/07/EN WP 136),” 20 June 2007. [Online]. Available: http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf. [Accessed 31 May 2018].
- --, “Opinion 05/2014 on Anonymisation Techniques (0829/14/EN; WP216),” 10 April 2014. [Online]. Available: http://www.pdpjournals.com/docs/88197.pdf. [Accessed 3 April 2018].
- --, “Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679 (17/EN WP 248),” 4 April 2017 (Adopted). [Online]. Available: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236. [Accessed 27 May 2018]
- Conference of the Independent Data Protection Authorities of the Bund and the Länder in Kühlungsborn on 9-10 November 2016, *The Standard Data Protection Model* (v1.0) <https://www.huntonprivacyblog.com/wp-content/uploads/sites/28/2017/04/SDM-Methodology_V1_EN1.pdf> [Accessed 27 June 2018].
- CNIL, Privacy Impact Assessment (PIA): Methodology (how to carry out a PIA) (edition of June, 2015)
- --, Privacy Impact Assessment (PIA): Tools (templates and knowldge bases) (edition of June, 2015)
- --, Privacy Impact Assessment (PIA): Methodology (edition of February, 2018) <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf> accessed 25 May 2018
- --, Privacy Impact Assessment (PIA): Templates (edition of February, 2018) <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-2-en-templates.pdf> accessed 25 May 2018
- Information Commissioner's Office (ICO), “Anonymisation: managing data protection risk code of practice,” November 2012. [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf. [Accessed 3 April 2018].
- --, “Big data, artificial intelligence, machine learning and data protection (20170904; v2.2),” (in particular Chapter 3: Compliance Tools, pp. 58-61) 3 March 2017. [Online]. Available: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf. [Accessed 5 April 2018].
- --, “Determining what is personal data (v1.1),” [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf.
- --, “Guide to the General Data Protection Regulation (GDPR)”, [Online]. Available at: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/. [Accessed 4 June 2018]. Including: “Lawful basis for processing,” [Online]. Available: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/#ib3. [Accessed 30 May 2018].
- --, “What is personal data? – A quick reference guide (V1.1),” [Online]. Available: https://ico.org.uk/media/for-organisations/documents/1549/determining_what_is_personal_data_quick_reference_guide.pdf. [Accessed 28 March 2018].
- J. Clark, “Legislative Comment - GDPR series: building a compliance programme,” Privacy & Data Protection, vol. 17, no. 3, pp. 7-9, 2017.
- J. Graves, “Data flow management: why and how,” Network Security, no. 1, pp. 5-6, 2017.

- LINDDUN Privacy Threat Modelling, [Online]. Available: https://linddun.org/index.php. [Accessed 31 May 2018].
- M. Elliot, E. Mackey , K. O'Hara and C. Tudor , "UK Anonymisation Network (UKAN): The Anonymisation Decision-Making Framework," 2016. [Online]. Available: http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf. [Accessed 20 February 2018].
- M. Hansen, M. Jensen and M. Rost, "Protection Goals for Privacy Engineering," 2015 IEEE Security and Privacy Workshops (SPW), San Jose, CA, USA, 2015, pp. 159-166. doi:10.1109/SPW.2015.13
- N. Fulford and K. Oastler , "People, processes, technology - a how to guide to data mapping," Privacy & Data Protection, vol. 16, no. 8, pp. 6-8, 2016.
- R. Hu, S. Stalla-Bourdillon, M. Yang, V. Schiavo and V. Sassone, "Bridging Policy, Regulation and Practice?A techno-legal Analysis of Three Types of Data in the GDPR," in *Data Protection and Privacy*, Hart Publishing, 2017, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3034261 [Accessed 4 June 2018].

Next steps

During the course of the programme, the toolkit is a living document. The final version of the toolkit is due to be submitted in December 2019 at the end of the programme. The final deliverable – D3.9. – will focus on the legal and privacy aspects of transnational, cross-sector data sharing in open innovation.

– – –

7.2.4 Transnational, cross-sector data sharing

Refer back to the main D3.9 report.

7.2.5 Data Pitch strategies for anonymisation, pseudonymisation and re-identification risk

Note: The following extract from the D3.1 Legal and Privacy Toolkit v1 report was delivered in June 2017. Co-ordinators: Professor Sophie Stalla-Bourdillon and Alison Knight. Quality reviewer: Open Data Institute (ODI). This extract was written during the early stages of the Data Pitch programme before the Data Pitch challenges tracks were finalised.

– – –

Anonymisation/pseudonymisation strategy under the Project

Two approaches have been considered under the project in relation to the sharing of data relating to persons by the Data Providers to the Participating SMEs, as outlined below:

Approach 1

One approach open to the Consortium considered at the planning stage would have been to require that all data relating to persons shared under the Project would be aggregated (pre-sharing) to the extent that no person-specific data (from which individuals can be singled out) remains in the datasets available for analysis by the Participating SMEs.

While this was considered as a possibility, the implicit trade-off between data perturbation and latent data value (the greater the degree of the former, the less the potential for the latter) is worth acknowledging. Thus, while minimising the risk of negative impact upon individuals that might arise when anonymised data is later processed, anonymisation techniques are typically applied to data in ways that allow value to still be extracted from it post-anonymisation. In reality, this means that more data value can often be extracted when individual-level granularity is left in datasets for experimentation.

Approach 2

An alternative approach open to the Consortium which we believe to be legitimate - as a means to balance the requirements of data protection law with the need for preserving data utility, and therefore provide opportunity for value extraction from the secondary reuse of data relating to persons - is to allow pseudonymised data to be shared under the Project in certain cases that justify this approach. However, as mentioned, data relating to persons from which direct identifiers have been removed through pseudonymisation are still likely to be characterised as personal data under EU data protection law, unless effective measures are put in place to overturn this presumption.

Therefore, pseudonymised data that is proposed to be shared under the Project would only be acceptable for use in the Project when accompanied by the putting into place of appropriate safeguards (legal, technical, and organisational steps) for reducing re-identification risk to an adequately safe level. In particular, these should reduce the risk of Participating SMEs from re-identifying the data subjects in the datasets shared with them to an extremely low level. Putting in place such re-identification risk mitigatory measures will also facilitate compliance with the GDPR if it were later deemed to apply by a national data protection authority or court.[132]

At the same time, it is recognised that – under this alternative approach - given the residual risk of re-identification, the act of anonymisation/pseudonymisation and supporting measures cannot be treated as a one-off exercise. As re-identification risks can change over time, re-identification risk assessments and management of the results of such assessments should be applied iteratively during the life of the Project. In other words, a 'release and forget' approach to previously personal data that is not appropriate insofar as safeguards to mitigate re-identification risk must be kept in place during the lifetime of the experimental phase of the Project (after which time Participating SMEs would be required to delete data shared with them).

Accordingly, and considering the diversity of potential situations, a 10-point multi-factor strategy under this approach has been developed and is set out below. The strategy is to be read alongside the next section which contains practical guidance more generally for both Data Providers and Participating SMEs, including in scenarios where pseudonymised data is not relevant. This strategy will help to ensure both that re-identification becomes 'reasonably' impossible from any pseudonymised datasets for sharing, and also that any secondary processing of such datasets by a SME recipient would be compliant with the GDPR:

1) Technical and organisational measures should be put in place to make sure Participating SMEs do not have access to the additional information required for recovering direct identifiers. For example, Participating SMEs will be bound by confidentiality obligations and restrictions on reuse and re-identification.

2) The pseudonymised datasets for sharing with the Participating SMEs should be stored on company-secure servers, or on the servers of the University of Southampton. Where 'keys' that would enable the linking back of real-world identities to pseudonyms are retained, these should always be safely secured by the Data Provider using organisational and technical safeguards (to reduce the risk of illegitimate access to such information).[133]

3) Data Providers should only share data relating to persons to the extent that it is necessary to achieve a previously-delineated purpose or purposes.

4) Moreover, any indirect identifiers in the pseudonymised datasets for sharing should be removed or masked where these would not be strictly necessary for the participating SME to achieve the specified purpose(s). Similarly, possibilities for linking or inferring new information about data subjects from analysing the datasets – which could increase their risk of re-identification by the SME – should be muted as far as strictly necessary relative to the specified purpose(s).

5) The Participating SMEs should only be permitted to process the pseudonymised datasets for sharing for a specified analytics-driven purpose(s).

[132] In other words, this approach also takes into account that, while pseudonymisation (alone) might not transform personal data into non-personal data, it serves a useful security purpose (recognised by the GDPR) as a legitimate way to minimise the likely privacy harm that might befall data subjects when their personal data are processed. Therefore, its use is to be encouraged generally as part of a data protection by design approach as described in section 4 [of original report].

[133] This security mechanism would be strengthened through the legal terms of the contract agreed with them (e.g. prohibiting them contractually from sharing the key under any circumstances).

6) Where that specified purpose(s) does not relate to scientific research (technological development and demonstration, fundamental research, applied research, or privately funded research), it should be compatible with the initial purpose(s) for which the data was originally collected. Obtaining a clear description of the initial purpose and the legal basis justifying the initial collection and the challenge to be solved will help in assessing this compatibility.

7) As data subjects have a right to object to secondary processing on data relating to them in certain circumstances, data subjects should be informed of what is being proposed where the scope of the initial consent obtained from them does not extend to the specified purpose(s). Otherwise, another legal basis would be required to justify the processing, such as the 'legitimate interest' basis. This would requiring analysis by the Data Provider regarding whether the processing is necessary for the purposes of the legitimate interests pursued by them, and are not overridden by the interests or fundamental rights and freedoms of the data subject which require personal data protection. The purpose of this balancing exercise is to prevent disproportionate impact on individuals. In practice, carrying out this exercise will also require a full assessment of the facts and context of each case as relevant.[134]

8) If the specified purpose(s) would involve data subjects being subject to 'profiling' analytics – and measures or decisions are subsequently planned to be taken vis-a-vis individual customers or groups of customers, based on profiles that might be created by the Participating SME – we advise that Data Providers carry out a DPIA. This would again require a full assessments of the facts and context of each case, and should be done before the measures or decisions are taken (and, if possible, even earlier before the Participating SME receives the relevant data from them).[135] Any high risks of impact to data subjects identified under a DPIA should then be addressed through the Data Provider taking different kinds of safeguards to mitigate such risks in proportion to the level of harm predicted. This might include technical and organisational measures taken to ensure 'functional separation' through data silo'-ing (i.e. data used for research purposes not being made available to support decisions that could be taken with regard to individual data subjects). Notwithstanding, when an organisation specifically wants to analyse or predict personal preferences, behaviour and attitudes of individuals from data they propose sharing - which will subsequently inform 'measures or decisions' that are taken with regard to those customers - free, specific, informed and unambiguous 'opt-in' consent should be obtained.[136] Ultimately, if these safeguards cannot be met, the Data Provider would have to rethink the data being made available for reuse and/or the specified purposes of that reuse.

9) Participating SMEs will be required to comply with, and provided training on compliance with, data protection law under the GDPR. They will be required to comply with data protection law and prohibited from re-identifying data subjects.

10) Participating SMEs will also be required to destroy the data at the end of the acceleration stage of the Project in which they are involved. Further, publication of research results should, as a rule, be possible in such a way that only aggregated (and/or otherwise fully anonymised) data will be disclosed.

Mosaic effects

So-called 'mosaic effects' denotes a broad concept underpinned by the idea that data may be linked to other information (to create a 'mosaic') in ways that increase privacy risks (potential 'effects') to those individuals about whom the data relates. For example, data techniques can reveal intimate personal details because of opportunities for data fusion – the merging of data, in particular across disparate datasets and information sources. Like joining together different pieces of a jigsaw puzzle,

[134] To note, that depending on their capacity, some national data protection authorities are happy to assist with, or assess results of, this exercise.

[135] In other words, it is recommended that such a robust and detailed impact assessment should be completed *prior* to the disclosure of information and making it available for reuse.

[136] For the consent to be informed, and to ensure transparency, data subjects should ultimately also be given access to 'profiles' relating to them, as well as to the logic of the algorithm that led to the development of the profile. Furthermore, the source of the data that led to the creation of the profile should also be disclosed and the decisional criteria.

for example, data analytic techniques are specifically aimed at merging multiple data sets to reveal complex patterns and infer new knowledge. Where data relates to persons, such fusion might permit the revelation and inference of new information in terms of what can be concluded about those people (potential **privacy attribute disclosures**).

In the context of data to which techniques have been applied to personal data to hide the identity of the data subject, the possibility of data fusion also has relevance because it may also be possible to work out who that person is by joining together the modified data with other auxiliary information (potential **identity attribute disclosures**). In that context, mosaic effects has been defined as *"[t]he process of combining anonymous data with auxiliary data in order to reconstruct identifiers linking data to the individual it relates to"* [99, p. 7].[137] Thus, personal details may be discerned or risk becoming discernible even from ostensibly 'anonymous' data through linking and combining multiple data points about the same person. The underlying presumption in such scenarios being that individual identifiers in these datasets would not otherwise allow a data subject to be re-identified. This is especially relevant to large scale data sharing and repurposing because seemingly 'anonymous' data sets can often be combined and analysed to reveal restricted or sensitive information. In other words, rather than the information shared being sensitive, sensitivity may lie in the inferences that are drawn from its processing and the way in which those inferences are drawn, potentially giving cause for concern.

Such concerns are especially topical in the context of research around 'big data' analytics and re-identification risk, such that this theme has been considered to deserve explicit consideration and treatment in the Project. Such trending concerns are because large scale data analytics involving sharing and repurposing of seemingly 'anonymous' datasets can often result in the revelation of new personal details about a person to whom they relate. This is despite the fact that individual identifiers in information that has been subject to anonymisation techniques would not – in and of themselves – enable its subject to be re-identified.

In legal terms, as alluded to, this issue of a potential mosaic effect arises alongside the issue of determining whether data protection law – which applies to any processing of personal data relating to living persons – would in fact apply to a specific data processing activity. There is concern particularly for data analytics that personal data may be discerned from data despite efforts being made to anonymise personal data by stripping it of identifiers. Consequently, data protection rules would then apply to the processing of this data set, which may not have been envisaged when it was planned to be processed. In broad terms, this would mean that organisations processing de-identified data typically could not use that data for purposes beyond those for which it was originally obtained where these purposes are incompatible, and that data could not be kept indefinitely.

Mosaic effects mitigation strategy under the Project

For the Project, concerns could be raised that mosaic effects might arise either as a result of integrating different data sets being shared under the Project and finding commonalities, or by linkage of individual datasets to other sources of information that are outside the control of the Consortium. As mentioned, in particular, there is a risk that data protection law might consequently apply in relation to all challenges where the processing of data relating to persons is involved where individuals can be singled out from the relevant data shared. Special concern is also reserved for the possibility that sensitive personal data might be inferable from specific datasets. As mentioned, categories of sensitive personal data raise a high level of concern, such that a stricter standard of data protection obligations should always apply to their processing.

[Paragraphs omitted] […]

Specifically, the following risk-scenarios from which mosaic effects might flow from these procedures are concerning:[138]

1. The possibility of combining data about people within a Data Provider dataset with external sources of information, including information brought to the Project for analysis in combination with

[137] Compare a similar non-UK specific definition, "*whereby personally identifiable information can be derived or inferred from datasets that do not even include personal identifiers, bringing into focus a picture of who an individual is and what he or she likes*" [196, p. 8].

[138] The separate concern that about combining data about people within the same Data Provider dataset whereby linkages could be made to multiple data points in relation to a singled out individual has been dealt with in the above section under pseudonymisation strategy.

the Data Provider dataset. This could include data collected by the Participating SME, data that has been shared with the Participating SMEs by third party data providers, and/or data made publicly available, in relation to the same singled out individual.

2. The possibility of combining data about people within a dataset provided by a Participating SME with external sources of information, including data publicly available[139] as well as closed datasets made available to the Participating SMEs by third party data providers, in relation to the same singled out individual.

The strategy adopted to address the possibility of 'mosaic effects' arising at the acceleration stages of the project, as described, encompasses anonymisation and pseudonymisation best measures as outlined above, the latter to be applied to data from which individuals can be singled out under the Project in any event.[140] Following this approach should reduce some types of mosaic effects to a low level, including in particular the imposition of contractual obligations, such as confidentiality, on the Participating SMEs.

Also, as part of the overall, comprehensive risk-management approach adopted under the Project to reduce –not just re-identification risk but also - the likeliness and severity of any harm potentially caused by the secondary processing of data to its subjects, the following notification procedure is in place. In the event that a Participating SME wishes to provide datasets for use under the Project, it will be obliged to provide a description of each such dataset to the Consortium. Such description will include: the original and ownership of such data; the rights that the Participating SME has to use such data; and, whether such data has been subject to any anonymisation or pseudonymisation procedures. Following such notification, the Consortium will assess the potential for mosaic effects arising from the combination of such data and the Data Provider's dataset, or arising from the combination of discrete datasets that the Participating SMEs wishes to self-supply. The notification will only be rejected if such effects are deemed reasonable likely to arise (e.g. if such data sources might conceivably relate to the same individuals).

Turning theory into practice

As mentioned, the legal framework for sharing and re-using data presents as a complex picture, giving rise to challenges in assessing and managing associated risks. In particular, the legal implications of a third party using data in relation to which rights already exist – and for purposes other than that for which it was originally obtained – are key issues. This is because:

- Different types of law act concurrently in relation to a data sharing arrangement (e.g. IPR as database right or copyright), contractual rights and duties, and data protection regulation where personal data is involved or likely to be involved in secondary processing of data relating to persons). In other words, legal rights and duties arise in a multi-layered way and may differ between countries (including within the EU bloc).
- The requirements for legal and regulatory compliance, on the one hand, and innovative efficacy on the other hand, are sometimes hard to reconcile.

The identification of a separate data handling protocol for Data Providers and Participating SMEs is valuable and this is set out in guidance form in sub-sections […] below. In this first version of the toolkit, the issue of data protection risk mitigation is the key focus, supplemented by contractual provisions (whereby the Consortium imposes contractual requirements on Data Providers and Participating SMEs to ensure that they act in compliance with legal analysis already carried out based on EU best practices).

To note for the Project, data will only be permitted to be shared with Participating SMEs once a legal review has been carried out, and a legal agreement signed by, Data Providers. This will require individual discussions with the Consortium about defining the conditions for the data's reuse in light of legal compliance, and setting out the scope of the purposes for which each dataset will be

[139] To note, the mere fact that data relating to persons has been made publicly available does not lead to an exemption from data protection law. The reuse of such data remains subject, in principle, to data protection law if it is personal data.

[140] In other words, data that has been pseudonymised that is proposed to be shared under the Project would only be acceptable when accompanied by the putting into place of appropriate safeguards (legal, technical, and organisational steps) for reducing re-identification risk to an adequately safe level.

processed (and only processed) under the Project.

Data sharing methodology for Data Providers - managing the legal risks associated with data sharing in practice

Data Providers will need to review the types of data that they propose to share with the SMEs in advance of sharing to assess what types of legal issues arise in relation to that data – in particular, whether or not such data falls within the definition of personal data or not. When anonymising personal data, organisations should assess the risks that the data could be re-identified given the nature of the data, the context in which the data will be used, and the resources available to those with access to the data.

Where they consider that they data types do fall within the definition of personal data, or there is a reasonable risk that they might be construed as personal data, consideration must be given to:

- How to reduce the risk of re-identification of the data to be shared; and,
- If it is not possible to reduce that risk, or where pseudonymised data must be shared in order to achieve the challenge they set for the SME in relation to the data they provide, how to ensure that the requirements of the GDPR would be met upon reuse and how to minimise any risk of harm to data subjects flowing from the secondary processing. These could include compliance with extra requirements, such as performing a DPIA.

Best practice recommendations are suggested in the next sections to be followed by Data Providers in respect of the stages of consideration of data protection risks.

Assessing the effectiveness of – and managing - anonymisation/pseudonymisation techniques

Before the acceleration stage of the Project where the SME start to utilise the data, there are two, interlinked areas being assessed under the Project to reduce re-identification and harm risk flowing from the further processing of shared data when it relates to persons. These involve the Data Providers: ensuring the implementation of effective anonymisation practices (bearing in mind that no anonymisation technique is a fail-safe solution to the problem of re-identification risk); as well as putting in place adequate non-technical elements (organisational and legal measures).

Each Data Provider will be asked to complete a Consortium-prepared questionnaire (a copy of which can be found in Annex B [to original report] below), aimed at, inter alia, determining what data relating to persons is intended to be shared and the extent of the risks that might arise if such data were processed under the Project. A key section of that questionnaire relates to determining what anonymisation techniques have been applied to data relating to persons intended for sharing by the Data Providers, together with any other steps they have taken to reduce re-identification risk (such as the secure storage of any 'key' that would be required to back-track a pseudonym to a recognisable individual).

Responses to such questionnaires are analysed by the relevant Consortium member against EU guidance on best standards for ensuring the effective anonymisation of personal data and associated steps that can be taken to reduce the risk of data subjects being re-identified from transformed data (taking into account the 'means test' standard for determining 'identifiability' as set out in Recitals 26 of the Directive and the GDPR (see sub-section 5.1)).

In particular, as described by the Art.29 WP in 2014 in its Opinion on Anonymisation Techniques ('WP216' [63], summarised in Annex C below [in original report]), the adequacy of the anonymisation solution applied to data by Data Providers will be considered in terms of the risks of any future data recipients being able to:

- single out an individual in a dataset;
- link two records within a dataset with respect to the same individual; or
- any information in such dataset about an individual (in particular, the inference of sensitive personal data about them).

This analysis also addresses the possibility of mosaic effects that might arise where different data sets may be combined, or shared data linked to other unknown external sources that are out of control of the Data Providers. Another recommended piece of regulatory guidance is advice given by the UK ICO in particular in its Anonymisation Code of Practice [73] (summarised in **Annex D** below [in original document]).

Data Providers should consider the above guidance on a case-by-case basis relevant to the datasets they propose to share and the pre-set challenges (specified objectives) attached to each dataset, alongside possible harm mitigation measures that can be taken. In particular, the amount of linkability compatible with a challenge determines the anonymisation (or pseudonymisation) technique appropriate to be applied.

We also recommend a dynamic approach to anonymisation/pseudonymisation that involves consideration of the relevant data environment. In other words, Data Providers should assess the risks that the data could be re-identified given the nature of the data, the context in which the data will be used, and the resources available to those with access to the data. In particular, the purposes to be achieved by way of the processing of modified dataset should be clearly set out not least because they play a key role in determining the likelihood of residual re-identification risk arising. Such risk should then be reduced to an acceptable level.

It is an important consideration that the weaker the level of anonymisation techniques applied to personal data (so-called 'sanitisation', or indeed where only pseudonymisation is applied to such data), the stronger organisational and legal measures are needed to ensure that the overall risk levels associated with secondary processing remain low.

An overview of different types of de-identification methods

It is very important to take great care, at the initial stage of producing, disclosing and making available for reuse, any information derived from personal data. Several techniques can be considered for application to de-identify personal datasets and reduce any risks of re-identification of an individual prior to making the data available for reuse. As discussed above, full or partial anonymisation can be relevant to the safe use or sharing of data between organisations. When full anonymisation and use of aggregated data (at a sufficiently high level of aggregation – the most definitive solution to minimising the risks of inadvertent disclosure) is not possible, data will often at least need to be partially anonymised and additional safeguards may also be required, as discussed below.

It is helpful to distinguish different scenarios for further analysis ranked according to the strength of privacy protection provided (from the weakest to the strongest):

- **Scenario 1:** situations where directly identifiable personal data are needed due to the nature of the research and other solutions are not possible without frustrating the purpose of the processing, and further provided that other appropriate and effective safeguards are in place). This scenario is not outside the scope of the Project.

- **Scenario 2**: situations involving indirectly identifiable personal data: lower level of aggregation, partial anonymisation, pseudonymisation or key-coded data.

- **Scenario 3:** unidentifiable personal data: data are anonymised or aggregated in such a way that there is no remaining possibility to (reasonably) identify the data subjects.

The robustness of each technique is considered taking into account the three re-identification risk categories described by the Art.29 WP WP216 reformulated as the following questions:

(i) is it still possible to single out an individual?

(ii) is it still possible to link records relating to an individual?

(iii) can information be inferred concerning an individual?

Scenario 2

This scenario covers situations of partial de-identification where full anonymisation is not practically feasible due to the nature of the processing (e.g. where there is a need to use more granular data which as a side effect, may allow indirect identification). It takes into account that full anonymisation is increasingly difficult to achieve with the advance of modern computer technology and the ubiquitous availability of information. Thus, re-identification of individuals is an increasingly common and present threat.

In practice, there is a very significant grey area, where a data controller may believe a dataset is anonymised, but a motivated third party will still be able to identify at least some of the individuals from the information released. Furthermore, data controllers must be aware that the risk of re-identification can change over time, (e.g. powerful data analysis techniques that were once rare are now commonplace). Addressing and regularly revisiting the risk of re-identification, including identifying residual risks following the application of the following techniques, therefore remains an important element of any solid approach in this area. Good practice requires organisations to carry out a periodic review of their policy on the release of data and of the techniques used to anonymise it, based on current and foreseeable future threats. Notwithstanding a good practice approach, there remains the risk that data protection law continues to apply where such datasets are processed and this risk will need to mitigated by additional safeguards beyond the de-identification techniques described under the next sub-heading.

<u>Removing identifiers</u>

The first step of de-identification is to remove clear identifying variables from the data (name, date of birth or address).

Although some identifiers are removed, datasets may still retain a relatively high potential for re-identification as the data still exists on an individual level and other, potentially identifying, information has been retained. For example, some address postcodes have very small populations and combining this data with other publicly available information, can make re-identification of data subjects living in such postcodes a relatively easy task.

<u>Key-coding</u>

The technique involves consistently replacing recognisable identifiers with artificially generated identifiers, such as a coded reference or pseudonym (e.g. a randomly selected number). This allows for different information about an individual, often in different datasets, to be correlated.

This method has a relatively high potential for re-identification, as the data exists on an individual level with other potentially identifying information being retained. Also, because key-coding is generally used when an individual is tracked over more than one dataset, if re-identification does occur more personal information will be revealed concerning the individual. For that reason, key-coding is not considered a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.

<u>Reducing the precision of the data</u>

Rendering personally identifiable information less precise can reduce the possibility of re-identification. For example, dates of birth or ages can be replaced by age groups.

Related techniques include suppression of cells with low values or conducting statistical analysis to determine whether particular values can be correlated to individuals. In such cases it may be necessary to apply the frequency rule by setting a threshold for the minimum number of units contributing to any cell. Common threshold values are 3, 5 and 10.

Introducing random values (‘adding noise’) is more advanced and may also include altering the underlying data in a small way so that original values cannot be known with certainty but the aggregate results are unaffected.

Various other techniques (such as keyed-hashing, using rotating salts, removing outliers, and replacing unique pseudonyms) may also be used to reduce the risk that data subjects can be re-identified, and subsequently, that any measures or decisions can be taken in their regard.

Scenario 3

Full anonymisation (including a high level of aggregation) is the most definitive solution. It implies that there is no more processing of personal data and that data protection law is no longer applicable. However, it is a high standard to meet, in particular because it requires that any reasonable possibility of establishing a link with data from other sources with a view to re-identification be excluded.

<u>Aggregation</u>

Individual data can be combined to provide information about groups or populations. The larger the group and the less specific the data is about them, the less potential there will be for identifying an individual within the group. An example is aggregating postcodes at a regional level.

Importance of additional risk assessment and mitigation measures (safeguards and controls) beyond de-identification techniques

The above analysis shows that applying techniques to personal data is a key tool in achieving different levels of de-identification, but most of these procedures have their challenges and limits. The analysis also shows that once a first assessment has been completed in terms of the possibilities and limits of effective de-identification, a second step of assessing the need to complement these techniques with additional (organisational and technical) safeguards will often need to followed.

This second stage is about applying these other safeguards in order to adequately protect the data subjects. Risks flowing from secondary processing of data relating to persons must be reduced to an acceptable level after they are identified to prevent re-identification and reduce the future likelihood of harm. Safeguards may involve technological, legal, and administrative measures, and often a combination of each.

As an essential guide, it should be kept in mind that the easier it is for the data subject to be identified, the more additional safeguards will be needed. Furthermore, the more consequential potential adverse impact on the data subject if identified would be, the more should be done to limit the possibilities of re-identification and the more additional safeguards may be required.

Among the appropriate safeguards which may bring additional protection to the data subjects, the following could be considered and applied to the data provided by each Data Provider in the Project:

- taking specific additional security measures (such as encryption);
- in case of removing identifiers and key-coding data, making sure that data enabling the linking of information to a data subject (the 'keys') are themselves also coded or encrypted and stored separately to reduce the risk of illegitimate access to such information.[141]
- entering into a trusted third party arrangement; this model is increasingly being used to facilitate large-scale research using data collected by a number of organisations each wanting to anonymise the personal data they hold for use in a collaborative project. Trusted third parties can be used to link datasets from separate organisations, and then create anonymised records for researchers to work on.
- restricting access to personal data only on a need-to-know basis, carefully balancing the benefits of wider dissemination against the risks of inadvertent disclosure of personal data to unauthorised persons. This may include, for example, allowing read-only access on controlled premises. Alternatively, arrangements could be made for limited disclosure in a secure local environment to properly constituted closed communities. Legally enforceable confidentiality obligations placed on the recipients of the data, including prohibiting publication of identifiable information, are also important.

To note, in high-risk situations, where the inadvertent disclosure of personal data would have serious or harmful consequences for individuals, even the strongest type of restriction may not be suitable. That is because, even with safeguards in place, these would not be considered adequate to prevent undue impact on the data subjects.

De-identification and impact mitigation checklist

The following points should be considered:

- Determining of any specific unique (such as national identity number), or quasi-unique (such as names, date of birth), identifiers in the relevant data.

[141] As mentioned, the formal legal definition of 'pseudonymisation' under the GDPR (Article 4(5) describes it to mean, "*'the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person".*

- Cross-referencing to determine unique combinations like age, gender, and postcode.
- Acquiring knowledge of other publicly available datasets and information that could be used for list matching.

Even with such variables missing other factors should be considered:

- Motivation to attempt identification
- Level of details (the more detail the more likely identification becomes)
- Presence of rare characteristics
- Presence of other information (a dataset itself may not include any data that can identify an individual, but it may include enough variables that can be matched with other information).
- What a potential breach of privacy could mean for related individuals. More specifically, the level of potential impact will be dependent on the likelihood that identification could occur from the sharing of data but also the consequences of such sharing in terms of what it will be reused for. Factors that will help assess the level of likely impact to result from a processing activity include consideration of: the purposes of the proposed secondary processing; the proportionality of the secondary processing operations proposed in relation to such purposes; an assessment of the risks to the rights and freedoms of data subjects; and, any measures envisaged appropriate to address such risks.
- Any challenge set by a Data Provider requiring data analytics involving profiling should be considered carefully. In practice, it would need to be presumed that any secondary processing activities on data relating to persons involving profiling would require not just compliance with the GDPR (as presumed wherever there is a risk that personal data are to be processed under the Project), but also extra caution. In particular, obligations regarding the provision of information to the data subject in respect of the carrying out of profiling activities are more onerous compared with other processing activities (as described in section 4.3.1.1 above). Thus, consents and notices for profiling should be reviewed carefully. Moreover, the GDPR requires controllers to conduct a DPIA any time "*a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, **including profiling**, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person*." In this context, a data controller subject to the obligation to carry out the DPIA *"shall maintain a record of processing activities under its responsibility"* including the purposes of processing, a description of the categories of data and recipients of the data and *"where possible, a general description of the technical and organisational security measures"*. Data Providers will have to carefully consider whether a DPIA is needed for this Project from 25 May 2018.
- **What profiling activity could mean for related group interests.** Although this may not be so relevant to this project, for completeness, if data controllers plan to make decisions at a group-level (i.e. potentially affecting all individuals within a group), likely impact on collective interests should also be considered by them.[142] Even without decisions being planned in respect of a singled-out group, there may still be inherent risks of discrimination, or other negative impact liable to flow, from the making of profiling assumptions (that is, the detecting of patterns in data between a particular type of person grouped into various categories by profilers, and a type of behaviour or characteristic). We recommend such risks of harm also be considered in carrying out DPIAs, such as the possibility of incorrect assumptions being made about individuals categorised in line with an inferred group profile.

Other safeguards under the Project where data relating to individuals are to be shared

Where personal data are processed, data protection law requires that such data must be obtained only for specified lawful purposes and not further processed in a manner incompatible with those (original) purposes.

[142] For guidance on how privacy and data protection may be interpreted as referring to collective interests, see e.g. [197].

Data Providers will be asked to share details in response to the Consortium's questionnaire about the lawful purpose for which they currently process any personal data to be anonymised/pseudonymised and shared under the Project. Potentially personal data that does not have a lawful purpose (under the Data Protection Directive (Article 7) and the GDPR (Article 6)) will not be shared under the Project. Regarding the lawful reuse of anonymised/pseudonymised data (assuming it were in fact deemed to be personal data, such as by a data protection authority in a relevant EU Member State, or a court), where relevant any consents given by the data subjects to the original processing activity will be considered.

However, relying upon consent as legal basis for secondary purposes under the Project is recognised as challenging. This is in light of the restricted definition of 'consent' under the GDPR – in reminder (see footnote 5), it is defined as meaning a "*freely given, specific, informed and unambiguous indication of the data subject's wishes" signifying agreement to the processing of personal data relating to him or her"* (Article 4(11)). Whereas, consent will *not* be presumed to be freely given according to Recital 43 GDPR unless separate consent opportunities for different personal data processing operations have been proffered. These requirements are unlikely to be satisfied where pre-obtained data sets are repurposed to unlock yet unrevealed value (such as in the case of exploratory challenges). In other words, it is not easy to implement the principle of purpose limitation effectively in cases where processing purposes cannot be clearly defined or clearly foreseen as consent will not be presumed. An alternate legal basis to justify secondary personal data processing must be considered (see the related discussion in section 4.3.1.1 above [in original document]), along with the establishment of appropriate data protection safeguards.

For the avoidance of any doubts, in the event that data protection law applies to the processing carried out during the acceleration stage, the contracts entered into for the Project will also require that the Data Providers acknowledge their obligations as data controllers, including in respect of those who process (potentially) personal data on their behalf, and the Participating SMEs acknowledge that they may be construed as data processors (or, indeed, as joint data controllers of such data).

Assessing the adequacy of – and managing - data security measures

Article 32(1) GDPR follows a risk-based approach regarding imposing an obligation on data controllers to ensure the security of personal data processing. When assessing the appropriate level of security required, the data controller must take into account the risks that are presented by a relevant processing activity (in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed). This reflects the new risk-based approach taken by the GDPR in general to data protection law compliance.

Article 32(1) also sets out a number of specific "appropriate technical and organisational measures" controllers and processors could take, by way of suggestions for the contents of a data security policy. These are:

- The pseudonymisation and encryption of personal data.
- Measures ensuring the ability to ensure the ongoing confidentiality, integrity, availability and resilience of systems and services processing personal data.
- Measures ensuring the ability to restore the availability and access to data in a timely manner in the event of a physical or technical incident.
- A process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

Regarding security measures to be put in place under the Project, Data Providers will be asked to complete a technical questionnaire so that the Consortium can suggest the most appropriate option for the hosting of their (personal or non-personal) data. Such suggestions will be framed in light of the relevant contextual elements under consideration from which an assessment of any attendant risks (their type, likelihood and severity) will be carried out.

Four possible options for storing and providing access to data sets shared by the Data Providers by the Participating SMEs have been considered:

1) The University of Southampton hosts the data. Multiple levels of security (data hosting facilities using secure infrastructure managed by the Consortium at the University of Southampton) can be offered depending on what is deemed appropriate to the level of re-identification assessed to be present. These include: connection to the public internet through Janet, the UK university network; or, defining a secure zone Data Pitch via a standard proxy - an academic infrastructure hosting all the academic services and laboratories (this can be supplemented by the addition of a further proxy to create an isolated and secure zone for Data Pitch processing which is managed independently).[143]
2) The Data Provider hosts the data.
3) The commercial cloud hosts the data (under the direction of the Data Provider).
4) The participating SME (chosen to process the relevant data) hosts it.

Ultimately, the Consortium realises that the option of choice for the secure provision of data relating to individual to be shared by the Data Providers will be led by their preferences and their current legal compliance measures taken as data controllers. However, the Consortium will advise the Data Providers in their choices as far as possible, in particular when pseudonymised data is shared, to ensure that they are aware of the extent of their legal compliance obligations under existing and incoming data protection rules.

Options (1) or (2) are the Consortium's preferred option for hosting data relating to persons under the Project.[144] This is because of the risks in the other options as set out below:

- **Option (3) -** Hosting potentially personal data in a commercial Cloud carries security and other data protection risks.[145] While the Data Provider may consider that such risks are mitigated sufficiently when they manage the private Cloud storage themselves – in accordance with their duties as data controllers of such data under data protection law – the risks associated with the use of professional Cloud providers are normally not so mitigatable. For that reason, we are less inclined to recommend that Data Providers share access to potentially personal data via – for example non-EU (e.g. US) - cloud providers, in particular where it can be assumed that they use servers based outside the EU. Alternatively, recommend that the data provider has carried out a DPIA in this respect which they can evidence to the Consortium, in particular in respect of where data may be stored outside the EU (including regarding data 'adequacy' arrangements with non EU countries).
- **Option (4) –** This is discouraged as a means to host potentially personal data, unless the chosen SME already has a strong data analytics infrastructure, and a DPIA exercise has been carried out and can be evidenced to the Consortium with respects to the adequacy of the security over data currently hosted. At the very minimum, the Participating SMEs must

[143] The most secure level - hosting on a separate, BIL3 and ISO27001-compliant network in the University's Data Centre with a requirement for physical authentication for network access – has been discarded. Fulfilling this requirement is considered inappropriate because the Participating SMEs will be based around Europe. Furthermore, this type of physical-only access (and workstation isolation) is deemed suitable for the processing of highly sensitive data, of the type which will not be allowed to be shared under the Project.

[144] Under option (2), the Consortium also recommend that access to data relating to persons only be given to authorised persons at the Participating SMEs, as well as the adoption of passwords and other relevant further security mechanisms against data leaks or intrusion, such as access solely through APIs and encryption. To note, the ICO Blog: Why encryption is important to data security provides a useful discussion on how encryption works and the importance of identifying suitable forms of encryption techniques, bearing in mind the sensitivity of the data and how it will be stored and processed. It cites - by way of example - that the encryption techniques used for data storage will differ to those used for transferring personal data. To note, as security, like privacy, is an on-going process, not a one-time project, it is conceivable that a Data Provider will be required to re-evaluate access conditions during the course of the Project.

[145] The Art.29 WP [199, p. 19] has made a number of recommendations, which it describes as "a checklist for data protection compliance by cloud clients and cloud providers" (Opinion 5/2012 on Cloud Computing […]). It explains that these arise primarily from the fact that cloud computing clients relinquish exclusive control of their data, which can lead to: lack of portability and interoperability; loss of integrity of the data due to sharing of cloud resources; disclosure of data to law enforcement agencies outside the EU in contravention of EU data protection principles; loss of ability to intervene owing to the complex chain of outsourcing; inability of the cloud provider to help the controller respond to data subjects' requests; possibility that the cloud provider might link personal data from different clients; and, lack of transparency about how the cloud service intends to process data, so that the controller cannot take proper measures to ensure data protection compliance. The danger is considered increased if the client remains unaware that the cloud service involves a chain of processors (each subcontracting to the next), that processing is being conducted in different countries (which will determine the applicable law in the event of a dispute), or that the service will result in the transfer of data to countries outside the European Economic Area. Related useful guidance has been produced by the ICO here: [198].

have control mechanisms in place and evidence an organisational culture that encourages compliance with data protection law (see the next section on Participating SME requirements in relation to data protection law).

Where the Consortium feels that the re-identification risk presented in a particular case justifies caution, the Consortium's advice regarding the use of options (3) and (4) will be strict, based on a number of factors relating to technical and legal constraints, as well as the possibility of external attacks bearing in mind the nature of the data intended for sharing and its potential sensitivity. Furthermore, as described in section 5 [of the original report], this advice will be backed up by the 10-point strategy approach where pseudonymised data is shared.

Finally, at no time will any data relating to individuals be released (in whole or part) in the public domain under the auspices of the Project except in accordance with data protection law. Moreover, a reasonable level of documentation will be required to demonstrate that adequate security controls in place.

[Two blank sections omitted].

Data reuse methodology for Participating SMEs – managing the legal risks associated with data reuse in practice

Participating SMEs will be required to show understanding of the different types of legal issues that might arise in relation to the data that is to be shared with them, as well as sign a legal agreement with the Consortium obliging them to observe conditions for the data's reuse in light of legal compliance and setting out the scope of the research purposes for which each dataset will be processed (and only processed) under the Project.

Set out below are the Project's best practice recommendations in respect of this exercise. In this first version of the toolkit, they focus on data protection law compliance and related risk mitigation.

Data protection law compliance

Steps will be taken by the Consortium – via this toolkit, associated support, and the legal agreement that they will sign – to raise Participating SMEs' awareness of EU data protection law requirements in case it becomes relevant.

Furthermore, the Consortium will require all SMEs applying to take participate in the Project to sign an Ethics Statement and Declaration of Honour as part of the application process in line with H2020 guidelines. These will refer to the principles that apply to the processing of personal data that must be followed. Such declarations will also form part of the contract signed with the Consortium by those Participating SMEs who are successful in their applications, which can later be checked during the milestone reviews or in the final graduation.

Data controllers or data processors?

Data relating to persons shared by the Data Providers with Participating SMEs must first be subject to anonymisation/pseudonymisation techniques. However, there is a risk that even data so modified may yet still be deemed personal data.

For the avoidance of any doubts, in case data protection law applies to the processing of such data carried out during the acceleration stage, the Participating SMEs must acknowledge that they may be construed as data processors,[146] but also potentially – in the alternate - as joint data controllers of any personal data shared with them. Therefore, the responsibilities of both processors and controllers must be understood.

Where a participating SME is acting in a data processor role for a Data Provider, the Data Provider will be obliged under the GDPR to enter into a contract directly with them to ensure that they follow their data protection law obligations. For example, SMEs must offer sufficient guarantees to

[146] To note […] the GDPR introduces direct compliance obligations for processors. Whereas under the Data Protective Directive processors generally are not subject to fines or other penalties, under the GDPR they may be found liable to pay fines (including, as mentioned, fines of up to €20 million or 4% of global annual turnover for the preceding financial year, whichever greater, where they are directly responsible for a data protection law breach).

implement appropriate technical and organisational measures when they may be dealing with personal data.[147]

Higher standards of data protection are required when the Participating SME would be acting in the role of data controller. This role may be assumed, in the context of the Project, in relation to any personal data that they are permitted (by the Consortium) to analyse alongside the shared data during the acceleration stage. To this end, Participating SMEs should consider whether they have assessed the legal risks (and compliance obligations beholden on them) associated with any data relating to persons that they currently hold. In particular, subject to some exceptions, organisations currently processing personal data should be registered (publicly) with the relevant data protection authority (such as the ICO in the UK) first. Details of such registration will be required by the Consortium within the legal agreement entered into between the Consortium and the Participating SMEs.

Key issues for consideration

The processing of personal data must comply with the data protection law principles set out in section 4 above and with the contractual terms set out in the legal agreement signed with the Consortium. These include conditions regarding what Participating SMEs are allowed to do in the reuse of data shared with them, such as the fact that the data shared under the Project must only be processed for the purposes of the Project. Furthermore, Participating SMEs must agree not to attempt to identify any data shared with them that has been subject to anonymisation/pseudonymisation processes.

Any data relating to persons will also only be shared within the context of specific data experiments where the broad objectives for the secondary processing has been scoped by a defined purpose or purposes (a challenge). These challenges will not require the Participating SMEs to take any decisions or measures regarding any particular individuals, and this activity should be avoided at all costs. If such decisions or measures were taken, the Participating SMEs will be required to inform the Consortium immediately.

It should be noted that, any data controller intending to carry out processing activities which are likely to result in a "*high risk*" to data subjects (taking into account the nature, scope, context and purposes of the processing), should carry out a DPIA. In certain circumstances a Participating SME may be a data controller and in that case the requirement for a DPIA should be assessed carefully. To do this a consideration of Article 35(3)(a) GDPR should be made which requires data controllers to conduct a DPIA any time "*a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person*."

Any processing by Participating SMEs of personal data would also need to be justifiable in light of a specified lawful purpose (legal basis) for its reuse, at least where such use would be deemed incompatible with the original purposes for which personal data was processed. Furthermore, all appropriate safeguards for the rights and freedoms of the data subject should be put in place, including technical and organisational measures proportionate to the processing aim pursued, to safeguard the fundamental rights and interests of the data subject.

Finally, all Participating SMEs will also be required to enter into a confidentiality agreement, either separately or as part of the legal agreement with the Consortium whereby they will not be permitted to disclose any data identified as confidential that is shared with them.

Managing data security

Assessing the level of security required

As previously stated, organisations should adopt a risk-based approach to deciding what level of security is needed when they process personal data. Thus, Participating SMEs acting as data controllers or data processors must ensure a level of security appropriate to both:

[147] Article 28(1), GDPR states: "*Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject*".

- the harm (damage or distress) that might result if there was a data breach (the unauthorised or unlawful processing or accidental loss, destruction or damage of personal data); and,
- the nature of the personal data held (for example, how valuable, sensitive or confidential it is).

To assess this, Participating SMEs should carry out risk assessments by way of a formal process to identify, document and manage security risks.[46] For example, such assessments would typically also include a review of the purposes for which personal data are processed internally, or to be shared via external mechanisms, as well as taking into account:

- The nature and extent of an organisation's premises and computer systems.
- The number of staff and the extent of their access to personal data.
- The state of technical development.
- The cost of implementing security measures.

Managing security risks

Once security risks have been assessed, Participating SMEs should take reasonable steps to mitigate significant risks. This requires ensuring, not only the correct physical infrastructure and technical tools, but also that appropriate data security policies and procedures are in place. To this end, Participating SMEs that are also data controllers are obliged to take reasonable steps to ensure the reliability of any employees who have access to personal data. In addition, Participating SMEs should identify internal individuals who are responsible for ensuring data security.

As part of any general registration requirement under data protection law, Participating SMEs that are also data controllers may be required to notify their national data protection agency of the security measures that they have put in place. This requirement for registration would be limited to one year. The Participating SMEs should also be able to respond to any data security breach swiftly and effectively.

Finally, further obligations are also placed on Participating SMEs that are also data controllers and who employ a data processor. As mentioned, they must choose a data processor providing sufficient guarantees in respect of the technical and organisational security measures governing the processing to be carried out, and take reasonable steps to ensure compliance with those measures. However, it is unlikely although not impossible for Participating SMEs to utilise a data processor as part of the Project.

– – –

7.2.6 Conclusion

Develop a compliance strategy

Despite the complexity of the multi-layered legal framework for data sharing and (re)usage, you must be able to ensure that your data sharing and (re)usage activities are lawful and ethical from the outset. Especially, since non-compliance can have severe consequences, including litigation and reputational damage. It is therefore important that as part of your compliance strategy you consider the following points:

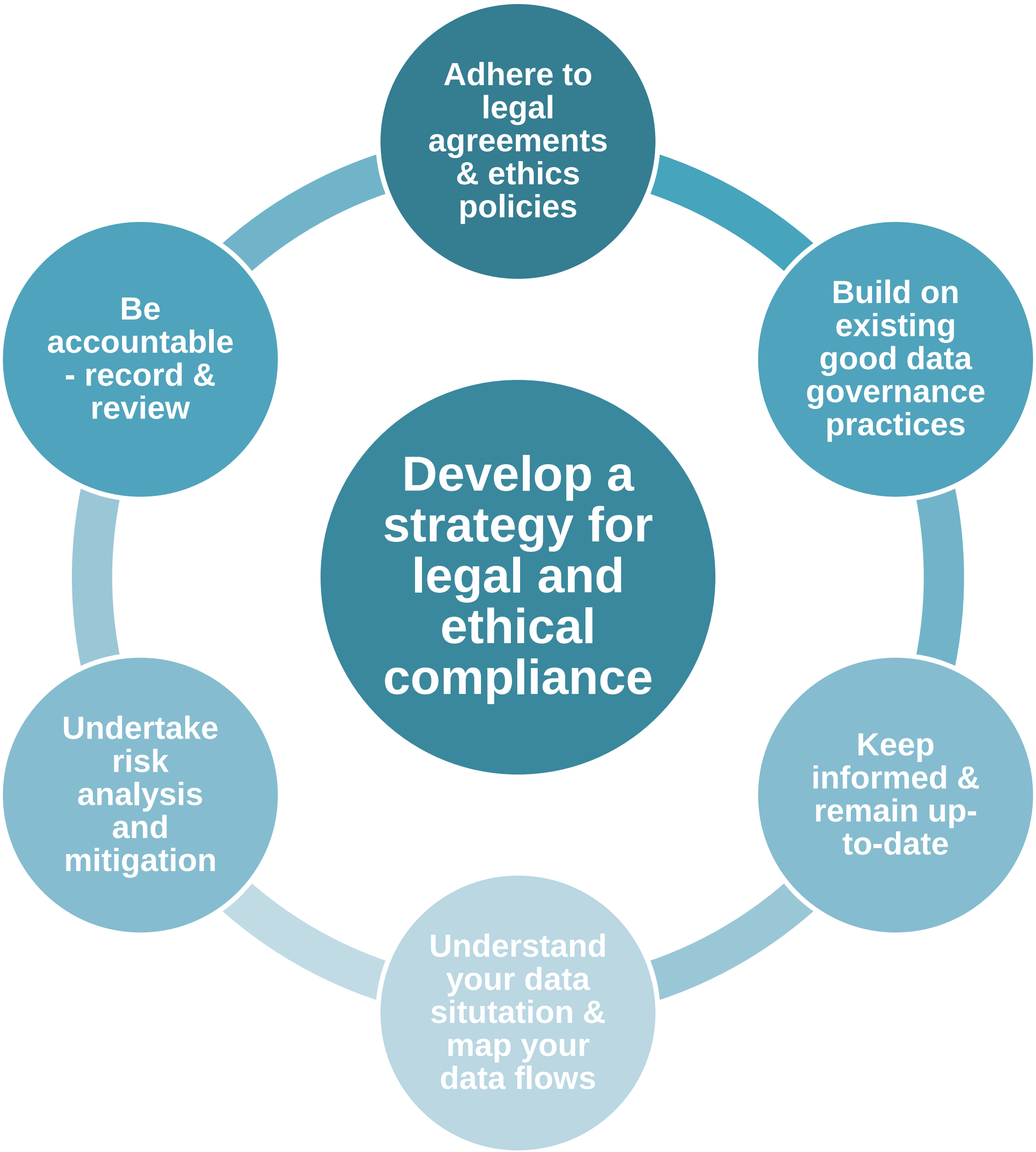

Adhere to legal agreements & ethics policies
Build on existing good data governance practices
Keep informed & remain up-to-date
Understand your data situtation & map your data flows
Undertake risk analysis and mitigation
Be accountable - record & review
Develop a strategy for legal and ethical compliance

Figure 15 Develop a strategy for legal and ethical compliance diagram

Adhere to legal agreements

Ensure that you adhere to any legal agreements you have signed and/or licensing arrangements that you enter into as part of a data sharing scheme.[148] In particular, you should check the nature of the limitations included which could reduce your planned re-usage of the data to be shared with you, such as the allocation of intellectual property rights between the data provider and the data recipient.[149]

Build on pre-existing good data governance practices

You should build on established risk assessment and management strategies for data-related activities within your organisation where such strategies are in existence. Such strategies may range from data protection and privacy governance frameworks to more detailed governance and management structures, e.g. which focus on information architecture, data accuracy, security, and

[148] For more information on data sharing agreements see: Data Legality Report v1 [60] and Data Legality Report v2 [59] that provide an overview of the creation of the Data Pitch contractual portfolio, including the three key contractual templates utilised by the programme.

[149] One key distinction to bear in mind in order to adequately allocate intellectual property rights is the distinction between the algorithm produced to process the shared data and the output of the processing activity through the means of the algorithm, which could be described as enriched or output data.

regulatory compliance.

Keep informed and up-to-date

It is crucial you possess an adequate level of understanding about the legal framework for data sharing and (re)usage in general – in particular, data protection laws, electronic privacy laws, intellectual property laws, competition laws, confidentiality laws and contract laws – plus any sector-specific legislation. Furthermore, it is important to remain cognisant of best practice standards that surpass minimum standards prescribed by the law, such as ISO standards.

Understand your data situation and map your data flows

Since the same dataset can attract different legal responsibilities, rights and liabilities under different sets of circumstances, it is vital that you understand the precise purpose(s) for and context of your planned data sharing and/or (re)usage activity. While data flow mapping is not a panacea for legal-compliance, it is a useful tool to employ so that: gaps between the regulatory framework and how data are processed in practice are revealed; (potential) high-risk data processing activities are identified and risks can be mitigated; individuals are well-positioned to make good decisions about data processing; and, the areas that require further legal training and guidance are exposed. After organisations have mapped their data flows, the next step will be to create (i) a risk register, (ii) a catalogue of risks; and (iii) associated control measures in order to demonstrate appropriate mitigation of risks. DPIAs must be carried out where required.

You should keep keep-up-to-date with relevant legal and ethical developments, e.g. by accessing training and guidance provided by those who oversee the data sharing scheme and other authoritative sources. For instance, see the table in the section below that provides a non-exhaustive list of organisations providing useful guidance and notable publications on legal compliance and ethical conduct.

Undertake risk analysis and mitigation

A key theme throughout the toolkit is the importance of data sharing and re-usage protocols together with appropriately documented risk analysis and management as part of a wider audit trail for responsible data sharing, management and re-usage. As illustrated by the anonymisation, pseudonymisation and re-identification risk strategy outlined by the first report, the basics of data mapping for anonymisation assessment raised by the second report, and the need for transnational, cross-sector evaluation highlighted by the third report.

The establishment of a robust risk-based data governance framework that fosters responsible data sharing, management and usage is vital for any open innovation programme. Data-related activities are responsible where they are lawful, ethical and ultimately accountable to members of the open innovation programme (e.g. the Consortium, Data Providers and other Participating SMEs), data subjects, supervisory authorities, external auditors and, where possible, the wider pubic. Compliance is important not only to avoid the often severe consequence of non-compliance with applicable laws – but to incentivise (more) data sharing between organisations and, ultimately, increase innovation by building trust and confidence in the open innovation model.

Some key areas of focus for risk analysis and mitigation are:

- The processing of personal data.
- Transnational, cross-sector data sharing.
- The use of automated decision-making and profiling.
- Intellectual rights management and clearance.
- The sharing and (re)usage of confidential information.

Be accountable – record and review

It is therefore essential that all organisations involved with a data sharing scheme retain an audit trail of steps taken to ensure legal and ethical compliance. This audit trail, for external and internal accountability, should demonstrate that:

a) The different areas of the law relevant to the planned data sharing and (re)usage activities, as well as data ethics, are considered and properly addressed across the lifetime of the programme;

(b) Such legal analyses are periodically reviewed to ensure they remain accurate, and are revised where there has been a material change in facts;

(c) A risk assessment has taken place; and

(d) Intellectual property rights clearance and management is appropriately recorded.

Key messages at a glance: a quick checklist on data sharing and (re)usage in open innovation from an EU perspective

This list of seventeen key legal messages for Participating SMEs to consider when sharing and re-using data builds on the initial list provided by the first toolkit report [3, p. 58]. This list has been expanded in order to take into account the further key messages raised by the two subsequent toolkit reports.

Some practical steps to consider when you are sharing and re-using data in the course of an open innovation programme (and elsewhere):

Figure 16 Table 3. Quick checklist on data sharing and (re)usage in open innovation from an EU perspective

Table 3. Quick checklist to help confirm whether data sharing, management and re-use is lawful	
1.	Comply with any legal agreements you have signed and/or licensing arrangements that you enter into as part of your participation in the open innovation programme. In particular, you should check the nature of the limitations included which could reduce your planned re-usage of the data being shared with you, such as: a. The allocation of intellectual property rights between the Data Provider and the Data Recipient.[150] b. Any constrains on transnational and/or cross-sector data sharing and usage.
2.	Conduct all your open innovation activities ethically, and adhere to the ethics statement and declaration of honour you have signed (or their equivalent).
3.	Identify and adhere to all applicable sector, process and/or service specific constraints and requirements, such as sector-specific codes of best practice, industry standards, and other legislative requirements.
4.	Respect third party intellectual property rights. Ensure you have the authority to share data and/or rights to use all data for your open innovation activities through an appropriately documented rights management and clearance process.
5.	Keep the Consortium informed about any SME (Self-Sourced) Data that you plan to use as part of your open innovation activities.
6.	Assess whether your open innovation activities would involve any data transfers to or from third countries or international organisations, and ensure these are lawful. Where personal data is involved, ensure the transfer to the third country or international organisation satisfies EU data processing requirements, and one of the three following routes in order to guarantee an adequate level of data protection: (i) adequacy decisions, (ii) appropriate safeguards, or (iii) derogations.
7.	Consider whether your open innovation activities would involve any data transfers between EEA member states, and ensure these are lawful. In particular, you should identify and comply with any applicable national legal variations.

[150] One key distinction to bear in mind in order to adequately allocate intellectual property rights is the distinction between the algorithm produced to process the shared data and the output of the processing activity through the means of the algorithm, which could be described as enriched or output data.

8.	Anonymised data must not be re-identified: a. You should make sure that the data does not allow individuals to be identified if combined with other information available to you (i.e. mosaic effects).[151] b. You should make an appropriate evaluation of the risk of re-identification where processing anonymised data.[152] c. You should periodically review the risk of re-identification and mosaic effects.[153]
9.	Implement data security measures that comply with the legal agreement you have signed, and appropriate to the specific context of the processing activity, including the level of sensitivity and risk.[154]
10.	Avoid the risk of harm to individuals who are the subjects of anonymised or pseudonymised data. a. You should consider the harm that might result from the accidental loss or unauthorised processing of data that are shared, managed and re-used as part of the open innovation programme. b. You should implement appropriate and effective safeguards that take into account the nature of the data being processed and any possible sensitivity.
11.	Ensure your open innovation activities would not involve any anti-competitive practices.
12.	Any breach of data protection law must be reported immediately to the Consortium and the relevant Data Provider must act accordingly as soon as possible to mitigate any risk of harm to data subjects.
13.	Conduct an appropriate and effective risk analysis and mitigation, including an anonymisation assessment. Data flow mapping is one method, which organisations sharing and/or re-using data can use, to support and demonstrate legal compliance with the GDPR and other applicable laws. It can further help to reveal further transnational and cross-sector considerations. After an organisation has mapped their data flows, the next step will be to create (i) a risk register, (ii) a catalogue of risks and (iii) associated control measures in order to demonstrate appropriate mitigation of risks.
14.	Open innovation activities that would involve solely automated decision making that produces legal or similarly significant effects are strictly prohibited, unless the decision *"(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; [/] (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or [/] (c) is based on the data subject's explicit consent."* – Article 22(2) of the GDPR.
15.	Open innovation activities that involve automated decision making and/or profiling should be explained to the data subjects involved.

[151] To note, it is not always necessary to have biographical details in data – such as a requirement for an individual to be named – for it to be deemed personal data. Identification (and re-identification) may be effected in other ways, notably through singling out individuals from others in a dataset via indirect identifiers.

[152] The risk of re-identification through data linkage is unpredictable because it can never be assessed with certainty what data are already available or what data may be released in the future. On the one hand, while it may not be possible to determine with absolute certainty that no individual will ever be identified as a result of the disclosure of anonymised data, a certain amount of pragmatism needs to be adopted. It involves more than making an educated guess that information is about someone.

[153] The likelihood and severity of re-identification risk occurring can also change over time (e.g. with the introduction of new technologies that can link data) and, therefore, re-assessments should be carried out periodically and any new risks managed. This should include trying to determine what additional information - personal data or not – could become available that could be linked to the data to result in re-identification.

[154] General and data-specific safeguards are set out in the legal agreements governing the reuse of data shared. Above and beyond fulfilling their contractual obligations, there is no 'one size fits all' solution to data security – each organisation should adopt a risk-based approach to its assessment and management in conjunction with implementing advice given by the Consortium. Different measures (and combinations of measures - legal, technological, and organisational) may be appropriate depending on the processing activity and other data environment contextual factors.

16.	Ensure all data-related activities are accountable, and retain an audit trail of steps taken to ensure legal and ethical compliance. This audit trail should demonstrate that: **a.** The different areas of the law relevant to the open innovation programme and the data sharing, management and re-usage activities and data ethics are considered and properly addressed across the lifetime of the programme. **b.** Such legal analyses are periodically reviewed to ensure they remain accurate, and are revised where there has been a material change in facts. **c.** A risk assessment has taken place. **d.** Intellectual property rights clearance and management is appropriately recorded.
17.	Keep up-to-date on relevant legal developments. Access training and guidance provided by the Consortium, and other authoritative sources (see following section), to reinforce existing knowledge or gain further knowledge on best practices for legal compliance and ethical conduct.

Further resources

Figure 17 Table 4. Organisations providing useful guidance and notable publications

Organisations providing useful guidance and notable publications	**Link to website**
Ada Lovelace Institute	https://www.adalovelaceinstitute.org/
Article 29 Data Protection Working Party archives (1997-November 2016)[155]	https://ec.europa.eu/justice/article-29/documentation/index_en.htm
Centre for Data Ethics and Innovation, UK	https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation
Commission Nationale de l'Informatique et des Libertés (CNIL), France	https://www.cnil.fr/
Competition and Markets Authority (CMA), UK	https://www.gov.uk/government/organisations/competition-and-markets-authority
EU Agency for Cybersecurity (ENISA)	https://www.enisa.europa.eu/
EU Agency for Fundamental Rights (FRA)	https://fra.europa.eu/en
EU Intellectual Property Office (EUIPO)	https://euipo.europa.eu/ohimportal/en
European Commission	http://ec.europa.eu/ (in particular see: https://ec.europa.eu/info/topics/justice-and-fundamental-rights_en ; https://ec.europa.eu/info/topics/competition_en ; https://ec.europa.eu/digital-single-market/en ; https://ec.europa.eu/programmes/horizon2020/).
European Data Protection Board (EDPB)	https://edpb.europa.eu/
European Data Protection Supervisor (EDPS)	https://edps.europa.eu/
Information Commissioner's Office (ICO)	https://ico.org.uk/
Intellectual Property Office (IPO), UK	https://www.gov.uk/government/organisations/intellectual-property-office

[155] Note: since 25 May 2018, the Article 29 Data Protection Working Party no longer exists. Refer to the European Data Protection Board for up-to-date guidance.

Interdisciplinary Centre for Law, Internet and Culture (iCLIC), UK	https://www.southampton.ac.uk/ilaws/index.page
National Data Protection Supervisory Authorities in the EEA (including CNIL and ICO).	For a list of current national data protection supervisory authorities in the EEA, including links to their websites – see: https://edpb.europa.eu/about-edpb/board/members_en
Organisation for Economic Co-Operation and Development (OECD)	http://www.oecd.org/
The Alan Turing Institute, UK	https://www.turing.ac.uk/
The Open Data Institute (ODI), UK	https://theodi.org/
The Royal Society	https://royalsociety.org/
UK Anonymisation Network (UKAN)	https://ukanon.net/
Web Science Institute, UK	https://www.southampton.ac.uk/wsi/index.page
World Intellectual Property Organization (WIPO)	https://www.wipo.int/portal/en/index.html

Summary

An effective strategy for handling data processing issues, together with useful guidance and training to promote awareness of key legal and ethical considerations – as provided by the toolkit – are fundamental components of any data sharing scheme. Moving forward, it is hoped that (more) organisations will be motivated to join and/or establish other data sharing schemes similar to Data Pitch – and as a result, share (more) closed data in trustworthy data environments that are legally and ethically compliant by default and design.

7.3 Appendix C. Supporting documents

7.3.1 Data Provider Questionnaire

Note: It was mandatory for potential Data Providers to the Data Pitch programme to complete the following Data Provider Questionnaire – the Data Pitch Consortium then reviewed given responses.

– – –

The datasets to be provided are likely to span many different types of data types, including proprietary data, associated with various levels of sensitivity.

Therefore, we would like some brief information about the type of data that your organisation would be happy to provide to the project, and possible degrees of data sensitivity in that overall data bundle. Using this information we can better assess what types of risks might arise and need to be managed.

ABOUT THE DATASETS YOU ARE CONSIDERING SHARING

1. What type and format of data would you be happy to share under the project?

2. How were these datasets created/collected?

3. What do they relate to (e.g. sector, product/service, territory, etc.)?

4. In relation to each dataset you would be happy to share under the project, would you currently classify them as (i) '**closed'** (internal access only); (ii) '**shared'** (third party access only); or (iii) '**open**' (where anyone can access).

5. Have you any constraints/strong will on data storage (e.g. in relation to where you would prefer data for sharing under the project to be stored, and around data access)?

To note, our data hosting and experimentation facilities are designed to manage datasets of all types, and will provide clear details about which organisation owns each dataset, usage restrictions, and the specific security level required for it.

In this context, we propose using pre-defined rule sets to specify the use, sharing, and anonymisation requirements (or pseudonymisation processes), of each dataset based on its sensitivity and level of openness as preferred by your organisation.

LEGAL OVERVIEW

Please note that any organisation that is considering entering into a data sharing arrangement with us must also do the following in respect of such data to be provided, and in respect of any national legal systems (including EU law) that apply:

- Highlight any regulatory requirements that might arise upon such data disclosure. In particular, this includes compliance with data protection law (for guidance, see e.g. the UK's Information Commissioner's Office Data Sharing Code of Practice).
- Highlight whether any intellectual property rights apply (for example, copyright restrictions, or data originally provided under a duty of confidence) that might arise upon such data disclosure.
- Highlight whether any other legal issues might arise upon such data disclosure (for example, because it is subject to contractual legal restrictions governing its usage).

Therefore:

- ❖ **If any of the data is likely to contain personal data, please answer the questions in the next section compliance with data protection law.**
- ❖ **If intellectual property (IP) rights are likely to apply to any of the data, please answer the questions in the following section regarding IP law.**
- ❖ **If you think that any other legal issues might become relevant if the data were shared as envisioned under this project, please answer the questions in the final section regarding contractual restrictions upon data usage, or any other possible legal issues that might be raised upon disclosure.**

DATA PROTECTION AND PRIVACY LAWS

Please provide details as follows:

1. Does any of the data contain personal data? In other words, according to EU law, does it contain *"any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"*?

Guidance: In relation to determining whether someone is identifiable from data (e.g. where it is not immediately obvious because the names and other identifiers have been removed), the key question for consideration is whether there are means that could likely reasonably be used to identify that person. In particular, under the new EU General Data protection Regulation (GDPR) coming into effect on 25 May 2018, the following guidance is provided (Recital 26, our emphasis regarding important, new changes from the existing EU Data Protection Directive):

> *"The principles of data protection should apply to any information concerning an identified or identifiable natural person.* ***Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person****.* ***['Pseudonymisation' is defined at Article 4(5) as, "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional***

> ***information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person"]. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.*** *The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes".*

If no to question 1, please answer question 10 below.

If yes, to question 1, please answer questions 2-9 below:

2. Please describe the type of personal data (e.g. names, addresses, telephone numbers, or financial information?) and how it was collected.

3. Where are the personal data currently stored and processed (including, if a third party is relied upon for these activities, where are they based)?

4. At a national level, what country's data protection laws applies to the processing of such personal data?

5. Under the Project, we want to mitigate the risks that personal data would be shared. Therefore, what specific measures will you put in place to mitigate the risk of identification from the data? Please give details. Would the data still retain individual-level elements – such that the data subject could still be singled out from the data via indirect identifiers?

6. Was data subject consent obtained when such personal data was gathered, or was another legal basis used to justify its collection,[156] and for what purpose?

7. If consent was obtained at the initial point of personal data collection, was consent also obtained to re-use that data for other purposes (e.g. for research purposes)?

8. If the personal data was collected from a third party, did that third party give a warranty or other appropriate form of assurance as to their compliance with data protection/privacy laws?

9. Does any of this personal data contain or imply sensitive personal data? In other words, according to EU law, does it contain information about the data subject's *"racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"*? If yes, please give details.

10. If the data you wish to share is not believed to be personal data (but it does relate to persons – that is, it refers to them, is used to determine or influence the way in which an individual is treated, or is likely to have an impact on their rights and interests), please specify why you think that those persons are not identifiable from the data. We are particularly interested in this answer if the data contains individual-level elements – such that individuals could still be singled out from the data by the SMEs analysing the data under the Project?

INTELLECTUAL PROPERTY (IP) LAWS

[156] The list of legal grounds that can justify the processing of personal data under the GDPR (Article 6) are: ***(a)*** *the data subject has given consent to the processing of his or her personal data for one or more specific purposes; (****b****) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;* ***(c)*** *processing is necessary for compliance with a legal obligation to which the controller is subject;* ***(d)*** *processing is necessary in order to protect the vital interests of the data subject or of another natural person; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (****f****) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.*

IP is an umbrella term which is used to describe a range of legal rights that attach to certain types of information and ideas and to their particular forms of expression. IP rights fall into two general categories:

- **Registered rights** that are granted on application to an official body, and include **patents, registered trade marks and registered designs**; and,
- **Unregistered rights** that arise automatically, give protection against copying or using the right, and include **copyright, database rights, unregistered design rights, rights in unregistered trade marks and confidential information**.

Please provide details of the following, including the relevant legal jurisdiction:

1. Any IP rights applying to the data you would be happy to share under the project. If so, please specify what type they are and what they apply to.

2. Ownership of such IP rights (including territory of right)

3. Any registrations of such IP rights (including the territory of registration and the relevant market of registration)

4. Any anticipated difficulty in granting licences (or sub-licences) to use such IP rights under the project (for example, on-going litigation)?

5. Whether any of the data you would be happy to share under the project was given in confidence. Under UK law, for example, a common law duty of confidence will arise where the information in question has a quality of confidence about it and was given in circumstances importing an obligation of confidence.

<u>CONTRACTUAL RESTRICTIONS AROUND DATA USAGE, OR ANY OTHER LEGAL ISSUES THAT MIGHT ARISE UPON DATA PROVISION AND SUBSEQUENT USAGE</u>

1. Please provide details of any contractual restrictions around the usage of the data that you would be happy to provide under the project, or any other possible legal issues related to such data usage, and specify the relevant territory. For example, if the data has been provided to your organisation by another party, please provide a brief overview of the contractual terms or service level agreements (SLAs) used for that purpose.

2. Are there any strong requirements you have regarding the further usage of any of your data sets dictated by law or for other reasons? What are these, in relation to what data, and what usage?

– – –

7.3.2 Record of Information about Self-Supplied Data

Note: During the Second Call of the Data Pitch programme, SMEs (re)using self-sourced data were asked to complete the following form.

– – –

Record of information about self-supplied data

[Company name]

Brief overview. As a participating SME in the Data Pitch programme, you are required to provide information to the Data Pitch consortium about the nature and source of each self-supplied dataset you intend - and have permission - to analyse as part of your challenge. After you have completed this form please send it via email to: [email]. If you would like further information or assistance with this form please do not hesitate to contact [email].

Contact details for further questions:

Date:

Number of self-sourced datasets I intend to (re)use:

	Brief description of dataset	Dataset attributes	Source of dataset	Does this dataset contain personal data?	Have any pseudonymisation or anonymisation measures been applied to these data? If so, please provide a brief description.
Dataset 1					
Dataset 2					
Dataset 3					
[Please delete/add rows to this table as you deem necessary]					

Please keep this record up-to-date. If you intend to (re)use further datasets, make sure: (1) these datasets are added to this record, and (2) you amend the version number in the footer.

– – –

7.3.3 Declaration of Honour for Participating SMEs to sign

Note: It was mandatory for potential Participating SMEs to sign the following Declaration of Honour.

– – –

1. As legal representative of [insert legal entity name], I declare that the entity is not:

a) bankrupt or being wound up, is having its affairs administered by the courts, has entered into an arrangement with creditors, has suspended business activities, is the subject of proceedings concerning those matters, or is in any analogous situation arising from a similar procedure provided for in national legislation or regulations;

b) having powers of representation, decision making or controlling personnel being convicted of, or having been convicted of an offence concerning their professional conduct by a judgment which has the force of res judicata;

c) having been guilty of grave professional misconduct proven by any means which the contracting authority can justify including by decisions of the European Investment Bank and international organisations

d) failing to be compliant with obligations relating to the payment of social security contributions or the payment of taxes in accordance with the legal provisions of the country in which it is established or with those of the country of the contracting authority or those of the country where the contract is to be performed;

e) having powers of representation, decision making or controlling personnel having been the subject of a judgment which has the force of res judicata for fraud, corruption, involvement in a criminal organisation or any other illegal activity, where such illegal activity is detrimental to the Union’s financial interests;

f) subject to an administrative penalty for being guilty of misrepresenting the information required by the contracting authority as a condition of participation in a grant award procedure or another procurement procedure or failing to supply this information, or having been declared to be in serious breach of its obligations under contracts or grants covered by the Union's budget.

2. I declare that the natural persons with power of representation, decision-making or control over the aforementioned legal entity are not in the situations referred to in b) and e) above.

3. I declare that I

a) am not subject to a conflict of interest and will take all reasonable measures to prevent any situation where the objectives of the Data Pitch project might be compromised due to undeclared shared interests;

b) have not made false declarations in supplying the required information to the project formally detailed as Data Pitch, and have not failed to supply the required information;

c) am not in one of the situations of exclusion, referred to in the abovementioned points a) to f).

4. I certify that I:

a) am committed to participate in the aforementioned project as part of the legal entity detailed above;

b) have stable and sufficient sources of funding to maintain its activity throughout its participation in the aforementioned project, and will provide any counterpart funding necessary;

c) have or will have the necessary resources as and when needed to carry out its involvement in the above mentioned project.

d) will comply with my responsibilities and obligations under the Data Pitch project, including those set out in the Data Sharing Agreement.

e) will respect any third party rights in relation to data provided for processing under the Data Pitch project.

f) will abide by international, EU and national laws and regulations that might apply to the substance, or outcome, of data sharing arrangements as relevant to activities that I/my entity will be involved in under the Data Pitch project.

g) will not share or disseminate data received through the Data Pitch project without the explicit prior consent of the data provider and any others with proprietary rights in relation to that data.

h) will take all reasonable measures to safeguard data provided to me/my entity for use in the Data Pitch project against possible misuse and unauthorised access.

i) will abide by international, EU and national laws imposing privacy and data protection requirements (including, in anticipation for its coming into effect, the General Data Protection Regulation **(**GDPR**)** (Regulation (EU) 2016/679)) as relevant. In particular, personal data shared under the Data Pitch project will not be re-used for purposes outside the project without the explicit prior consent of the data controller.

j) will act in good faith as far as reasonably possible under the Project and fully apply the principles of the Ethics Statement.

5. I declare that, to the best of my knowledge, I am eligible to apply for the Data Pitch call and all the information I have provided is true.

Name

Signature

Date

– – –

7.3.4 Ethics Statement for Participating SMEs to sign

Note: It was mandatory for potential Participating SMEs to sign the following Ethics Statement.

– – –

This Ethics Statement underpins the Data Pitch project in setting out specific rules and standards of conduct expected from recipients of Data Pitch funding. Ethical conduct means acting consistently in a way that is ethical and fair and encouraging others to do likewise.

The standard of behaviour expected is additional to compliance with relevant legal rights and obligations arising automatically by virtue of law applying to each participant. It is also not intended to exclude or replace responsibilities agreed under contract with the Data Pitch consortium (in case your application is successful), as well as the certifications/declarations set out in the Declaration of Honour.

As legal representative of [insert legal entity name], I certify that [insert legal entity name] will adhere to the following principles as far as reasonably possible under the Data Pitch project:

1. act in good faith;
2. respect human rights;
3. ensure research quality and integrity;
4. be able to show that our findings are independent and non-discriminatory to any groups of individuals;
5. not misrepresent credentials;
6. demonstrate authenticity and validity of authorship;
7. respect confidential information;
8. secure any confidential information provided to prevent its misuse or unauthorised access;
9. only share confidential information where necessary and only where the prior informed consent of anyone potentially affected by the disclosure of such information has been received;
10. respect the privacy of any people identified from the findings of the Data Pitch project as far as possible;
11. avoid any conduct that may cause anyone harm, and seek relevant individuals' informed consent for any activities that might affect them directly;
12. determine the applicable laws that may apply to our activities under the Data Pitch project and plan our activities in accordance with such laws as early as possible;
13. not collect or otherwise process any personal or sensitive data not essential for our Data Pitch activities;
14. be fully transparent to the Data Pitch consortium about the purpose, methods and intended possible uses of our Data Pitch activities, and what risks, if any, are involved.
15. seek advice promptly from the Data Pitch consortium where we believe ethical and/or legal risks may be raised by our activities.

Name

Signature

Date

– – –

7.3.5 Note on toolkit dissemination

Interactive forms of dissemination

Given the principal purpose of the toolkit is to offer guidance on the key legal and privacy aspects of data sharing, management and usage of (closed) data that can be understood by non-legal protection specialists, it was essential to engage with Participating SMEs to obtain feedback during the Data Pitch programme. Legal guidance workshops and webinar with Participating SMES therefore helped to enrich the development of the toolkit by revealing any practical gaps that require further attention as well as ensuring that the toolkit is user-friendly.

The provision other more interactive forms of dissemination is important for further engagement with toolkit users. For instance, we decided to create seven legal decision-trees in order to communicate some of the key aspects of the GDPR in a simple way to Participating SMEs. A decision was taken to make these decision-trees interactive through the development of a prototype e-learning tool on data protection and the basics of mapping data flows (accessible via http://pz-wptest.synote.io/ [last accessed 8 October 2019]) in order to enable tailored responses to a user's planned data processing activity.

Toolkit-related publications and conferences

The contribution to conference programmes and publications has been crucial for gaining external feedback through peer review as well as raising the profile of the toolkit and the wider Data Pitch programme. We therefore disseminated aspects of the toolkit, during the course of the Data Pitch programme, by contributing to the following conference programmes and publications:

Figure 18 Table 6. Dissemination of the toolkit during the Data Pitch programme (2017-2019)

Table. Dissemination of the toolkit during the Data Pitch programme (2017-2019)		
	Type	**Reference**
1.	Contribution to Book Section[*]	Hu, R., Stalla-Bourdillon, S.,* Yang, Mu., Schiavo, V. & Sassone, V. (September 1, 2017). Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR, in *Data Protection and Privacy: The Age of Intelligent Machines*, edited by R. Leenes, R. van Brakel, S. Gutwirth and P. De Hert, Hart Publishing, 2017. Available at SSRN: https://ssrn.com/abstract=3034261.
2.	Member of Conference Discussion Panel[*]	Panel Discussion on *The Law and Science of De-identification* organised by the Future of Privacy Forum (Sophie Stalla-Bourdillon* was a panel member), 10th International Conference on Computers, Privacy & Data Protection – The Age of Intelligent Machines, 25-27 January 2017, Brussels, Belgium (p. 29) http://www.cpdpconferences.org/assets/cpdp2017.pdf [last accessed 8 June 2018].
3.	Journal Article	Stalla-Bourdillon, S., & Knight, A. (2017). Anonymous data v. personal data—a false debate: an EU perspective on anonymization, pseudonymization and personal data. *Wisconsin International Law Journal*, *34*(2), 284-322.
4.	Conference Discussion Panel with Other Members of the Data Pitch Consortium	Panel Discussion on *Data Protection and Innovation Acceleration* organised by Data Pitch, 11th International Conference on Computers, Privacy & Data Protection - The Internet of Bodies, 24-26 January 2018, Brussels, Belgium, (p. 26) http://www.cpdpconferences.org/assets/CPDP2018_PROGRAM_FINAL.pdf [last accessed 1 June 2018]
5.	Member of Conference Panel Discussion [*]	Knight A.* & LaFever G. (April 18, 2018). Just Because You're GDPR Compliant Does Not Mean You Can Use Your Data, *IAPP European Data Protection Intensive Conference 2018*, Retrieved from: https://iapp.org/conference/iapp-europe-data-protection-intensive-2018/sessions-dpi18/?id=a191a000001JieyAAC [last accessed 1 June 2018].
6.	Conference Paper with Other Members of the Data Pitch Consortium	Stalla-Bourdillon S., Thuermer G., Walker, J. & Carmichael L. 2019. Data protection by design: Building the foundations of trustworthy data sharing. In Proceedings of Data for Policy (DfP). Note: a revised edition is forthcoming for publication in an academic journal.

8. Glossary

For the purposes of the toolkit, the following terms are defined as follows:

(Data Provider) Data: These data are: (a) shared by a Data Provider under a signed Data Provider (Data Sharing) Agreement; and subsequently (b) re-used for innovation purposes by successful SMEs - as part of the Data Provider Challenges track* - pursuant to a signed SME Contract.*

Anonymisation Practice: A technical and/or organisational method employed to de-identify a dataset (i.e. remove personally identifiable information from a dataset) [63, p. 11]. There are two main approaches to anonymisation [63, pp. 11-19]: (i) *"randomization techniques"* – e.g. *"noise addition"*, *"permutation"* and *"differential privacy"* and (ii) "*generalization techniques*" – e.g. *"aggregation and K-anonymity"* and *"L-diversity and T-Closeness"*. *Note: this term is not to be confused with the higher benchmark set by the legal standard for anonymisation (see definition below).****

Assessment of Anonymisation Practices: A process of evaluation undertaken by data owners to ensure their past, current and planned data practices and activities are legally and ethically compliant.***

Big Data: Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.**

Big Data Analytical Techniques: Computational methods applied to big data (such as artificial intelligence systems, including machine learning) to reveal patterns, trends, and associations.

Controller: An organisation which alone or jointly with others determines the purposes and means of the processing of personal data.**

Cross-Sector Data Sharing: One or more parties communicate, disclose or otherwise make particular data available to one or more other parties in distinct areas (e.g. another industry).[157] ****

Data Ethics: *"a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)."* Source: Floridi and Taddeo [100].

Data Flow Mapping: The creation of a graphical representation that charts the actual and potential movement of a (particular version of) a dataset as it is collected, managed, shared and (re)used across various data environments. An approach employed for the creation of data situation models.***

Data Owner: An individual and/or organisation who is involved with, and therefore able to make decisions about: the collection, management, release and/or (re)usage of a dataset. E.g. a data provider who releases data within an open innovation programme or an SME who re-uses these data. *Note: this term is widely used in ICT and business sectors – it is not a legal term.****

Data Provider Challenges: The Data Provider determines the challenge and provides certain data (i.e. Data Provider Data) for the successful applicant to reuse as part of their solution to the specific problem raised. Furthermore, the successful applicant is able to include other datasets (i.e. SME Data) if required.*

Data Providers: Those organisations that agree to supply their data under the open innovation

[157] For instance, "cross-sectoral" is defined by Lexico [149] as: *"[r]elating to or affecting more than one group, area, or section"*. E.g. see [157] for further background information on cross-sector data sharing – in particular data collaboratives.

programme.**

Data Sharing: One or more parties *"communicate, disclose or otherwise make particular data available"* [1, p. 7] to one or more other parties. Data sharing can (i) occur within a specific organisation or between third party organisations; and (ii) be "systematic" or "exceptional" [2, p. 9].[158] ****

Data Situation Model: A representation of a (particular version of a) dataset and the relationships with its various environments – past, current and planned – that can be used as a basis for anonymisation assessment. *Note: "data situation" is a term utilised by the UK Anonymisation Network (UKAN) see:* [62, p. 130].***

Data Subject: A living individual who is the subject of personal data.**

Direct Identifier: Data that directly identifies a single individual.**

Indirect Identifier: Data that indirectly identifies a single individual.**

Legal Standard for Anonymisation: *"[…] personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable […]"* Source: Recital 26 of the GDPR [12].***

Mosaic Effect: When it is possible to determine an individual data subject's identity without having access to obvious (direct) identifiers by correlating data pertaining to the individual across numerous datasets or intra-dataset; whereas individual identifiers in these datasets would not otherwise allow a data subject to be re-identified.**

Open Innovation Challenge: Applicants to the programme are able to propose an innovative solution that (re)uses data that are self-sourced (i.e. SME Self-Sourced Data).*

Open Innovation: Organisations work with external partners (e.g. data providers and SMEs) and/or obtain insights from external sources in order to develop high impact, cutting-edge ideas, methods, products and/or services (e.g. in the course of an accelerator).[159] ****

Participating SMEs: Those organisations that are selected to process data under the acceleration stage of the open innovation programme.**

Processor: An organisation which processes personal data on behalf of the controller.**

Profiling: *"any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".* Source: Article 4(4) of the GDPR [12].****

Re-identification Risk: The probability in which an individual/individuals could be identified from specific data [74, p. 26], [75, p. 24], and the likely resultant harm or impact if re-identification were to occur [75, p. 24].[160] The Article 29 Working Party [63, pp. 11-12] identifies three principal ways in which individuals may be re-identified, by: (1) singling-out an individual; (2) linking records relating to an individual; and (3) inferring information concerning an individual.[161] ***

[158] Note that this definition is based on: (i) ICO Data Sharing Code of Practice [2, p. 9]: *"the disclosure of data from one or more organisations to a third party organisation or organisations, or the sharing of data between different parts of an organisation […] two main types of data sharing: [/] systematic […] and, [/] exceptional […]"*; and, [1, p. 7]. For more information on the key types of business-to-business (B2B) data sharing models see: [150, p. 5] and [151, pp. 60-65].

[159] For instance, "innovate" is defined by Lexico [153] as: *"[m]ake changes in something established, especially by introducing new methods, ideas, or products"*. Furthermore, Henry Chesbrough [154, p. 1] defines open innovation as follows: *"the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively"*. Note that "open" innovation does not necessarily equate to "free" innovation – as it is common for some open innovation programmes to involve licensing fees and other financial agreements [152, p. 9]. For further information on open innovation see e.g.: [113] for a brief history of open innovation; [110] on innovation accelerators; [118] for an European Commission (EC) list of online resources on open innovation; [115] for an interview with Henry Chesbrough on open innovation; and, [155] & [156] on open innovation in practice.

[160] In the words of Marion Oswald [75, p. 24], it is an assessment about *"the possibility of something bad happening"*.

[161] Moreover, Boris Lubarsky [178] outlines the three principal methods of re-identification: (a) *"insufficient de-identification"*; (b)

Secondary Processing: Any processing of data after its initial collection.**

Sectoral Challenges: The Data Pitch consortium sets the challenge that requires the successful applicant to provide their own data (i.e. SME Self-Sourced Data) in order to solve the problem raised.*

SME Data: These data are: (a) gathered and/or collected by a successful SME; and/or (b) obtained from a third party by an SME. These data can be (re-)used for innovation purposes alongside (Data Provider) Data - as part of the data provider challenges track - by successful SMEs (subject to conditions) under a signed SME Contract.*

SME Self-Sourced Data: These data are: (a) gathered and/or collected by a successful SME; and/or (b) obtained from a third party by an SME. These data are (re-)used for innovation purposes - as part of the sectoral or open innovation challenges tracks by successful SMEs (subject to conditions) under a signed SME Self-Sourced Data Contract.*

Transnational Data Sharing: One or more parties communicate, disclose or otherwise make particular data available to one or more other parties in third countries or international organisations.[162] ****

Source of definitions:
* Definition from Data Legality Report v2 [59, p. 7].
** Definition from D3.1 Legal and Privacy Toolkit v1 [3, p. 10]
*** Definition from D3.5 Legal and Privacy Toolkit v2 [59, p. 7]
**** Definition from D3.9 Legal and Privacy Aspects of Transnational, Cross-Sector Data Sharing in Open Innovation

"pseudonym reversal"; and, (c) *"combining datasets"*.

[162] For instance, "transnational" is defined by Lexico [148] as: *"[e]xtending or operating across national boundaries"*.