Deliverable

D1.2 Cities Guide to Legal Compliance for Data-Driven Decision Making It. 1

Project Acronym:	DUET	
Project title:	Digital Urban European Twins	
Grant Agreement No.	870697	
Website:	www.digitalurbantwins.eu	
Version:	1.0	
Date:	2 October 2020	
Responsible Partner:	GSL	
Contributing Partners:	AIV, PLZ, DAEM	
Reviewers:	Lieven Raes (AIV) Nils Walravens (IMEC) Andrew Stott	
Dissemination Level:	Public	x
	Confidential – only consortium members and European Commission	

This project has received financial support from the European Union's Horizon 2020 Programme under grant agreement no. 870697

Revision History

Revision	Date	Author	Organization	Description
0.1	5.8.2020	Kletia Noti, Tomas Pavelka	GSL	Initial structure
0.2	14.9.2020	Kletia Noti, Tomas Pavelka Lieven Raes Zdenek Malik Dimitra Tsakanika	GSL AIV PLZ DAEM	First draft pilot questionnaires input
0.3	21/9/2020	Lieven Raes	AIV	Review
0.4	28/9/2020	Kletia Noti, Tomas Pavelka	GSL	Second draft
0.5	2/10/2020	Geert Mareels	AIV	Review
1.0	2/10/2020	Kletia Noti, Tomas Pavelka	GSL	Final version

Table of Contents

Executive Summary

Smart cities must ensure that data-driven decisions (1) conform to the applicable law (are regulatory compliant) and (2) manage legal liability risks related to defective decisions or defective data. Breach in either area may result in public enforcement (penalties), and/or private enforcement (damages claims, restrictive orders), and reputational damage.

This deliverable provides a first version of the emerging "Cities Guide to Legal Compliance for Data-Driven Decision Making", an easy to understand guide for cities on the legal necessities for data-driven decision or policy making.

A note is made that this deliverable should not be regarded as legal advice strictly speaking; organisations' legal departments or external attorneys qualified in the concerned jurisdictions should be consulted with respect to any particular legal matter.

The document comprises a theoretical part:

- **Summary of EU-level legislative framework** for data-driven decision making, including the GDPR, the ePrivacy legislation, a free flow of non-personal data framework, and intellectual property and trade secrets legislation, related to data ownership.
- **Set of definitions and concepts** used by the guide; this has an ambition to contribute to the emerging glossary of terms horizontally used by the DUET consortium. These definitions comprise GDPR terminology, which is necessary precisely to delineate personal data and mixed datasets that have the status of personal data, in order to ensure GDPR compliance at each step of decision making where personal data is involved. Other definitions are related to terms used by DUET (original data, existing data, context data, historical data), which cut into legal liability risk assessment in one way or the other. Several data or processing categories may trigger additional legal necessities and should therefore be carefully defined, e.g., special categories of personal data (biometric data, health data, etc.), trade secrets, location data, profiling methods, etc.
- **Overview of core overarching principles**: adherence to these principles helps ensure regulatory compliance and also a better management of liability risks. As a starting point, the use and consultation of data is, in GDPR terms, considered as a kind of data processing. This has a range of implications. First, a decision making process using any kind of personal data will be subject to the GDPR. Second, if there is a string of processes leading to the decision, these may be considered separate kinds of processing, and be subject to standalone GDPR requirements (e.g., a sufficient legal basis and specified purpose may be required for a GDPR compliant processing at each such stage). Finally, the principles on which the GDPR is based help manage legal liability risks even where no personal data is involved. E.g., observing the data minimisation principle means that only strictly necessary amount or detail of data is used to drive a process, and thus limits the exposure to risk in cases of flawed or inadequate data.

  Similarly, the privacy by design and by default (which includes the data minimisation imperative) and the precautionary principles (if in doubt, apply the stricter legal standard) help achieve full legal compliance and avoid any residual risks, even though this deliverable also warns against application of excessive standards not strictly required by the law, so that any choking effects that might hinder achievement of data driven processes' full potential are avoided.

There are also certain principles chosen by DUET as a consortium of partner organisations to steer its handling of data: first, that the ownership of data goes hand in hand with the responsibility for data management, and second, that personal profiles of human beings should not be commercialised.

which is followed by a practical part:

- **Summary of pilot cities organisations responses to data-related questionnaires with a commentary**. An initial set of questions responded by pilot organisations in Athens, Region Flanders, and Pilsen helped to map the inner data processes and intentions as regards DUET purposes. We have found the following:

  Each partner organisation appears to have sufficient responsibilities in data acquisition and management so that it can (co-)determine the datasets usage for DUET purposes. On the other hand, no pre-established procedures for handling data in decision making processes seem to have been established by the pilot organisations, which underscores the impact potential of this deliverable.

  Each partner is familiar with handling personal data and has access to a Data Protection Officer and a legal department, which is an indication that handling of personal data for DUET purposes (even incidental) should not raise manifest project risks. In any event, responses show that there is no intention – at the moment – to make an extensive use of personal datasets for DUET purposes. Each pilot organisation has some experience with data sanitization techniques including anonymization and pseudonymization.

  The pilot organisations identified several data sources and activities and provided initial information, on the basis of which we can take a preliminary view as regards their regulatory risk potential. In respect of count data generated by sensors for measuring air quality or noise levels, or magnetic loop sensors for counting traffic, we consider that these sensors would not typically generate personal data. Accordingly, use of such data for further decision making processes should not raise clear GDPR-related risks. On the other hand, data generated by ANPR (automatic plate number recognition cameras) and floating car data (FCD) may involve personal data and their further processing could be subject to GDPR requirements (or trigger ePrivacy law requirements). However, it appears - at this stage - that any ANPR data used by DUET will be anonymised, and FCD data will be provided to DUET by a third party and also fully anonymised, which reduces risk of privacy non-compliance.

  We found also that DUET will predominantly rely on existing third-party data, which means that some (careful) assumption may be made about the data regulatory compliance in cases where the data is shared by reputable third-party providers or public authorities.

  We identified two other specific data areas that warrant further interest to confirm the scope of applicable legal necessities: data related to terminal equipment of end-users (e.g., mobile phones or connected vehicles), and data in electronic communication services. Processing of WiFi and other unique identifiers of these devices may trigger ePrivacy law requirements. For the upcoming deliverables in this stream, we will investigate the degree of partner involvement in these transmissions or the original data collection, because mere purchasing of such data already pre-processed by third parties should be prima facie risk free. Similarly, if DUET is not directly involved in processing of data in cities' machine-to-machine dataflows conducted via public electronic communication networks, the ePrivacy law requirements on these services should not add further regulatory burden on DUET.

The emerging "Cities Guide to Legal Compliance for Data-Driven Decision Making". The first version, which is to be gradually improved and targeted by forthcoming deliverables (July 2021, May 2022), provides a logically organised set of areas to check for legal necessities of a smart city data-driven decision making process. Following an introduction (including a necessary legal disclaimer), the guide elaborates on the following interest areas:

What decision and who makes it: decision type, hierarchy and reach of the decision matter; GDPR compliance at each step if personal data involved; higher privacy impact risk with new technology and large scale data operations; legally binding measures/policies may require further impact on fundamental freedoms assessment.

Type of decision-making process: processing is typically automated and triggers GDPR if personal data is involved; emphasis must be put on accountability and security of automated systems; risk of data re-identification; algorithms should be open and fair.

Ownership issues (decision, data, IP): distribution of liability; legal gap for AI systems; legal basis for use of data (ownership, user license, open access) data ownership and responsibility principle; IP ownership; Article 26.1 of the Grant Agreement (IP ownership of DUET activities results).

Data factual quality (properties): factual data properties impact on legal necessities; GDPR imposes certain data properties quality standards; ANPR data example.

Data legal quality: typical legal defects: GDPR breach; ePrivacy rules breach (state-level differences); license infringement; IP rights infringement; original data (primary responsibility of your organisation) vs. 3rd party data (primary responsibility of the 3rd party organisation); full data audit; limited data audit; location data (preference for anonymous data principle); web/social media data; ANPR data, smart data.

Risks in further data processing: regulatory non-compliance, liability for damages, flaws in processing data within the decision-making process; personal data (legal basis, country-specific derogations); location data; IoT data; liability for defective data; liability risk limitation (contractual, non-contractual liability).

Purpose limitation: specifying the data processing purpose; original purpose, re-use purposes, compatibility test; compatibility presumptions; EU Member State purpose exemptions; terminal equipment data (ePrivacy rules: consent based re-use); non-personal data purpose setting.

Data minimisation, adequacy of data use: GDPR principle of data minimisation; privacy by default; data extent and granularity; large-scale data processing additional requirements; pseudonymization and encryption; excessive IoT data; inside organisation application (Article 39.2 of the Grant Agreement).

Accountability, fairness, transparency: organisation must be able to demonstrate regulatory compliance (GDPR, ePrivacy, other applicable laws); information to data subjects; IoT transparency; measures to ensure accountability (data protection policies, processing agreements, documentation of data processing activities, record and report data breaches); DPIAs AI systems transparency and fairness; security by design.

Collaboration and trust: standardisation; standard licenses; contextual controls; contracts for personal data processing; use-restriction marking; data retention periods.

1. Introduction

DUET is creating a “Cities Guide to Legal Compliance for Data-Driven Decision Making”, an easy to understand guide for cities on the legal necessities for data-driven decision or policy making.

Any decision or policy making based on data, data models or data analytics by a smart city should represent legally compliant solutions that reflect the smart city’s objectives as well as public values, and serve the public good and interests. In particular, a smart city’s handling of data for decision-making must (1) conform to the applicable legislation in the area (regulatory compliance, “safe harbour”) and (2) take into account potential risks of legal liability for any harm done by wrong decisions based on data or data analytics or caused by the defect in the data itself (legal liability prevention and management). Consequences of a breach in either area may expose any organisations concerned to legal enforcement by public authorities as well as private actors, and potentially lead to administrative or criminal penalties, restrictive court orders, damages claims, and – last but not least – reputational damage.

The aim of the deliverable and the emerging guide is to help smart cities (and at this earlier stage, concerned DUET partners) ensure, by means of a logically structured guidance, to pay attention to adequate regulatory compliance and liability risk assessments at various stages of a data driven decision-making process. This document is the first in a series of three planned deliverables concerning the emerging “Cities Guide”- the main reason for planning future versions of this guide is the gradual development in understanding DUET piloting activities as well as the DUET architectural design, which are in their early stages. Our learning curve about smart cities’ practical needs in the decision-making arena is currently based on a dialogue with the DUET pilot cities teams via targeted questionnaires. First iterations of these questionnaires have already informed this deliverable. Accordingly, we expect that the emerging guide may at first be helpful internally to inform DUET’s activities, but towards the end of the project, we may consider recasting the final document in a more abstract fashion to create a set of best practices shareable with other interested smart cities, stakeholders, or the general public[1].

This document is divided into a *theoretical part* providing a summary of the applicable legislative framework (with references to the more detailed overview given by deliverable D1.1) (Section 2), a set of working definitions of terms and concepts used by the emerging guide, and an overview of core overarching principles for data-driven decision making and data handling (Section 3). This is followed by a *practical part*, which, first, provides a commented summary of information gathered to this date from pilot cities about their data handling approaches within DUET activities (Section 4), and second, a step-by-step guide that can be followed to inform risk assessment in a data-driven decision making process (Section 5). Cross-references seek to connect the theoretical and practical parts of this document, as well as tie the pilot activities examples to the guidance, which is intended to be more widely applicable, and thus more high level. Section 6 concludes and outlines future work in this stream of deliverables.

This document and the emerging guide is complementary to, and does not replace, existing or future legislation (in more detail described in D1.1), and is intended to complement the Data Management Plan developed as a part of the WP8 deliverables. Ethically responsible decision making with regard to data and cloud infrastructure will be addressed by forthcoming deliverables D1.5 and D1.6 (Ethical Principles for using

[1] We take note of Article 29.1 of the Grant Agreement.

Data-Driven Decision in the Cloud). That said, the emerging guide does at times go beyond the basic level of "legal necessities" and seeks also to provide directional guidance on certain ethical aspects of data-driven decision making. This is done mainly in the areas where DUET wishes to adhere to higher than legal standards, but also where the currently applicable legislation leaves certain gaps (such as in the area of AI, robots and machine learning governance), and where we would advise against exploiting these gaps. This is because doing so might result in exposing the stakeholders to some kind of legal liability, or at least reputational risk, at the end of the day.

Legal issues around cybersecurity, sketched in Chapter D1.1, will be dealt with in more detail in the forthcoming deliverables on ethics (D1.5 and D1.6 (Ethical Principles for using Data-Driven Decision in the Cloud)). This document touches on security only insofar far there are legal necessities imposed by the GDPR and ePrivacy legislation.

2. Legal framework

2.1 Legal notice

This document and the emerging guide is complementary to, and does not replace, existing or future legislation in further detail described in deliverable D1.1 (Legal Landscape and Requirements Plan). Readers should make use of references back to deliverable D1.1 in order to get a fuller picture of the applicable law in the areas covered in the emerging guide.

It is important to note that this document or deliverable D1.1 do not, and are not intended to, constitute legal advice; instead, all information, content, and materials in these documents are for informational purposes only within the scope and objectives defined for the respective DUET project deliverables. Given that these documents got finalized at a certain point in time, information in these documents may not constitute the most up-to-date legal or other information at the cut off date. Readers of these documents and their organisations should contact their in-house team members or an attorney qualified in the concerned jurisdictions to obtain advice with respect to any particular legal matter.

In addition, we understand that, as regards data collection (and presumably also other data management activities), "*every partner is responsible for the behaviour of all team members, which may also include subcontracted organisations.*" (D8.3, p. 18). At the time of the first version of the Data Management Plan (D8.3), it was not yet clear, which DUET partners are responsible for the various steps of data management; this potential responsibility gap is indicated in the "future work" section (p. 27). D8.3 mentions a possibility of creating a working group of Data Protection Officers from each participant organisation (*ibid*), which, from the legal point of view, is recommended as well.

2.2 Applicable legislation

Smart cities may have many constituent actors: municipalities and other public bodies and authorities, public interest groups, private interests groups, commercial companies (suppliers or collaborators), and possibly also concerned citizens and the general public.

In the eyes of the law, each such actor may bear a share of responsibility for the decisions it makes or partakes in, and may be subject to a wide range of applicable laws, by-laws, soft law, court or administrative decisions or self-regulation at a local, national, EU-wide or international level.

It is in principle a duty of each actor to determine what legal obligations it is subject to at the relevant time and in the relevant circumstances. The following is an overview of EU-wide regulation and legislation that will typically be applicable to most data-driven activities within DUET's pilots.

2.2.1 GDPR

Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (the **General Data Protection Regulation**, or the **GDPR**) lays down the principles, rules and procedures to be followed by any actors involved in processing of personal data related to individual natural persons residing in the EU, or any personal data processing by actors established in one or more EU Member States.

According to the GDPR, the use or consultation of personal data is considered "data processing". This means that data use for decision making will fall within the GDPR's scope of application, to the extent they are personal data. (see further **Section 3.3.1**).

Even though the GDPR governs only processing of personal data (or mixed datasets where personal and non-personal data are inextricably linked), the basic principles for legal, fair and transparent data processing may be recommended for management of any data, including non-personal data, because they may directly impact on the overall legal risk exposure of the organisations handling data. These principles are further set out in **Section 3.3.2.**

For additional information on the GDPR, readers are guided to **Section 2 of deliverable D1.1.**

2.2.2 ePrivacy legislation

Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (**ePrivacy Directive**) complements the EU privacy framework by providing a set of special rules protecting electronic communication data.

Insofar as relevant for DUET's data management purposes, the Directive's provisions apply to a) processing of data in connection with the provision of publicly available electronic communication services[2], including location data of such services' users or subscribers[3], and to b) storing of information (data), or gaining of access to information (data) already stored, in the terminal equipment of end-users[4]. By way of example, activities involving collection of location data by telecom operators or access to data stored in terminal equipment by help of an app installed on users' equipment (such as mobile phone apps), or by tracking

[2] Article 3 ePrivacy Directive.

[3] Article 9(1) ePrivacy Directive.

[4] Article 5(3) ePrivacy Directive.

technologies processing data from terminal equipment, will typically fall within the scope of application of these rules.

It is important to note that the ePrivacy rules apply not only to personal data, but to any data in electronic communications (even though typically they will include data belonging to, or about, an individual user), and not only to data of natural persons, but to data of (or on) legal persons as well[5].

In terms of interplay with the GDPR, the Directive is in a relationship of speciality: provisions of the ePrivacy Directive serve to "particularise and complement" the GDPR. The GDPR, in turn, provides with respect to the ePrivacy Directive that it (GDPR) does not impose additional obligations in relation to processing of personal data in connection with the provision of publicly available electronic communication services in relation to matters for which they are subject to specific obligations with the same objective set out in the ePrivacy Directive.

The ePrivacy rules may pose some difficult questions about their applicability to smart city activities, particularly in the context of IoT systems, which could be considered as "electronic communication services" when they channel data over public networks such as the Internet. We cannot exclude at this stage that activities will be identified, in respect of which DUET would be advised to initiate consultations with national Data Protection Authorities.

In addition, the ePrivacy Directive is considered outdated and a proposal for new ePrivacy Regulation is pending in the legislative process. This may bring about important changes regarding the scope and substance of these rules. For more details on these developments and on selected aspects of the ePrivacy Directive as such, see **Section 2 of deliverable D1.1**.

Finally, the applicable ePrivacy rules are still only in the form of a directive, which means that the directly applicable national legislation transposing these rules must be consulted in any particular matter. There may be country-specific differences also in the way in which Member States have opted to derogate from these rules for various public policy reasons.

2.2.3 Regulation on free flow of non-personal data

DUET systems will make use of large quantities of non-personal data. **Regulation (EU) 2018/1807 on a framework for the free flow of non-personal data in the European Union** provides a legal framework for the free flow of non-personal data in the EU. Examples are machine-generated data or commercial data, which are either non-personal in nature or refer to personal data that has been made anonymous.

The regulation aims at removing data localisation requirements by individual EU Member States, make data available to competent (public) authorities for performance of their duties in accordance with the law, and facilitate data porting and promote open standards, and help cooperation between authorities of different EU Member States.

[5] Judgment of the Court of Justice of the EU in Case C-673/17 Planet 49.

This regulation reflects in the emerging guide to the extent it governs the relationship with the GDPR in the area of mixed datasets. For additional information on the Regulation, readers are guided to **Section 2.2.1 of deliverable D1.1.**

2.2.2 Open data/Public Sector Information legislation

The EU has adopted rules to encourage Member States to facilitate the re-use of data from the public sector with minimal or no legal, technical or financial constraints. **Directive (EU) 2019/1024 on open data and the re-use of public sector information** will replace rules introduced by earlier Directive 2003/98/EC on the re-use of public sector information (PSI Directive, with its major amendment in 2013).

The Directive defines certain groups of data and documents that should be "open by design and by default", i.e., data in open format, including dynamic data and APIs, that can be freely used, re-used and shared by anyone for any purpose, for private or commercial purposes.

The open data legislation reflects in the emerging guide mainly in the context of standard licenses and license limitations. For additional information on the directives, readers are guided to **Section 2.2.1 of deliverable D1.1.**

2.2.2 Intellectual Property Rights / Databases / trade secrets / national regimes

The IP-related legislation reflects in the emerging guide mainly in the areas of decision and data ownership, and risks assessment of a decision making potentially infringing on third party IP rights. For additional information, readers are guided to **Section 4.2 of deliverable D1.1.**

3. Issues, concepts and definitions, overarching principles

3.1 Issues

Datasets and models may suffer from deficiencies in various properties. Any decisions made on the basis of inadequate, low quality or even defective data may translate, if not amplify these deficiencies in the resulting decisions and lead to regulatory non-compliance and legal liability.

Smart cities are likely to face issues in the following data properties:[6]

- *definition-related properties*
  - *relevance: the usefulness of the data in the context of your business.*
  - *clarity: the availability of a clear and shared definition for the data.*
  - *consistency: the compatibility of the same type of data from different sources.*

[6] Developing High Quality Data Models, EPISTLE, 2003.

  - *timeliness: the availability of data at the time required and how up to date that data is.*
  - *accuracy: how close to the truth the data is.*

- *properties related to both definition and content*
  - *completeness: how much of the required data is available.*
  - *accessibility: where, how, and to whom the data is available or not available (e.g. security).*
  - *cost: the cost incurred in obtaining the data, and making it available for use.*

The GDPR, for instance, recognises that issues with data accuracy or completeness may lead to unfair and non-transparent handling of personal data, and impact on data subjects' right to privacy. We consider the same to be true also with regard to handling of non-personal data, use of which for decision making may result in a range of sub-optimal outcomes, ranging from low-quality, uniformed, decisions through to wrongful acts causing direct or indirect harm and resulting in an organisation-wide or individual-level legal liability. Similar repercussions can be caused by making decisions based on out-of-date, inconsistent, unclear, or irrelevant data.

In addition, data-driven decision making is not risk-free even when good quality data is available. This is because, e.g., the decision-making process itself may fail to meet legal necessities, such as GDPR requirements for further personal data processing, or may simply be badly designed or executed. Residual risks may be present in the simple fact that a decision or policy always aims at achieving a certain impact, and where there is impact on others, liability may follow.

We trust that these risks can be mitigated if the organisations and individuals concerned opt to adhere to the overarching principles when handling data (**Section 3.3**), and follow the emerging guide on legal necessities for data-driven decision making (**Section 5**) below.

Data skills and data literacy may be a separate issue. The data literacy issue may have direct implications for the quality of any decision made on the basis of data and data analytics. As the European Commission observed*,* big data and analytics are top of the list of critical skills shortages. In 2017, there were approximately 496 000 unfilled positions in the area of big data and analytics in the EU27[7]. We consider that analysis of this issue goes beyond the ambitions of this deliverable, and will need to be tackled by smart cities by means of proactive and responsible human resources policies.

3.2 Definitions

This section is intended to be used in conjunction with the emerging guidance and provide those who consider themselves insufficiently acquainted with basic issues and definitions in the legal aspects of data handling. The main reason why we suggest a separate section for definitions is that many smart city professionals and DUET colleagues will be fairly familiar with these terms, and so the guide can be more succinct without these definitions provided in-line or by help of footnotes.

This section can also contribute to an emerging glossary of terms related to DUET data management. We have sought to ensure consistency with terms and definitions used in related working streams (WP8). We anticipate a further discussion in this direction with the concerned partners.

[7] A European strategy for data, COM(2020) 66 final, p. 10.

3.2.1 Data categories according to their protection

Table 1: Data according to protection

Category	Definition	Comment/example
Personal data	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[8].	**Examples**: name and surname; a home address; an email address such as name.surname@company.com; an identification card number; location data (for example the location data function on a mobile phone); an Internet Protocol (IP) address; a cookie ID; the advertising identifier of a mobile phone; data held by a hospital or doctor, which could be a symbol that uniquely identifies a person. Conversely, personal data are not (for example): a company registration number; an email address such as info@company.com; anonymised data. **Comment**: the question of whether data relate to a certain person is something that has to be answered for each specific data item on its own merits[9].
Special categories of personal data	Personal data directly or indirectly revealing: racial or ethnic origin; political opinions; religious or philosophical beliefs; trade union membership; genetic data; biometric data; data concerning health; data concerning a natural person's sex life or sexual orientation[10].	"Genetic data" means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question; "biometric data" means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data; "data concerning health" means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status[11].
Non-personal data	Data other than personal data.	Guidance defines these data by origin either as: data which originally did not relate to an identified or identifiable natural person, such as data on weather conditions and air pollution generated by sensors installed on wind turbines or data on maintenance needs for industrial machines; and

[8] Article 4(1) GDPR.

[9] WP29 Opinion 4/2007 on the concept of personal data.

[10] Article 9(1) GDPR.

[11] Article 4 GDPR.

		data which were initially personal data, but were later made anonymous[12].
Mixed dataset	Dataset or a model that contains at least one personal data point.	**Comment**: Mixed datasets represent the majority of datasets used in the data economy and are common because of technological developments such as the Internet of Things (i.e. digitally connecting objects), artificial intelligence and technologies enabling big data analytics.[13] GDPR must be observed for the personal data part of the set[14]. **Example**: data related to the Internet of Things, where some of the data allow assumptions to be made about identifiable individuals (e.g. presence at a particular address and usage patterns).
Mixed dataset with inextricably linked personal and non-personal data	Situation whereby a dataset contains personal data as well as non-personal data and separating the two would either be impossible or considered by the controller to be economically inefficient or not technically feasible[15].	**Comment:** if at least one personal data point is inextricably linked to the non-personal data in a given mixed dataset, the whole dataset falls under the definition of "personal data"[16]. Separating the dataset may decrease the value of the dataset significantly. In addition, the changing nature of data (dynamic data) makes it more difficult to clearly differentiate and thus separate between different categories of data. In practice, mixed datasets will generally be considered personal data[17]. **Example:** when buying CRM and sales reporting systems, the company would have to duplicate its cost on software by purchasing separate software for CRM (personal data) and sales reporting systems (aggregated/non-personal data) based on the CRM data.

[12] Communication from the Commission to the European Parliament and the Council - Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union - COM(2019)250.

[13] Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union; COM(2019) 250 final, page 8.

[14] *Commission Staff Working Document, Impact Assessment accompanying the document Proposal for a Regulation of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union* (SWD(2017) 304 final), part 1/2, p. 3, 'regardless of how much of personal data are included in mixed datasets, GDPR [the General Data Protection Regulation] needs to be fully complied with in respect to the personal data part of the set.

[15] Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union; COM(2019) 250 final, page 10.

[16] Article 2(2) of the Free Flow of Non-Personal Data Regulation: "In the case of a dataset composed of both personal and non-personal data, this Regulation applies to the non-personal data part of the dataset. Where personal and non-personal data in a dataset are inextricably linked, this Regulation shall not prejudice the application of Regulation (EU) 2016/679."

[17] Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union; COM(2019) 250 final, page 10.

Confidential data	Data subject to an obligation of professional secrecy regulated by Union or Member State law, including a statutory obligation of secrecy.[18]	
Trade secrets	Information which meets all of the following requirements: (a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) it has commercial value because it is secret; (c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret[19].	**Examples**: undisclosed know-how and business or technological information. **Comment**: The definition of trade secret excludes trivial information and the experience and skills gained by employees in the normal course of their employment, and also excludes information which is generally known among, or is readily accessible to, persons within the circles that normally deal with the kind of information in question[20].
Business confidential data	For DUET purposes, intermediate versions of DUET consortium project data and datasets are deemed business confidential, irrespective of the license that the consortium establishes for final datasets[21].	It is important to distinguish this concept from the concept of “data confidentiality and integrity” (see **Table 5**), trade secrets, confidential data (see above), and other intellectual property (IP) related concepts.
Aggregate data	Aggregation refers to a data mining process in statistics. Information is only viewable in groups and as part of a summary, not per the individual. Aggregate data may, but also may not be personal data depending on the circumstance[22].	**Comment:** Aggregate-level data is useful for answering research questions about populations or groups of people. This reduces privacy risks, but aggregation of a sample that is too small can lead to privacy issues. GDPR puts emphasis on the fact that aggregate data, statistical results or the personal data are not used in support of measures or decisions regarding any particular natural person.[23] **Example:** aggregate counts of people in an office space can be used in combination with other data, such as weather data, to create an energy-efficiency program so consumption is controlled, with the goal of saving money and reducing energy use.
Anonymized/de-identified data	anonymisation means the process of changing data/documents into anonymous data/documents which do not relate to an identified or identifiable natural person, or the process of rendering	**Comment:** 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.

[18] Article 14(5)(d) GDPR.

[19] Article 2(1) of the Directive 2016/943 (EU) on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.

[20] Ibid, recital 14.

[21] D8.3 (Data Management and Modelling Plan), p. 21.

[22] U.S. Federal Trade Commission: Is aggregate data always private? (Available at https://www.ftc.gov/news-events/blogs/techftc/2012/05/aggregate-data-always-private).

[23] Recital 162 GDPR.

	personal data anonymous in such a manner that the data subject is not or no longer identifiable[24].	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[25]. Sufficiently robustly anonymized data are not personal data.
Pseudonymized data	"Pseudonymisation" means 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[26].	**Comment:** Pseudonymization is not a method of anonymization. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure. Pseudonymized data is personal data. Encryption can be considered among the pseudonymization techniques.

3.2.2 Data categories according to origin and purpose[27]

Table 2: Data according to origin and purpose

Category	Definition
Original data	Data produced by a DUET partner organisation (e.g., during a dissemination action or a pilot activity).
DUET existing data	Existing data already in possession of the DUET consortium and/or individual members of it prior to the project's initiation.
Existing third party data:	Data sourced/procured by the DUET consortium and/or individual members of it during the project's timeline.
IoT data	Smart cities use numerous resources such as sensors, cameras, mobile devices, etc. to collect data, route them through gateways and networks and eventually story them in a database[28].
Historical IoT data	Type of sensor data. The historical data set is a large volume of data typically indexed according to time and geographical dimensions. The historical data is mainly used to train digital twin models and visualize the past.
Context IoT data	The context data contains values as currently measured by the different devices. The context data is used as input for simulations and to visualize the present.

[24] Article 2(7) of the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

[25] Recital 26 GDPR.

[26] Article 4(5) GDPR.

[27] *See also* D8.3 (Data Management and Modelling Plan), p. 6 and 34.

[28] D8.3 (Data Management and Modelling Plan), p. 38.

Location data	Data indicating the geographic position of a person or the terminal equipment of a person (user)[29].
Modelling data	Modelling data contains all data related do models and interactions[30].
Smart data	Data or datasets extracted from larger amounts of data (big data, raw data) using algorithms according to certain structures, in order to provide meaningful information, understandable to the user, in order to help users achieve meaningful results[31]. They may combine data coming from sensors, social media and other human-related sources and thus may contain personal data, including special categories of personal data.
Provided personal data	Data provided directly by the individuals concerned (such as responses to a questionnaire).
Observed personal data	Data observed about the individuals (such as location data collected via an application).
Derived/inferred personal data	Derived or inferred data such as a profile of the individual that has already been created (e.g. a credit score).
Dynamic data	Data/documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data[32].
Research data	Data/documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results[33].
Metadata	Metadata is data that provides information about other data[34]. For example, draft ePrivacy Regulation defines electronic communications metadata as “data processed by means of electronic communications services for the purposes of transmitting, distributing or exchanging electronic communications content; including data used to trace and identify the source and destination of a communication, data on the location of the device generated in the context of providing electronic communications services, and the date, time, duration and the type of communication”[35]. According to ISO, geospatial metadata “provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services”[36].

[29] Recital 14 of Directive 2002/58: Location data may refer to the latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded.

[30] D8.3 (Data Management and Modelling Plan)

[31] This definition is digested from narrative at p. 22 od D8.3 (Data Management and Modelling Plan).

[32] Article 2(8) of the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

[33] Article 2(9) of the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

[34] https://en.wikipedia.org/wiki/Metadata.

[35] Proposal for a regulation concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), European Council, 6 March 2020.

[36] ISO 19115:2013 "Geographic Information – Metadata".

3.2.3 Types of data processing operations

According to the GDPR, processing of data means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction[37].

This GDPR definition is intentionally broad in order to cover as many types of data processing as possible. It is meaningful to use this GDPR list (slightly consolidated and expanded) to look in more detail on these types of processing operations. Such approach may help readers to break down their internal data handling processes into steps at which various legal requirements may apply. Even though the GDPR applies only to personal data processing, this data operation typology can be used equally in the field of non-personal data and may inform the analysis.

Table 3: Data processing operations

Data operation	Description
Collection	Acquiring/creating data by asking questions and collecting responses (including via an online form), collecting data from sensors (other than recording), scraping the web.
Recording	Acquiring/creating over data by recording natural persons by means of audio-visual recording, taking photos, recording by a dictaphone, recording phone calls, keeping record of a meeting, recording that you have a person's consent for a particular type of processing of personal data.
Obtaining data from a third party data provider – open data/public information	Data accessible and open without any restrictions, or data accessible by an unlimited number of interested parties subject to applicable licensing conditions (open license access).
Obtaining data from a third party data provider/vendor (non-open data)	Purchasing data or otherwise individually negotiating access to data (individual access license).
Organisation and structuring	Sorting/grouping of data according to certain characteristics or logic, creating a filing system, creating a database.
Storage	Storing data in physical depositories or in the cloud, keeping data for longer, e.g. not erasing the data after they had been processed for a respective task. This can involve pseudonymization or encryption of data for secure storage.
Adaptation or alteration	Changing the nature, contents, quality of the data or metadata by correcting errors or updating the data. Typically done in order to maintain data accuracy. This activity may be done by you, but you may also allow users to alter data related to them via access to a personal account on a website.
Consultation and use	Use of data for making a decision, drawing a conclusion, forming an opinion, use of data (feeding) in algorithmic decision making or machine learning operations and systems.

[37] Article 4(2) GDPR.

Disclosure by transmission	Sharing of data with other organisations (other companies or authorities), but also within your organisation with different branches/sections/departments. This may include uploading of data to a cloud drive.
Dissemination or otherwise making available	Disclosure to the public or a wider group of recipients by means of e.g., webpage, generally available APIs, open database.
Alignment or combination	Integration or combination of data in a dataset (pre-existing or new), alignment of data so two datasets can interact.
Restriction	Marking of stored data with the aim of limiting their processing in the future[38].
Erasure or destruction	Erasure of data which are no longer needed for the envisaged purpose; removal from search index. Can be done individually or en masse, ad hoc or at regular points where different categories of data get erased. This can involve destruction of physical documents, media or hard drives.

3.2.4 Other data management and privacy related concepts

Table 4: Other data management and privacy related concepts

Concept	Description
Data subject	Identified or identifiable natural person, subject of personal data related to that person.
Privacy by design	Implementation, at the time of the determination of the means for processing and at the time of the processing itself, of appropriate technical and organisational measures to protect the rights of data subjects[39].
Privacy by default	Implementation of 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. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility[40].
Filing system	A structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis[41].
Personal data breach	Breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed[42].
Terminal equipment	Equipment directly or indirectly connected to the interface of a public telecommunications network to send, process or receive information; in either case (direct or indirect), the connection may be made by wire, optical fibre or electromagnetically; a connection is indirect if equipment is placed between the terminal

[38] Article 4(3) GDPR.

[39] Article 25(1) GDPR.

[40] Article 25(2) GDPR.

[41] Article 4(6) GDPR.

[42] Article 4(12) GDPR.

	and the interface of the network[43]. Examples: mobile phones, laptops, tablets, connected devices (connected vehicles).
Data Protection Officer/DPO	A person tasked by an organisation (a data controller) with ensuring compliance with the GDPR and other privacy related tasks.
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
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[44].
Processor	Natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller[45].
Data Protection Impact Assessment/DPIA	An assessment to evaluate the origin, nature, particularity and severity of risk to the rights and freedoms of natural persons[46].
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[47].
Machine-to-machine service/IoT services	Services involving an automated transfer of data and information between devices or software-based applications with limited or no human interaction[48]. These services may be considered an "electronic communication service"[49] and thus be subject to ePrivacy laws.

3.3 Overarching principles

The following set of principles is a digest from applicable law and DUET project policies, adherence to which should contribute to legally compliant data-driven decision making. Some selected principles do not strictly follow from the applicable law, but derive from a conscious choice by the DUET consortium to adhere to certain ethical standards, or may represent a suggestion to adhere to a legal standard even in situations in which the law is not strictly speaking applicable. An example of the latter approach is our suggestion to adhere to the GDPR principle of "data minimisation" even where one does not deal with personal data. This

[43] Article 1(1) of the Directive 2008/63 on competition in the markets in telecommunications terminal equipment.

[44] Article 4(7) GDPR.

[45] Article 4(8) GDPR.

[46] Recital 84 GDPR.

[47] Article 4(4) GDPR.

[48] Recital 12 of the proposal for a regulation concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), European Council, 6 March 2020.

[49] Article 2(4) of the Directive (EU) 2018/1972 establishing the European Electronic Communications Code.

may help to reduce the amount of data used for a particular decision to the necessary minimum, which may in turn reduce residual liability for wrong decisions.

3.3.1 Use or consultation of data is "data processing"

According to the GDPR, the use and consultation of, as well as other operations run on, personal data is considered processing of personal data[50]. This is so Irrespective of whether the processing is done manually (to the extent it is done for filing system purposes (3.2 Definitions) or partly or wholly by automated means, or by autonomous systems such as AI (which is a form of "automated means of processing").[51]
This has a threefold significance:

a) The GDPR is fully applicable to each process of data use and consultation, or any other operation that involves data processing. As 3.2.3 Types of data processing operations shows, most if not all conceivable data operations will be included in the scope of "data processing" activities. Therefore, any use of personal data in decision-making will likely qualify as personal data processing.

b) In the eyes of the law, various decision-making stages of processing can be considered as a "separate kind of processing"[52]. This means that all GDPR requirements can apply in full with regard to that separate processing. For example, the fact that certain data has originally been collected and processed lawfully (i.e., with a sufficient legal basis and for a specified purpose, among other requirements) does not automatically imply lawfulness of any further (separate) processing steps, which will be part of a decision-making process. If a data driven-decision making involves a string of processes, each of which may make use of personal data, the GDPR will likely apply through the entire decision-making process and compliance must be ensured at every step.

c) GDPR principles of personal data management may be considered a good practice also for handling non-personal data in a decision-making process. For example, following the data minimisation principle may help reduce risk of wrong decisions and legal liability even where no personal data are involved.

3.3.2 GDPR principles on data processing

There are seven core GDPR principles, which apply whenever personal data is processed, including when dealing with a mixed dataset[53]. It may also be meaningful to extend adherence to these principles even where non-personal data is concerned, because it may contribute to full legal compliance and achieve the adequate ethical standard in the decision-making process and decrease the overall exposure to legal liability risks. Such an approach should, however, be balanced against practicalities and possibilities of working with non-personal data, and should not lead to unnecessarily prohibitive or inhibitive policies, but rather underpin the use of non-personal datasets to their full potential.

[50] Article 4(2) GDPR.

[51] Article 2(1) GDPR: This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.

[52] *See*, e.g., EDPB Guidelines 3/2019 on processing of personal data through video devices, point 51: "Any disclosure of personal data is a separate kind of processing of personal data for which the controller needs to have a legal basis in Article 6."

[53] Article 5 GDPR.

Table 5: Core GDPR principles

Principle	Comment
Lawfulness, fairness and transparency	Personal data must be processed in a lawful and transparent manner, ensuring fairness towards the individuals whose personal data is being processed. Lawfulness of processing means that personal data processing is lawful only if and to the extent at least one of the legal bases for processing set out by the GDPR applies[54].
Purpose limitation	Personal data may be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. There is a set of purposes, for which further processing is deemed compatible with the original purpose; these include archiving purposes in the public interest, scientific or historical research, and statistical purposes.
Data minimisation	Personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.
Accuracy	Personal data must be accurate and where necessary kept up-to-date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.
Storage limitation	Personal data is stored for no longer than necessary for the purposes for which the data is processed. Longer keeping may be permissible for archiving purposes in the public interest, scientific or historical research, and statistical purposes.
Integrity and confidentiality	Organisations handling the data must use appropriate technical and organisational measures that ensure an appropriate security of personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage.
Accountability	Organisations handling the data are responsible for, and are able to demonstrate compliance with the above principles.

3.3.3 Privacy by design

When personal data is involved, it is mandatory for data controllers to implement privacy safeguards into the code, method, manner or technique of data collection and processing. These privacy safeguards serve to implement in design the above-stated GDPR principles of data processing, such as data minimisation, or implementation of appropriate technical and organisational measures, such as pseudonymization, or encryption by default.

3.3.4 Privacy by default. Data minimisation principle

The data minimisation principle means that only those data are used or consulted for decision-making that are necessary. Processing of any excess data is unnecessary, thereby creating unnecessary risks, which may vary from hacking to unreliable inferences resulting in incorrect, wrongful, and potentially dangerous decisions.[55] The European Commission also noted that *"generating and processing less data limits the*

[54] Article 6 GDPR.

security risks. Therefore the compliance with data minimisation measures also provides for security safeguards."[56] Adhering to the data minimisation principle can therefore be recommended as a good practice for handling non-personal data as well.

Privacy by default concept is closely linked to the data minimisation principle, and in the GDPR terms it requires implementation of 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[57].

3.3.5 Precautionary principle

The precautionary principle is, broadly speaking, an approach to innovations with potential for causing harm when extensive scientific knowledge on the matter is lacking[58]. In the context of legal necessities, it can be interpreted as that where applicability of a stricter legal standard is unclear to the matter at hand, the stricter standard should be followed. For example, for DUET's or smart city's data management purposes, when it is unclear whether a decision is made based on personal or non-personal data, it should be assumed that personal data are at stake and the decision-making process should be fully GDPR-compliant.

3.3.6 Data ownership goes hand in hand with an organisation's responsibility for data management

DUET has chosen to adhere to a principle that as far as DUET research results are concerned, "*ownership goes hand in hand with the responsibility for data management*" (see D8.3 (Data Management and Modelling Plan), p. 10). We consider that this principle may be extrapolated here as a rule of thumb when it comes to data ownership and, by extension, decision ownership that includes that data.

We note, however, that in the eyes of the law, data ownership does not always result in responsibility for data management (e.g., personal data controller or processor may be another organisation), for the data-driven decision or any related legal liability.

3.3.7 Non-commercialisation of personal profiles

DUET has chosen to refrain from commercialisation of personal data. D8.3 (Data Management and Modelling Plan) declares at p. 7 that "*DUET consortium rejects the commercialisation of the personal profiles of human beings as a non-ethical practice***"**. In practice, this means that no licensing rules will be set on non-anonymised data (i.e., personal data), which means (we assume) that no such data will be shared with or sold to third parties or the general public.

We note that such requirement does not strictly follow from the applicable law, and personal profiles may under certain conditions be commercialised. The GDPR broadly recognises processing of personal data for

[55] Mireille Hildebrandt, *Primitives of legal protection in the era of data-driven Platforms,* " Geo L. Tech. Rev. 252 (2018).

[56] European Commission Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection C(2020) 2523 final.

[57] Article 25(2) GDPR.

[58] https://en.wikipedia.org/wiki/Precautionary_principle#European_Union. Applicability of the "precautionary principle" as a general principle of EU law may follow from certain case-law of the Court of Justice of the EU (e.g., Case T-74/00 *Artegodan*).

direct marketing purposes as a legitimate purpose, even though it gives data subjects a right to object against such processing[59]. We understand, therefore, that the above-stated principle constitutes an ethical principled choice above the applicable legal standard adopted by the DUET consortium.

3.3.8 Anonymised data preference principle

Using anonymous data has the advantage that no personal data is processed and the GDPR does not apply to such processing.

Randomization and generalization (e.g., aggregation or K-anonymity) are the main anonymization techniques. Using data processed with such techniques may significantly reduce risk of impact on individuals' privacy and of regulatory non-compliance[60]. The emerging DUET Data Management Plan also recognised this: "*Non anonymised data" is a temporary status as mentioned in the DUET Data Management Lifecycle. This temporary status must be as limited as possible since it creates a kind of a "grey zone" in terms of data management*[61]". Using personal data unless strictly necessary is also in line with the principle of data minimisation. Smart cities are therefore advised to follow that principle where possible. Where data cannot be fully anonymised (or it is at odds with the processing purpose), pseudonymization (e.g., encryption) techniques can be used instead, but note that such data continue to be treated as personal data. Pseudonymization is not a method of anonymization. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.

There are typically three risks related to sufficient robustness of anonymization techniques:

- *Singling out*, which corresponds to the possibility to isolate some or all records which identify an individual in the dataset;
- *Linkability*, which is the ability to link, at least, two records concerning the same data subject or a group of data subjects (either in the same database or in two different databases). If an attacker can establish (e.g. by means of correlation analysis) that two records are assigned to a same group of individuals but cannot single out individuals in this group, the technique provides resistance against "singling out" but not against linkability; and
- *Inference*, which is the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes[62].

In addition, certain data types (such as location data or IoT data) carry higher risk of re-identification (particularly with help of big data analytics; in such areas, there may be higher requirements on what constitutes anonymised data and on the processes of correct anonymization. The emerging guide provides further information on this topic.

For detailed information on anonymization, pseudonymization and other data sanitisation techniques and their risks, consult the guidance: **Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques**.

[59] Recital 47 GDPR.

[60] E.g., in its Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, the European Data Protection Board emphasises that *"when it comes to using location data, preference should always be given to the processing of anonymised data rather than personal data.*"

[61] D8.3. (Data Management and Modelling Plan), p. 16.

[62] Article 29 Working Party Opinion 05/2014 on Anonymisation Techniques.

4. Commented summary of pilot city organisations' responses to data questionnaires

This section provides a summary of initial responses to questionnaires on data handling processes within pilot organisations (in case of Flanders, including IMEC) for the DUET project purposes. These first questionnaires were fashioned as relatively open ended, hence the responses differed in focus and granularity of information.

Subsections with summaries of the teams' responses are accompanied with a short commentary from the perspective of legal issues described above in the theoretical part, and further set out in the emerging "Cities Guide" in **Chapter 5** below; cross references to the rest of this document are given where appropriate.

We took note of the caveat that the initial information provided by partner organisations may be subject to change in line with the future development of pilot activities, acquisition of new datasets, change of approach, etc.

4.1 Short description of each pilot city organisation and its data management role

Athens

DAEM S.A. (City of Athens IT Company) is the oldest and most significant Local Government IT Company, in operation since 1983. It aims at providing Cloud based multi-platform e-Governance to local government organizations, public administration and other authorities and organizations. The development and promotion of new innovative services which are fundamental to the smart and sustainable city idea, lies at the heart of DAEM's interest and is a strategic objective at the city level.

More specifically, DAEM supports public administration - mainly the City of Athens but also other municipalities in Greece - and designs software and applications targeting city services. Hence, DAEM acquires technical expertise in software development and engineering, IT infrastructure development and maintenance, development of open source IT products, solutions and applications.

Flanders

Informatie Vlaanderen (Dutch, translated: Informatie Vlaanderen) is part of the Flemish Government, in Belgium. AIV is a public body tasked with support in the areas of digitization of data in e-government, GIS and public information. Informatie Vlaanderen is responsible for several project-related domains:

- Policy support on digitization, information acquisition, information access;
- Responsible for service creation, service management and access in the GIS and e-government domains;
- Solution realization and optimization in cooperation with the government, private organizations and
- companies in the fields of GIS and e-government;
- The realisation of a central information platform of services for citizens, organizations and companies - https://joinup.ec.europa.eu/community/epractice/case/magda-20-platform;
- Supporting government organizations (direct or indirect, via other intermediate organizations) of citizens, organizations and companies;

- Supporting government organizations to improve their internal processes and services by simplification
- and digitalisation to enhance a better level of service to their (external) customers;
- Delivery of a central data exchange platform to maximize the usage of government information focussed on information about persons, organisations and companies;
- Information exchange on government services including real estate data of the Flemish government;
- Delivery and access to geographical information via an efficient GDI (Geo-Data Infrastructure) platform including all the inspire datasets and other (authentic) geo-data sources (many of them are real estate oriented or can be used in the field of real estate context);
- Establishing a digital framework to support and stimulate the realisation of a single and unique information infrastructure.

Pilsen

Správa Informačních Technologií města Plzně (SITMP), as a part of the city of Plzeň, is a public company responsible for ICT of the city Plzeň. SITMP manages key business software of city organisations: SAP, MS, AGENDIO, GIS, eSPIS, SOL, etc. SITMP provides services for city hall, city parts, city police and around 100 city companies. It operates IoT on the LoRa platform in Pilsen, using it for city companies and for research.

SITMP has 210 million CZK yearly turnover and 108 employees, its technical capabilities include 6000 PCs, 150 km of optical metropolitan network, around 1400 requests from customers/monthly, 2 data centres (level TIER 3).

4.2 Access to data, data ownership and data management

The initial input indicates that all pilot city organisations share the responsibility for, or at least have some say in, acquisition, purchase, processing, storage, and dissemination of data needed for the DUET project. This should enable them to (co-) determine data usage for DUET purposes. Decision making process steps described below in which such competence is required are discussed by the guide in B. What decision and who makes it; D. Ownership issues (decision, data, IP); J. Accountability, fairness, transparency; and K. Collaboration and trust. On the other hand, no pre-established procedures for handling data in decision-making processes seem to have been established for or within the pilot organisations, which underscores the impact potential of this deliverables set.

Athens

DAEM as a private organization supporting the City of Athens services, is mainly active in developing software for the Municipality and also offering backend maintenance. Accordingly, the city and citizens data is not owned or processed by DAEM. In parallel, DAEM maintains and administers the hardware that hosts the services, namely servers etc. Any other ownership of data by DAEM is under compliance with the legislation and GDPR regulation. For that purpose, the Legal Department and DPO are monitoring all activities. Specifically, referring to the EU Projects Sector, any possible data included in funded projects is ensured to follow the law compliance and GDPR both internally by DAEM and its departments, and also by the projects consortia according to the directives of the EU for each project.

In its responses to the questionnaire, DAEM indicated that it may seek to acquire (traffic) data for DUET purposes, which indicates some control over data management at least with regard to DUET activities.

Flanders

There are basically four ownership models: (i) Datasets owned 100% by AIV; (ii) Datasets with shared ownership AIV + 3rd party; (iii) Datasets opened up by the AIV for a third party; and (iv) Datasets bought from the private market. The idea is to open as much as possible datasets as open data or as services with an SLA against cost-price. Agreements are made with all datasets where a third party is involved. AIV has traditionally a big impact on the data availability and distribution in Flanders. AIV acts as an enabler and is delivering services for other government agencies to increase data distribution. AIV has also a legal role as service integrated and trusted third-party. AIV plays a key role in the distribution of (personal) data between agencies, controlling security, logging and in general privacy.

A steering committee led by AIV with other government agencies involved, decides about the data disclosure conditions and the recognition of a dataset as an authentic source. Negotiating the disclosure conditions with the private sector and 3rd parties who have invested a lot themselves in datasets (for ex the utility sector) can be difficult.

Regarding smart cities, only a part of the available data is owned by the government. Setting up a strategy of negotiating with the private sector as the public sector (as a whole) is still a challenge.

Pilsen

City data sources involve: (i) data created by the city by its own (road and air quality sensor data, 3D models of selected buildings); (ii) data managed and provided by the state, Pilsen Region or other state organisations, used by the city (cadaster, RUIAN, digital technical map, FCD); and (iii) data purchased (mostly through public procurements - e.g. 3D model of the city). The idea is to open as much as possible datasets as open data or as services with an SLA against cost-price. Agreements are made with all datasets where a third party is involved.

SITMP is responsible for the purchase, storage, distribution, access and administration of data for all city organisations. SITMP has a big impact on the data availability and distribution in Pilsen. SITMP acts as an enabler and is delivering services for other city organisations to increase data distribution. SITMP has also a legal advisory role as a trusted third-party and plays a key role in the distribution of (personal) data between users, controlling security, logging and in general privacy. In public procurements, SITMP defines the conditions for data ownership and licensing and thus can control how data is used afterwards.

4.3 Access to Data Protection Officers and other legal resources

All pilot cities organisations responded that they have, or have access to, a Data Protection Officer (DPO), and a lawyer/legal department.

This is essential for any compliant data driven decision-making, which involves personal data. The emerging guide seeks to provide some core information and background against which compliance with privacy protection requirements can be checked, but it is important that smart city teams coordinate closely with their DPOs and legal departments if they have any doubts.

The following table summarizes the pilot cities' perceived experience in handling of personal data:

Personal data processing experience	Occasional	Regular	DPO involvement
Athens	✓		✓
Flanders		✓	✓

Pilsen		✓	✓

The guide discusses the importance of setting the right escalation path for adequate risk impact management in B. What decision and who makes it.

4.4 Some potential use of personal data/sets for DUET purposes

Even though each pilot city organisation occasionally (Athens), or regularly (Flanders, Pilsen) collects and processes personal data, there appears no clear intention - at the moment - to make an extensive use of personal data/sets for DUET purposes.

The following table shows pilots' indicated intentions with regard to the use of personal data:

	Athens	**Flanders**	**Pilsen**
Create/collect personal data	Yes	Yes	Yes
Purchase data with personal data included	No (traffic data purchase possibility)	Possibly	No
Share personal data with DUET partners	No	Yes, on the basis of a processing agreement	No (e.g., ANPR data will only be shared anonymised)
Share personal data with other third parties	No (Athens will only disseminate project activities and outcomes, not datasets)	Yes, if strict necessity and with government agencies only, if legally possible	No

In addition, the following was understood from the pilots' responses:

All three pilot cities confirmed that they may be collecting data via questionnaires/user surveys/meetings. Such methods usually involve processing of personal data (at least before the surveys or recordings are anonymized, unless it is possible to make them anonymous at source). The GDPR may apply to the process of collection and any further processing of thusly collected data (see G. Risks in further data processing), unless they are made anonymous before further processing. E.g., **Pilsen** indicated that such information will be received only as anonymous.

Athens indicated that it may consider acquiring some traffic data, not yet specified at the time.

Flanders indicated that it may possibly engage in the collection/creation of data for DUET purposes, not yet specified at the time.

Pilsen indicated that it may collect/create data via the following means: air quality sensors, traffic sensors (magnetic loops, FCD) or speed control zones (ANPR data). Some remarks on these:

Air quality sensors and magnetic loop sensors (or other purely count sensors, such as noise level) typically do not generate personal data. However, considerations set out by the guide for use of non-personal data may still apply (see mainly I. Data minimisation, adequacy of data use, and the related 3.3.4 Privacy by default. Data minimisation principle).

ANPR data are typically considered personal data (although approaches may slightly differ among EU Member States) if they are linked or linkable to identifiable persons - e.g. car owners or drivers. ANPR images may further have defects (or even features) that reveal personal information (see also E. Data factual quality (properties)). **Pilsen** confirmed that these defects/features are indeed possible.

- **Flanders** indicated that it may seek access to a third party (police force) provided ANPR data and that this data will be anonymised. Flanders is further contemplating use of other plate number database vehicle information for concrete smart city applications, such as vehicle type, fuel type, euro norm, CO2 exhaust, or weight category.
- **Pilsen** acknowledged that it collects ANPR data that is personal data, but indicated that it will share with DUET only data that is anonymized.

FCD (floating car data) comes from connected cars (as the **Pilsen** team confirmed, the source is GPS location + mobile network connectivity). In case of **Pilsen**, this data will be collected and provided for DUET purposes by a third party (Ředitelství silnic a dálnic) on a license agreement basis, and will be provided anonymised. If an organisation participates in the original data collection via GPS location or mobile networks, it may be subject to ePrivacy law requirements (see 2.2.2 ePrivacy legislation), which does not seem to be the case with Pilsen at this stage.

Use of anonymized and/or aggregated, and particularly third-party provided ANPR, FCD or other data in the plate number database or should not raise any particular issues. See further 3.3.8 Anonymised data preference principle. However, to the extent there may be issues with robustness of the anonymisation method used, residual risks exist. See also F. Data legal quality and G. Risks in further data processing. The **Flanders** team raised one such vehicle related information use issue as an example: *specific categorisation actions will be necessary to avoid recognizing individual vehicles like the first truck driving on hydrogen*.

Collection, creation, and even purchase of personal data typically means that the organisation will be considered a “controller” for the GDPR purposes (see 3.2 Definitions).

Sharing of personal data with any third party, including among DUET partner organisations, must be done in full compliance with the GDPR, including when such data is shared (outsourced) for processing purposes only - such third party would be considered a “processor” (see 3.2 Definitions). For these cases, adequate data sharing and processing agreements must be put in place (as **Flanders** envisages to do). In case of joint controllership over personal data, adequate agreements must also be put in place. The guide addresses these issues in K. Collaboration and trust.

Data sharing (including non-personal data) is further considered by the guide in K. Collaboration and trust, and may further be extensively covered by the emerging **DUET Data Management and Modelling Plan (WP8)**.

4.5 Low potential use of special categories of personal data for DUET purposes

All pilot cities organisations responded that they do not - at this stage - envisage collecting or processing of any special categories of personal data (see 3.2 Definitions) for DUET purposes.

Flanders and **Pilsen** indicated that they may potentially use aggregated health data, for example for the air quality use cases. Such aggregation can take the form of, e.g., aggregation of data per city district (as indicated by **Pilsen**).

In case the aggregation is sufficiently robust and there is no reasonable risk of re-identification, such data should not be treated as personal data (including any special category of personal data) and its processing should not raise much privacy risks. However, see 3.3.8 Anonymised data preference principle section for possible issues with anonymisation and aggregation techniques, also discussed by the guide in F. Data legal quality and G. Risks in further data processing especially in the HPC, non-anonymized location data and sensor fusion risk contexts.

4.6 Data sanitization techniques

Big data typically requires some sanitization and cleaning in order to remove unnecessary personal data, excessive granularity, noise, etc. Techniques and challenges of these methods are briefly described in 3.3.8 Anonymised data preference principle. Robustness of, e.g., anonymisation or aggregation of data may be particularly challenging in certain contexts such as IoT data or location data - these issues are further described in the guide chapters on F. Data legal quality; G. Risks in further data processing and I. Data minimisation, adequacy of data use.

Pilot cities organisations indicated that they are at least theoretically acquainted with these techniques, which the following table summarizes. The **Athens** team is familiar with some anonymisation and data encryption techniques, but more detailed information is currently unavailable to the pilot team.

Data sanitization techniques used	**Aggregation**	**Selection (omitting)**	**Anonymisation/blurring of images**	**Pseudonymisation**
Athens	✓	-	-	✓
Flanders	✓	✓	✓	✓
Pilsen	✓	✓	✓	✓

4.7 Use of third-party data vs. original data

DUET activities will predominantly rely on existing data provided by third parties (even if some may be considered “colleague organisations”, such as other public authorities or bodies within the same government structure). This is evident - at this stage - from the overview of user stories and contemplated data sources presented by **D2.2 Scenario specifications of the DUET solution,** and from the information gathered from pilot organisations based on data questionnaires. There may be important exceptions, however, for example sensor data collected by **Flanders** (IMEC) or from sensors deployed by **Pilsen**.

Third-party data

Whether personal or non-personal, third-party provided data may carry less privacy impact risk and other risks for DUET, because the third party that is sharing or selling the data is primarily responsible for legal compliance of the dataset. The third-party provider should also specify the purposes for which the data may further legally be used. For the approach to third-party data, see mainly F. Data legal quality as regards a careful presumption of low risk, and see G. Risks in further data processing on the question of who may be liable for defective data.

Third-party data may only be used with a sufficient legal basis, i.e., they have been purchased (there is an appropriate user license), provided for free based on an individually negotiated license/access, or they are generally available for free (open access). The guide deals with the issue of ownership and possible intellectual property/trade secrets infringements in D. Ownership issues (decision, data, IP).

Third parties should only share personal data based on adequate data sharing or license agreements; the guide discusses this in K. Collaboration and trust.

Original data

The respective DUET partner organisation will be responsible for the original data it generated/collected and provided for DUET purposes. The **DUET Data Management and Modelling Plan (WP8)** may extensively deal with the issues of compliant collecting, processing and sharing of data.

4.8 Specific data source/purposes envisaged for DUET purposes

There are certain areas in which collection or use of data may pose various risks, including significant privacy risks. The initial information collected from pilot cities indicates that for DUET purposes, certain risky areas are not within DUET's purview at the moment; this includes the use of non-anonymised telecom data, scraping of social media data, or the use of profiling (see 3.2 Definitions). It has been posited also that the consortium wishes to avoid commercialization of individual person's profiles altogether (3.3.7 Non-commercialisation of personal profiles).

Certain DUET partners may, however, engage in use of specific data types that may call for a more detailed legal assessment of their individual user cases or their source. The following table provides an overview and is followed with a commentary on four data types: telecom data, telraam data (location data) terminal equipment data (location data), and IoT/machine-to-machine communication services.

	Telecom data	**Social media scraping**	**Profiling**	**ANPR data**	**Location data**	**Terminal equipment data**	**IoT/machine -to-machine services**
Athens	No	No	No	No	Yes (GIS city data)	No	No
Flanders	Yes - Proximus mobile phone data, BEMobile, Mobile data for detecting car traffic	No	No	Yes (anonymised by the police)	Yes (telraam, ANPR, FCD, traffic counting, air quality, parking sensors)	Yes - WiFi interface identifiers; CityMesh data	No
Pilsen	No	No	No	Yes (will be anonymised for DUET purposes	Yes (traffic, air quality, noise data, FCD)	No	Yes - open LoRa platform

ANPR, FCD, air, noise sensors and magnetic loops as traffic data sources were discussed above in 4.4 Some potential use of personal data/sets for DUET purposes.

Telecom data

Flanders (IMEC) submitted that it envisages processing of traffic data based on telecom operators' gathered mobile phone data or mobile car data. Per IMEC's (and AIV's) explanations, this data will be highly abstract, and therefore unlikely to raise any privacy risks. Telecom operators (as public electronic communication service providers) are subject to ePrivacy rules (see 2.2.2 ePrivacy legislation), which require that such data are further processed only when they are made anonymous (unless users' consent exists). The guide discusses this in F. Data legal quality.

Telraam data

Flanders (AIV) submitted that it intends to use third-party provided Telraam data. Telraam is a combination of sensors, low-resolution camera and a processing microchip[63]. According to the team, *Telraam interprets camera data to count pedestrians, cyclists, cars and lorries. The privacy-by-design solution, allows the owner of the device to only see the camera pictures when installing the device (to outline the camera) and only counts are transmitted to the central server. The raw camera data is not exposed to anybody (except during installation).* It follows that only anonymised data (counts) will be acquired from a third party for DUET purposes, and that in any event the described privacy-by-design solution limits collection of personal data by these systems to minimum. This is a good example of a privacy enhancing design, the guide picks this up in G. Risks in further data processing.

Scanning of unique WiFi identifiers (terminal equipment)

Flanders (IMEC) submitted that it may use traffic data based on scanning of users' phone WiFi interface identifiers (MAC addresses). This data will be used for the purposes of Digital Twin City Flows integration, and IMEC confirmed that no personal data will be involved, as the data is aggregated as counts within a certain interval. However, MAC addresses are typically considered as personal data, and as explained in 2.2.2 ePrivacy legislation, additional legal requirements may apply to electronic communications and terminal equipment data processing. Processing of WiFi and other unique identifiers of users' terminal equipment (e.g. phones) may trigger such responsibilities (in concurrence to applicable GDPR requirements, as these unique identifiers are typically considered personal data). Conversely, if DUET partners are not responsible for these transmissions or the original data collection, but only purchase such data already processed by third parties, these additional legal necessities are unlikely to arise. The guide addresses this in F. Data legal quality.

Machine-to-machine / IoT services

Similarly to the above comments on terminal equipment data, machine-to-machine communications in the emerging IoT/cities of things context (see 3.2 Definitions) can also be subject to the ePrivacy rules, because they may fall within the definition of publicly available electronic communication services. **Pilsen's** open LoRa platform offering to the interested public conveyance of data from IoT devices over a range of available networks (WiFi, Ethernet, 3G, LTE, Bluetooth, LoRaWAN) over to data centers[64] may be an example of such services. However, if DUET does not become directly involved in processing of data in cities' machine-to-machine dataflows conducted via public networks, the ePrivacy requirements on these services should not add further regulatory burden on DUET's activities. It would be the partner organisation's responsibility to provide to DUET or integrate in DUET systems only data that is compliant with the applicable law (the guide addresses this in D. Ownership issues (decision, data, IP)). The guide addresses machine-to-machine communications in G. Risks in further data processing.

We anticipate and will actively seek further discussion with responsible DUET partner organisations about these specific data sources.

[63] https://telraam.net/en/what-is-telraam.

[64] https://iot.plzen.eu/, available in Czech only.

5. "Cities Guide to Legal Compliance for Data-Driven Decision Making" (first version)

This section contains the first version of a guidance that is to become a "Cities Guide to Legal Compliance for Data-Driven Decision Making." Guidance given at each step in this first version is subject to additions and review as we gain more knowledge of DUET's work with datasets and models, and are provided more detailed information about decision-making processes at pilot organisations, with which we are working in parallel.

At this stage of understanding DUET's activities, a typical data-driven decision or policy-making consists of several steps, which usually follow in certain logical order. Issues of "what", "why", "who", "how" may be addressed by decision-makers in the following steps: (i) type of decision to be made; (ii) type of the decision-making process that will lead to that decision; (iii) who takes ownership over a particular decision involving use of data; (iv) identify risks that a particular decision may induce; (v) set out steps that lead towards finalizing the decision; etc. The following guide attempts to link these logical steps with selected risk areas of data handling, such as (i) issues with data ownership; (ii) data quality (factual properties and legal aspects); (iii) GDPR compliance of processing; (iv) data purpose limitation and data minimisation; (v) accountability in use of data; (vi) security of data and processes; etc.

This first version is written up in a simple textual form and a bullet-point format. Future versions of this guide may include more advanced graphics, diagrams, decision trees or similar, in order to provide a concise, effective and easy to use guidance in line with the original deliverable intention. In addition, future versions of this guide may include a separate subsection to deal, in more detail, with the following aspects of data management: collecting, storing, (processing), and sharing of data. At this stage, it appears that these aspects will be covered by WP8 working streams and in particular, the emerging Data Management and Modelling Plan. We will hold discussions in order to avoid overlaps and/or achieve consistency between the responsible WP leads going forward.

*   *   *

A. Introductory remarks

- Any decision or policy making based on data, data models or data analytics by a smart city should represent legally compliant solutions that reflect the smart city's objectives as well as public values, and serve the public good and interests.

- In particular, a smart city's handling of data for decision-making must (1) conform to the applicable legislation in the area (regulatory compliance, "safe harbour") and (2) should take into account potential risks of legal liability for any harm done by wrong decisions based on data or data analytics or caused by the defect in the data itself (legal liability prevention and management). Consequences of a breach in either field may expose the organisation concerned to legal enforcement by public authorities as well as private actors, and lead to administrative or criminal penalties, restrictive court orders, damages claims, and – last but not least – reputational damage. However, smart cities and stakeholders should not be discouraged to make data-driven decisions; on the contrary, evidence suggests that data-driven decisions tend to be better decisions.

- The following guide is not intended to constitute legal advice; instead, all information, content, and materials in this guide are for informational purposes only. Given that this document got finalized at a certain point in time, information in it may not constitute the most up-to-date legal or other information. Readers of this document should consult a legal officer at their organisation, or contact an attorney qualified in the concerned jurisdictions to obtain advice with respect to any particular legal matter.

- Finally, this guide should also not be taken as an exhaustive overview of all possible data-driven decision making legal issues. Rather, it presents a selection of risk areas, practical relevance of which is being confirmed with DUET piloting teams and via further learning process. Organisations' legal and Data Protection officers should be engaged to give tailored advice on any particular legal matter.

B. What decision and who makes it

Key aspects: hierarchy and reach matter; GDPR at each step if personal data involved; higher privacy impact risk with new technology and large scale data operations; legally binding measures/policies.

"*The Digital Twin provides a risk-free experimentation environment to inform stakeholders what they need to do with the assets in the real word in order to both achieve the most effective long-term policy outcomes, and short-term operational decisions.*" (from DUET presentation materials)

Data may inform decision-makers at various stages of a smart city development. Consider the type, scope and status of the to-be-made decision. Properly name roles and assign responsibilities. This may all co-determine the legal requirements on a compliant decision-making and also impact on potential legal consequences of any problem with the decision or how data is used to drive it.

- **Executive/operational/minor decision**. Smaller scale decisions may be subject to less stringent or copious legal requirements and carry lower risk of legal liability. However, consider this:
    - *Is personal data involved?* GDPR is applicable at each step that may be considered processing of personal data[65]. Even a single, isolated breach of GDPR may lead to enforcement and penalties, which can be very high and commensurate with the economic size of your organisation. → **engage your Data Protection Officer**
    - *How high in organisation hierarchy is the decision made?* An executive decision made high up may have wider impact and legal consequences down the line.
    - *How is data presented?* The way data is presented as the result of a decision-making process may influence future decision making processes, thus causing any potential issue to proliferate.

      → establish an escalation path to your superiors for cases where serious issues or threats of data breaches occur

[65] See Section 3.2.3.

- **Policy/strategic/major decision**. A wider impact decision can take the following forms: (i) organisation-wide policy document, (ii) decision embedded or integrated in a data model, (iii) decision used for machine learning or to set parameters of an AI system, or (iv) document shared with or sold to third parties, or published to (or even made binding on) the general public. Such decisions carry higher risk, will typically have higher legal requirements, and may cause a more wide-spread damage. Consider this:
    - *Is personal data involved?* Large-scale operations with personal data may have a greater impact on data subjects' privacy. GDPR may require so-called "Data Protection Impact Assessment" to be carried out in certain high-risk areas:
        - Use of new technologies (e.g., IoT applications);
        - Systematic monitoring of a publicly accessible area on a large scale (including with CCTV cameras or sensors able to collect personal data);
        - Systematic and extensive evaluation of personal aspects based on automated processing, including profiling (3.2 Definitions), and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
        - Processing on a large scale of special categories of data (3.2 Definitions) and data relating to criminal convictions and offences.

          → **engage your Data Protection Officer**. Consult appropriate guidance (**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**).
    - *Large-scale data = large-scale liability.* Any issues with data properties quality or legal quality (e.g., data collected or processed in breach of GDPR or licensing conditions) may proliferate in large-scale data operations, or when implemented in a decision with wider reach. Potential for organisation-wide legal liability for regulatory non-compliance and any damage caused.

      → **Identify the appropriate person/role/body within your organisation hierarchy to make or approve strategic-level and policy decisions**
    - *Legally binding measures/policy imposing a limitation to the fundamental right to privacy*. If you are a public authority and consider (and have the requisite power of) adopting a policy with legal (or quasi-legal) effects on the general public that also involves data processing, such type of decision may be subject to additional requirements imposed by EU and international law on states to ensure individuals' fundamental right to privacy. You may wish to consult the following step-by-step guidance**: EDPS, "Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit", 11 April 2017**, and **EDPS, "Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data", 25 February 2019.**

C. Type of decision-making process

Key aspects: processing is typically automated and triggers GDPR if personal data is involved; emphasis must be put on accountability and security of automated systems; risk of data re-identification; algorithms should be open and fair (ethical requirement).

- Human factor / manual decision making processes may be prone to human factor mistakes and lack of control.
    - *Is personal data involved?* Purely manual, non-automated processing of personal data falls out of scope of GDPR, unless it is intended to form part of a "filing system" (3.2 Definitions), or you create written records in a manual filing system.
    - *Difficult cases in distinguishing manual versus automated decisions*: a model presents visualised data (e.g., based on running a query on the Digital Twin system), and I make the decision on its basis. Is it manual or (partly) automatized?

      → the **precautionary principle** guides to adhere to the stricter legal and ethical standard, which in this case may mean that the decision making should comply with GDPR and higher standards for ensuring accountability, transparency and security of systems (see below).

- Automated decision-making processes may be more demanding in ensuring accountability and transparency (see furtherJ. Accountability, fairness, transparency).
    - *Is personal data involved?*
        - Automated processing of this data always triggers GDPR application. N.B. processing information in an electronic form (e.g., inputting data points into a computer software, such as MS Office) is considered automated.
        - *Risk of re-identification.* When non-personal data goes through machine learning or deep learning processing, there is still a chance it can be linked to an individual person thanks to the advanced computational capabilities. Data can thus dynamically become personal data again.
        - GDPR places special requirements on *automated individual decision-making* and *profiling* (3.2 Definitions), which may produce legal or similarly significant effects concerning data subjects (3.2 Definitions), for example e-recruiting practices without any human intervention. Data subjects must be informed about such automated decision-making upfront.

          → **engage your Data Protection Officer**
    - Only partly automated processes do not typically lower the standards required for the process, including legal necessities (e.g., GDPR remains fully applicable to the entire processing).
    - *Fully algorithmic/autonomous decisions*: model makes the decision, I only implement. Currently, there are significant gaps in AI regulation. (This will be further addressed by deliverable D1.5. (Ethical Principles for using Data-Driven Decision in the Cloud.)

      → **Ensure algorithms driving HPC analytics are open and fair**

D. Ownership issues (decision, data, IP)

Key aspects: distribution of liability; legal gap for AI systems; data ownership principle; IP ownership; Article 26.1 of the Grant Agreement.

Decision ownership is key for distribution of responsibility and (legal) liability for the decision to be made. Multiple factors may indicate or affect ownership from the legal perspective: data ownership, IP ownership; AI ownership. Liability distribution is particularly difficult in complex decisions involving multiple actors or complex IT systems.

In practice, the decision owner and data/rights owner and user may be one and the same organisation, but many other possibilities exist. For example: **organisation X** may be the decision owner (e.g., determines the purposes and means of the data-based decision making, it would be the "controller" in the GDPR sense (3.2 Definitions); **organisation Y** owns the data but makes them available for use to **X**; and **organisation Z** may be "using" the data on X's behalf (it would be the "processor" in the GDPR sense (3.2 Definitions). It is worth clarifying this upfront.

- *DUET activities*: assign clear decision ownership to the right DUET partner organisation.

- *Legal liability*. Owner of a decision will typically be liable for any legal consequences that decision (or the decision-making process) causes. While the primary liability is with the responsible organisation, a secondary liability may go after the individual persons/employees involved in the decision-making process.

- *Autonomous /AI systems*: make clear(er) who controls the AI used. Be transparent about these issues (see also K. Collaboration and trust).
  - There is currently a legal gap on how liability is assigned in case of autonomous systems. A rule of thumb could be the so-called *vicarious liability* – liable is the organisation that owns and/or controls the system.
  - Joint ownership and joint legal liability (and "controllership", if personal data is involved) is possible under the law; there may even be a degree of control by an end-user (of a consumer-oriented application, for example, where the end user is responsible for steps he/she wishes to take).

- *Data use.* In order to use data for decision-making, you must be either the (i) data owner, or (ii) have some other right to use the data (e.g., commercial license, open data license, free open access).
  - As regards DUET: *data ownership goes hand in hand with the responsibility for data management* (D8.3).
    - This may be a rule of thumb, but may not always correspond to how applicable law attributes data management to data ownership, and vice versa. A smart city may involve multiple interacting data flows or multiple data owners/"controllers". E.g., DUET system can use data owned by another organisation. IoT data may often be co-created and thus jointly owned.

- *Is personal data involved?* Even if you are the data owner, you may or may not also be the "controller" for GDPR purposes (3.2 Definitions). If you intend to use personal data for decision-making, you will typically be the "controller".
  - *If you are not the data owner*: make sure you have the right to use the data (commercial license, open data license, or the data is publically available with no conditions attached); if not, contact the data owner. For a list of access rules to existing DUET data ("background") → **see Annex 4 to D8.3**. (Data Management and Modelling Plan).
- *Data use when third party data are covered by IP rights or trade secrets.* Concerning data where your organisation is not an owner, and this data is covered by an IP right, or a trade secret (3.2 Definitions), on which third parties may have a claim, particular potential risks may emerge. For a thorough overview of those risks, please see Deliverable 1.1. In this context, the above-mentioned licencing contracts that Duet needs to enter into with third parties IP rights' holders (or third parties who are licenced the IP right from the rightholder) must specifically allow Duet using that data, covered by IP rights and/or trade secrets, in a way which is compatible with the IP owners' rights and which thus allows Duet's making use of that data in a way which is IP compliant. It is paramount that Duet partners have procedures and organisational aspects in place to enter into IP compliant licencing contracts. **Engage commercial lawyers in your organisation and, when in doubt, external IP attorneys qualified to practice in your jurisdiction with negotiating licencing. Carry out due diligence when engaging into contractual aspects with third party data users (who are not the direct owners of the IP or trade secret underpinning that data). In the context of negotiating licencing, depending on the rules of the pilots' jurisdictions, consider specific clauses in the context of licencing which shield Duet's partners from liability for third party IP rights' breaches. Consider also potential intellectual property coverage (i.e. liability insurance against these risks) to minimise risk of exposure for third party IP rights' breaches.**

- *Ownership over created Intellectual Property (IP) rights and trade secrets.* It is possible that a new IP right gets created in the course of making a decision or as its result. Ownership of such IP rights will typically accrue to the organisation making the decision or policy, but difficult cases may arise, particularly if the user doesn't own the data used to drive the decision-making process. Joint ownership is a possibility.
  - For DUET purposes, per **Article 26.1 of the Grant Agreement**, results (including any IP rights attached to them) are owned by the beneficiary that generates the results. Joint ownership occurs *if (a) two or more beneficiaries have jointly generated the results, and (b) it is not possible to: (i) establish the respective contribution of each beneficiary, or (ii) separate them for the purpose of applying for, obtaining or maintaining their protection.*
  - If unsure about IP rights implications of your decision → **engage legal officers at your organisation. IP attorneys qualified to practice in concerned jurisdictions may need to be engaged.**

E. Data factual quality (properties)

Key aspects: factual data properties impact on legal necessities; GDPR imposes certain data properties quality standards, ANPR data example.

Data quality may have a direct impact on the quality as well as legality of the decision or the decision-making process. Before starting the decision-making process, consider:

- *Data properties*:
    - relevance: the usefulness of the data for the specific decision-making process.
    - clarity: the availability of a clear and shared definition for the data.
    - consistency: the compatibility of the same type of data from different sources.
    - timeliness: the availability of data at the time required and how up to date that data is.
    - accuracy: how close to the truth the data is.
    - completeness: how much of the required data is available.
    - accessibility: where, how, and to whom the data is available or not available (e.g., security).

- *Is personal data involved?* When processing personal data, ensuring that the data processed is relevant, up to date, accurate and secure is mandatory by law (see GDPR principles of data minimisation, accuracy, and integrity and confidentiality).

- *Practical example with ANPR data*. There are significant amounts of information an ANPR (automatic number plate recognition) system can collect. For example, is the system just recording vehicle registration marks? Or is it recording images of vehicles, occupants or 'patch plates' as well? If it's the latter, make sure the amount of information being collected is justifiable.

F. Data legal quality

Key aspects: typical legal defects: GDPR breach; ePrivacy rules breach (state-level differences); license infringement; IP rights infringement; original data (primary responsibility of your organisation) vs. 3rd party data (primary responsibility of the 3rd party organisation); full data audit; limited data audit; location data (preference for anonymous data principle); web/social media data; ANPR data, smart data.

Data, datasets and models may suffer from various legal defects. If defective data is used to drive a decision, that decision may become tainted with risk of regulatory non-compliance, or the use of defective data may cause the decision-making process to become flawed and lead to wrong (harmful) decisions, thus creating a causal nexus between defective data and harm caused.

- *Kinds of legal defects*:
    - data collected or processed in breach of the GDPR or ePrivacy legislation (e.g., personal data collected without sufficient legal basis; non-anonymized location data shared by telecom operators, wrongly anonymised data);
    - data acquired or processed in breach of licensing conditions (contractual breach), or without sufficient license (e.g., misused, stolen data);
    - use of data infringing Intellectual Property (IP) rights of a 3rd party.

- *Original data* (3.2 Definitions). Your organisation is primarily responsible for legal compliance of such data. This should be primarily achieved through compliance with the data collection requirements laid down by the Data Management Plan. Prior to using original data, check that they are defects-free to reduce any residual risks.

- *Existing data.* The following complements the general guidance of the DUET Data Management Plan: *In case of reuse of existing data, i.e. owned by someone else (a third party or another DUET partner), the individual or joint responsibility is to* ***check the nature of data*** [⋯] *and* ***undertake the consequent actions*** [per the Data Management Plan] (D8.3)**.**
  - *DUET existing data* (3.2 Definitions). If unsure about legal compliance of this data, check with the data providing partner organisation. If unsure about GDPR-compliance of your organisation's existing personal datasets, conduct a *data audit*. This should involve identifying:
    - categories of personal data you process;
    - location where it is stored (hard disks, cloud drives, physical filing cabinets);
    - inflows (sources of personal data – cameras, sensors, web forms, email, phone calls);
    - outflows (third parties with whom you share data – public authorities, private enterprises, individuals, cloud storage companies).

    → engage your Data Protection Officer to oversee the audit
  - *Existing third party data* (3.2 Definitions).
    - *3rd party responsibility*. Third party organisations should primarily be responsible for the data they shared. Therefore, you may make a careful assumption that the data, when provided for further re-use, is free from legal defects, because the 3rd party provider/vendor is obligated by law to collect, process, and share data lawfully and for legitimate purposes only. This assumption may not apply, however, if the data provider is established outside of the EU, because it may be subject to lower legal standards.
    - *Limited data audit*. It may be practically difficult for a decision-maker to review the entire history of an acquired dataset. However, a limited data audit may help decrease the residual risk:
      - Does the data come from a reliable source? (reputable vendor or a public authority). Does the data come from a provider established in the EU?
      - Doesn't the data suffer from defects in important properties? (accurateness, timeliness, consistency, etc., see E. Data factual quality (properties)
      - Isn't the data manifestly unfit for the required purpose?
      - Is the data shared under reasonable licensing conditions?
      - Are there issues with correct anonymization or pseudonymization? (If you processed non- or insufficiently anonymised personal data by accident → **engage your Data Protection Officer**)
    - *Licensing conditions*. Further use of data may be limited by licensing conditions. This may include limitation of use purposes, manner of processing, data retention periods, or possibilities further to share or disseminate the data.

- *Location data.* → **Work with anonymous data, where possible** (anonymised data preference principle).
    - *Location data origin.*
        - *Telecom providers*. Less risk with data acquired from electronic communication services providers. They are obligated by law to anonymise location data before sharing them with 3rd parties. Such anonymous data can be re-used for further processing, including for modelling purposes.
        - *Location data collected from users' terminal equipment* (3.2 Definitions).
            - You may collect such data, e.g., by help of an app installed or cookie placed on an end user's mobile phone, → **you are primarily responsible under the law for lawful collection and processing of such data.** Any further processing (i.e. not the initial processing strictly necessary to provision of the end-user service/app) of non-anonymised location data is only possible with the terminal equipment's end user's consent (an additional consent with any such further re-use), or based on an exception provided for by the law → **engage your Data Protection Officer**
            - *Statistical counting/device fingerprinting.* The above principle (user consent required) currently also applies to activities such as tracking of physical movement by scanning of users' equipment's Wi-Fi interface or other unique identifiers (such as MAC address, which is typically considered as personal data). Legal necessities in this area may change (relax) in the future, however, so if you engage in such data collection activities → **discuss with your Data Protection Officer. Consult appropriate guidance (WP29 Opinion 9/2014 on the application of Directive 2002/58/EC to device fingerprinting)**
            - *Floating car data (FCD).* This time stamped geolocation and speed data may originate in connected vehicles (e.g., via their communication with the GPS or via mobile network connection). Collecting such data may trigger application of the ePrivacy rules (location data is collected via electronic communication service, or by accessing data on users' terminal equipment - mobile phone or the connected vehicle). Processing of anonymised FCD data should typically raise no issues, but beware of flaws in anonymisation (see below, and also 3.3.8 Anonymised data preference principle).
        - *Sensitive data.* HPC processing of location data may allow the data controller to draw far reaching inferences from the data, which reveal, e.g., life habits of data subjects. This creates risk of processing of special categories of personal data (3.2 Definitions) without sufficient legal basis. If you suspect this may be the outcome of your decision-making process → **engage your Data Protection Officer**
        - Note that where location data processing falls in scope of ePrivacy rules (2.2.2 ePrivacy legislation), there may be various country-specific differences in how that legislation has been transposed into the applicable laws.
    - *Possible flaws in anonymisation*. Seemingly anonymous location data may be vulnerable to re-identification. If your organisation is handling the anonymization process → **consult EDPB Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak.** (The European Data Protection Board recommends conducting a

reasonability test, which takes into account both objective aspects (time, technical means) and contextual elements that may vary case by case (rarity of a phenomenon including population density, nature and volume of data). If the data fails to pass this test, then it has not been anonymised and therefore remains in the scope of the GDPR).

- *Web / social media scraping* (“*voluntary data*”; public registries of persons). This is typically personal data → **engage your Data Protection Officer. Data Protection Impact Assessment may be required for large-scale data processing.** Posting on social media publicly does not give third parties sufficient legal basis for further processing of the published data without informing the data subjects.
    - Best available legal bases may be:
        - *Processing is necessary for compliance with a legal obligation to which the controller is subject*. This will typically be available for public authorities and other bodies entrusted with public administration tasks.
        - *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.* Potential argumentation lines include: public safety (e.g., municipal traffic management), economic well-being (e.g., optimization of services of general interest such as electricity, water and waste management), public budget interests. This set of examples is not exhaustive.
        - *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*.
        - *Data subjects’ consent*. There are country-specific differences in minimum age for statutory legal capacity to give valid consent with personal data processing (Belgium: 13 years, Czech Republic: 15 years, Greece: 15 years). It is generally advisable to avoid scraping under age minors’ data from the web, unless this is objectively justified, e.g., by the specific purpose of the decision/policy.
    - You may be required to provide certain information to data subjects, see **J. Accountability, fairness, transparency**
    - *Personal data from public registries (open data).* Open data legislation at the level of EU Member States may provide a list of documents or public data that may contain personal data (typically public registries), but which can be freely re-used for commercial or non-commercial purposes. No or low impact on data subjects’ fundamental freedoms and interests is typically presumed in such cases. If unsure whether you can re-use personal data in public registries → **check with your Data Protection Officer**

- *Automatic Number Plate Recognition data (ANPR).* ANPR data will be considered personal data to the extent it allows a vehicle and the related information (location, speed, photo of the driver) to be tracked to its registered owner, operator or the driver. Other data from the number plate recognition database may include vehicle type, fuel type, euro norm, CO2 exhaust, or weight category. Processing ANPR data fully anonymised should not raise privacy risks, but there may be specific cases in which they can (typically, if the anonymisation suffers from the “singling out” issue. See 3.3.8 Anonymised data preference principle).

- *Smart data* (3.2 Definitions). Self-check questions: do I understand the algorithms and underlying (raw) dataset used to produce the smart data? Are these algorithms fair? If you consider that the algorithms/machine learning processes or raw data used to generate smart data:
    - are manifestly intransparent, unfair, suffer from obvious factual or legal defects, have wrongly set parameters or flawed methodology → **escalate the issue to appropriate role within your organisation, engage your Data Protection Officer or a legal officer**
    - contain personal data that may have been processed incorrectly/unlawfully, you have suspicion of a personal data breach → **engage your Data Protection Officer**

G. Risks in further data processing

Key aspects: regulatory non-compliance, liability for damages, flaws in processing data within the decision-making process; personal data (legal basis, country-specific derogations); location data; IoT data; liability for defective data; liability risk limitation (contractual, non-contractual liability).

Main legal risks in data-driven decisions are regulatory non-compliance (risks of administrative or criminal sanctions) and liability for damages. Main ethical risks lie primarily in the field of soft sanctions, such as reputational damage, impact on project timeline and success; or indirect harm such as loss of opportunities, jobs, etc.

Any data-driven decision may result in such outcomes due to: (i) flawed objectives, values or methodology chosen for the decision-making process; (ii) flaws (factual, legal) in the data used for making the decision; or (iii) flaws in the further processing of data for purposes of the current decision-making process.

- *Objectives/values/methodology:* this issue is beyond the scope of this guidance focused on legal necessities, but a smart city should never lose these from sight in its policymaking activities.

- *Data flaws*: see E. Data factual quality (properties) and F. Data legal quality issues. The following steps assume that data selected to drive the decision are free from any factual or legal defects, including that personal data was originally collected in line with the GDPR.

- *Flaws in further data processing (the decision-making process).*
    - *Is personal data involved?* Further processing must fully comply with the GDPR. Three particular overarching principles (among others) must not be breached:
        - *Lawful processing.* The decision-making process is likely to be considered a separate kind of processing, for which the controller (3.2 Definitions) needs to have a legal basis under the GDPR. → **engage your Data Protection Officer**
        - *Purpose limitation*. The purpose of further processing must not be incompatible with the purpose for which data was originally collected. There are certain purposes which may be deemed compatible, such as scientific research or statistical purposes. See further H. Purpose limitation.
        - *Data minimisation*. Data for driving a particular decision must be adequate, relevant and limited to what is necessary in relation to the purposes of that decision. See further I. Data minimisation, adequacy of data use

- *Country-specific GDPR derogations.* Individual EU Member States may derogate from certain GDPR rules for specific purposes, e.g., national security, defence, public security, prevention of crime, other important objectives of general public interest (monetary, budgetary and taxation matters, public health and social security), etc.
    - A suggested gap analysis consists of these steps
        - Identify which EU Member States jurisdictions are applicable to your processing activities.
        - Conduct a gap analysis to understand what needs to be done to update your policies or decision making processes. It may be necessary **to engage privacy law attorneys qualified in the respective jurisdictions** to make a more detailed case-by-case assessment of these requirements.
        - States may change their laws from time to time – stay updated.
    - Examples of derogations or special rules in the pilot jurisdictions:
        - *Belgium:*
            - If the federal police transfer personal data to any other public authority or private organisation, this transfer must be formalised by an agreement between the federal police and the controller who receives the data.
            - In the event of processing by a federal authority, a specific Data Protection Impact Assessment must be conducted prior to the start of the processing activity, even if a general Data Protection Impact Assessment has already been conducted.
        - *Czech Republic*:
            - Provided that the processing is carried out for the purpose of a "protected interest", controllers and processors are exempt from the obligation to carry out a compatibility test (see also H. Purpose limitation) regarding the purposes of processing. These include various public policy interests and enforcement of private claims.
            - Public "controllers" are exempt from the obligation to carry out a Data Privacy Impact Assessment provided that the controller is obliged to carry out the processing of personal data by applicable law.
        - *Greece*:
            - Public bodies may process special categories of personal data (3.2 Definitions) to the extent necessary for reasons of substantial public interest, national or public security; or the implementation of humanitarian measures.
            - Disclosure of personal data, including special categories of personal data (3.2 Definitions), from public bodies to private entities is permitted to the extent necessary for the performance of tasks vested with the public body.
            - Obligations to provide certain information to data subjects are derogated if it would jeopardise the proper fulfilment of the tasks of the public body

        - Public bodies may process personal data for new purposes and don’t need to run the compatibility test, if necessary: to validate the data provided by the data subject, where there exists reasonable doubt as to its accuracy; for purposes of national security, public security or taxation; for the prosecution of criminal offences; to prevent serious violations of rights of third persons; and to generate official statistics.

- *(Geo)location data* → **work with anonymous data, where possible** (anonymised data preference principle)
    - *Flaws in anonymisation*. Anonymous location data are vulnerable to re-identification (see also F. Data legal quality).
    - *Revealing sensitive information.* E.g., HPC processing of geolocation data from connected vehicles may reveal life habits of data subjects, and create other far reaching inferences. This creates risk of processing of special categories of personal data (3.2 Definitions) without sufficient legal basis. If you suspect this may be the outcome of your decision-making process → **engage your Data Protection Officer**
- *IoT data* → **work with anonymous or aggregate data, where possible** (anonymised data preference principle)
    - *Aggregate data* (3.2 Definitions) carry less privacy impact risk. If aggregate data can be considered fully anonymized or de-identified, they are no longer personal data and the GDPR does not apply.
    - *Personal data containers* (*privacy-by-design and privacy by-default principles*). Where possible, have data collected and processed by devices locally. For example, use sensors that allow anonymization (e.g. blurring of images, aggregation, geo/masking) at source or shortly thereafter.
        - A good practice example is so-called “telraam data”, given by AIV in response to data questionnaires: *Telraam interprets camera data to count pedestrians, cyclists, cars and lorries. The privacy-by-design solution, allows the owner of the device to only see the camera pictures when installing the device (to outline the camera) and only counts are transmitted to the central server. The raw camera data is not exposed to anybody (except during installation).* For further details on video devices data issues, see **EDPB, Guidelines 3/2019 on processing of personal data through video devices**.
        - Note that personal data container solutions may not be feasible if cloud infrastructure is used to collect and process raw data from sensors/cameras (further aspects will be addressed by the deliverable D1.5. (Ethical Principles for using Data-Driven Decision in the Cloud).
    - *Sensor fusion risk.* Combining sensor data or data derived from different sources may lead to better and more precise information than would be possible when these sources are working in isolation. Increased possibility of risk that there is insufficient legal basis for such advanced processing of personal data, or that the purpose limitation or data minimisation principle will be exceeded. In addition, raw data can later be combined with other data incoming from other systems (e.g. CCTV or

internet logs). In such circumstances, some sensor data are particularly vulnerable to re-identification attacks.

- *Risk of excessive data collection.* IoT systems often lead to excessive data collection compared to what is necessary to achieve the purpose. Processing this data further may breach the *data minimisation principle*.
- *Historic* data may be less impactful also with regard to persons' privacy or commercial sensitivity. *Context and (near-)real time* data may, conversely, be more impactful. This may not be true in specific cases, but it may be useful to consider sensitiveness of the data processed in your risk analysis.
- *IoT/machine-to-machine communication services as electronic communication services*. Data that is transmitted through publicly available electronic communication networks may be subject to additional ePrivacy laws requirements (see 2.2.2 ePrivacy legislation), this may include machine-to-machine communication services/ "IoT" services (see 3.2 Definitions). Requirements include confidentiality of data flows, anonymisation, storage limitation and security requirements, and apply irrespective of personal or non-personal nature of the data processed. Collection and processing (including further processing) of, e.g., location data via these services is subject to special rules also. On the other hand, acquisition of data obtained from these sources by third parties for further processing, where the organisation plays no part in its original collection/transmission over communication networks, may escape application of ePrivacy rules, and will only be subject to the GDPR if personal data is involved.

- *Who is liable for defective data?* As a matter of principle, data owner/provider should be liable for the quality (factual, legal) of the data. Its use in your decision-making, may, however, bring your organisation within the scope of liability rules, if, for example, a third party relies on your decision or on the way the decision presents the data, and suffers damage.
  - *Examples of factors (co-)determining liability:* whether data is integrated into software – purely digital product (e.g., a licensable data model solution); whether data integrated into a device – product with some "physical existence" (e.g., a robot driven by data/software); whether data only used to inform a particular decision or a decision-making process.
  - *Contractual liability.* Where your organisation has a contract with the damaged party (e.g., licensee of your data model solution), it is likely that claims for damages linked to flawed data or flawed decision-making process will be targeted at your organisation. Further rights of redress against the flawed-data provider may be available to your organisation under the applicable law.
  - *Non-contractual liability.* Where no contract exists between your organisation and the damaged party, assignment of liability may be complex and there may be several extra-contractual liability regimes applicable in different EU Member States (as well as worldwide) at the same time. → **engage legal officer at your organisation if concerns exist**
  - *How to decrease risks of contractual or non-contractual liability*:
    - Impose appropriate licensing conditions, warnings and liability limitations on the recipients of your decisions (e.g., users of data model). For example, liability

limitation clauses may be included in your contracts with third parties. Liability limitations may be imposed in open access licenses as well.

- Comply with license conditions and limitations attached to the data used. Make sure that these conditions and limitations are transposed further (e.g., attached to the software or data model) if applicable to the use of your products by third parties.
- Risk-shifting agreements (e.g., insurance contracts) may be available to cover impacts of your decisions/products. (Note that an obligatory insurance scheme for certain categories of AI/robots may be introduced in the future.)
- *Precautionary approach principle* may be recommended particularly in cases of large-size data processing, which may cause wide-spread damage.
  → engage legal officer at your organisation if concerns exist

H. Purpose limitation

Key aspects: specifying the data processing purpose; original purpose, re-use purposes, compatibility test; compatibility presumptions; EU Member State purpose exemptions; terminal equipment data (ePrivacy rules: consent based re-use); non-personal data purpose setting.

Any data-driven decision will have its intended purpose. Identifying it will be necessary to deal with legal necessities for any personal data processing, but also serve as a benchmark against which to weigh the relevance and necessity of the data (even non-personal data) intended to be used (principle of data minimisation). Adhering to the *purpose limitation principle* helps to avoid the phenomenon of "function creep", which means use of data for a different goal than it was originally collected for. Ignoring this may result in a GDPR infringement (if personal data is involved) and may increase overall legal liability risks.

- *Is personal data involved?* Personal data cannot be processed unless for a specific purpose (*purpose limitation principle*). Such purpose must be specified (set prior to the processing), explicit (clearly communicated to the stakeholders, mainly the data subjects, at the time of collection/processing) and legitimate (must not follow illegitimate aims, such as unlawful discrimination). When re-using personal data for further processing (e.g., modelling) and decision-making, consider this:
  - *Purpose must still be specified*. This can be conflated with the phase of setting the objectives of your particular decision-making process/data model, but the purpose must still be sufficiently specified.
  - *Check sufficient legal basis*. If the original legal basis for data processing was consent, check if your current purpose was covered in the specified plausible purposes when the consent was given (at the data collection point). If not, a new legal basis must be identified (either a non-consent-based, or you may need to obtain new consents from the data subjects) → **engage your Data Protection Officer.**
  - *New purpose compatibility check*. If the new purpose is not covered by the original data subject's consent, further non-consent based processing is possible only if it fulfils the GDPR purpose compatibility check. **You may need oversight or approval of your Data Protection Officer**. The following questions must be considered:
    - any link between the purposes for which the personal data have been collected and the purposes of the intended further processing;

        - the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller;
        - the nature of the personal data, in particular whether special categories of personal data are processed (3.2 Definitions) , or whether personal data related to criminal convictions and offences are processed;
        - the possible consequences of the intended further processing for data subjects;
        - the existence of appropriate safeguards, which may include encryption or pseudonymization.
    - *Compatibility presumptions*
        - The GDPR deems three purposes for further processing purposes compatible, i.e., no need to run compatibility tests for these. Note that the *anonymised data preference principle* should still be observed unless it would prejudice the processing purposes.
            - archiving purposes in the public interest;
            - scientific or historical research;
            - statistical purposes.
    - *Country-specific derogations/exemptions*: Individual EU Member State's laws may also provide various legal exemptions from the GDPR standards on data processing for the purposes set out below. Where an exemption may apply to your processing, establish all the jurisdictions in which this processing takes place in order to be able to run an adequate gap assessment. Member States may change the law from time to time – it is advisable to keep track of concerned jurisdictions.
            - freedom of expression and information,
            - public access to official documents;
            - national identification numbers;
            - employee data;
            - professional secrecy obligations;
            - churches and religious associations.
    - Information to data subjects: the GDPR requires certain information to be provided prior to each new processing to the data subjects, see further J. Accountability, fairness, transparency

- *Terminal equipment data. (e.g., location data collected via a mobile app or from vehicles).* Re-use of data obtained from users' terminal equipment is <u>not</u> possible only based on the GDPR purpose compatibility test. End users' <u>additional consent</u> must be acquired before further processing of such data, e.g., for modelling purposes. → **engage your Data Protection Officer.** For example, telemetry data, which is collected during use of the vehicle for maintenance purposes may not be disclosed to motor insurance companies without the users consent for the purpose of creating driver profiles to offer driving behaviour based insurance policies.

- *Other data involved*. Even with no personal data involved, identifying purpose of data processing may be important to ensure:

  - Compliance with any data licensing conditions and limitations;
  - Ensuring only relevant and adequate data are used for the decision-making (*data minimisation principle*).

I. Data minimisation, adequacy of data use

Key aspects: GDPR principle of data minimisation; privacy by default; data extent and granularity; large-scale data processing additional requirements; pseudonymization and encryption; excessive IoT data; inside organisation application (Article 39.2 of the Grant Agreement).

Only those data should be used for decision-making that are necessary (*data minimisation principle*). Processing of any excess data is unnecessary, thereby creating unnecessary risks, which may vary from hacking to unreliable inferences resulting in incorrect, wrongful, and potentially dangerous decisions. This in turn may expose your organisation to unnecessary risks. Measures should be put in place that ensure data minimisation by default.

- Self-check: Am I using the maximum amount/extent/detail of data necessary for the intended purpose? Have I considered a less detailed/extensive dataset to achieve the same result? Is the data actually suitable to achieve the intended purpose?

- *Data granularity*: both extent and granularity of the data (e.g. data relating to individual persons as against aggregate data) matter in the data minimisation assessment.

- *Is personal data involved?* GDPR makes the by-default data minimisation principle mandatory.
  - Data minimisation and adequacy should be considered at each separate processing operation.
  - Large-scale processing of personal data may require a **Data Protection Impact Assessment → engage your Data Protection Officer**
  - The principle is mandatory also when further processing is for the *"compatible" purposes*, i.e., archiving purposes in the public interest, scientific or historical research and statistical purposes (see also H. Purpose limitation).
  - *Privacy-by-default* means that the controller (3.2 Definitions) must 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. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility.

- *IoT data*. IoT systems often lead to excessive data collection compared to what is necessary to achieve the purpose (this may breach the *data minimisation principle*).

- *Risk mitigations*:
  - *Pseudonymization* (3.2 Definitions) is a technical and organisational measure that the GDPR recognises can be helpful in safeguarding the data minimisation principle.

  - *Encryption and an anonymized/ aggregated data consultation process ("hybrid processing")*. It should be possible to encrypt personal information in a way that preserves the ability to run queries on the encrypted data. Analysts can ask questions that link together personal data, but they only ever see anonymized or aggregated results .

- *Inside organisation application*. Data minimisation principle applies also inside your organisation with regard to how much data is available to which teams or roles. → **Article 39.2 of the Grant Agreement** *(Processing of personal data by the beneficiaries): [⋯]**The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement**.*

J. Accountability, fairness, transparency

Key aspects: organisation must be able to demonstrate regulatory compliance (GDPR, ePrivacy, other applicable laws); information to data subjects; IoT transparency; measures to ensure accountability (data protection policies, processing agreements, documentation of data processing activities, record and report data breaches); DPIAs AI systems transparency and fairness; security by design.

Accountability and transparency rules are indispensable in order to allow leaping technological developments, such as big data and HPC analytics, to thrive in human rights-centric democratic societies.

- *Is personal data involved?* GDPR requires every organisation handling personal data to be able to demonstrate compliance with the GDPR rules and principles.
  - Full accountability is indispensable to facilitate enforcement of *data subjects' right*s regarding privacy and their personal data, including: rights to access their data; right to data rectification; erasure requests (right to be forgotten); and data portability rights.
  - *Transparency and fairness*. The GDPR asks controllers to be transparent about the functions and processing of personal data, and enable data subjects to monitor the data processing. The *fairness principle* specifically requires that personal data should not be collected and processed without the individual concerned being aware about each step of processing.
  - GDPR mandates *informing data subjects* (3.2 Definitions) also when controller (3.2 Definitions) intends to further process data for a purpose other than for which the personal data were originally obtained, and even where personal data have not been obtained from the data subjects. A specific set of information must be provided prior to that further processing (e.g., before you use the data to drive a decision). Unless one of the below four exceptions applies → **engage your Data Protection Officer to compile the necessary information package** (**Guidance is also available: WP29 Guidelines on transparency under Regulation 2016/679**)**.** Exceptions:
    - the data subject already has the information;
    - the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, or in so far the obligation to provide information is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the

controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available;

        - obtaining or disclosure of personal data is expressly laid down by Union or Member State law to which the controller is subject and which provides appropriate measures to protect the data subject's legitimate interests; or
        - where the personal data must remain confidential subject to an obligation of professional secrecy regulated by Union or Member State law, including a statutory obligation of secrecy.
    - *IoT transparency*. Sensors and cameras may intentionally be designed as non-obtrusive, as invisible as possible. This may undermine transparency vis-à-vis data subjects. Appropriate labelling and signs should be put in place by the sensor/camera operator. → **engage your Data Protection Officer if your organisation is responsible for the sensor/camera deployment**

- Measures that demonstrate compliance with the *accountability principle*. These measures should be implemented with the **help of your Data Protection Officer**.
    - *Adopt data protection policies*. For DUET purposes, WP8 aims at creating a **Data Management and Modelling Plan** covering the issues of data collection, processing, storage and sharing within the project lifecycles. For GDPR compliance purposes, these policies should be reviewed and updated where necessary (periodically or on an occasion of a significant change in approach or policy).
    - *Conclude written processor agreements*. Whenever you intend to use a processor (3.2 Definitions) to handle personal data on your behalf, a written agreement setting out each party's responsibilities and liabilities must be put in place → **engage your Data Protection Officer** (see also K. Collaboration and trust).
    - *Maintain documentation of data processing activities* (3.2 Definitions).
        - *Practical example:* analysts can ask questions that link together personal data in a data model. All the questions and answers should be recorded, creating an audit trail that allows regulators and courts to inspect how the data has been used and to penalize misuse.
        - In principle, there should be evidence of each personal data processing operation made, even if you "merely" wish to consult data or data models to inform a decision-making process.
        - It is a recommended good practice to maintain documentation even for non-personal data processing activities. This may help address issues of attributing legal liability to organisations and individuals, and decrease organisation-wide exposure to legal risks.
    - *Record and where necessary, report personal data breaches.* The GDPR sets out detailed requirements on data breaches reporting, should they nevertheless occur. Actors involved in personal data processing are obliged to prevent any such breaches from happening.
    - *Carry out Data Protection Impact Assessment (DPIA)* for uses of personal data that are likely to result in high risk to individuals' rights and freedoms. In particular, a DPIA may be required in cases of:

    - Use of new technologies (e.g., IoT applications);
    - Systematic monitoring of a publicly accessible area on a large scale (including with CCTV cameras or sensors able to collect personal data);
    - Systematic and extensive evaluation of personal aspects based on automated processing, including profiling (3.2 Definitions), and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
    - Processing on a large scale of special categories of data (3.2 Definitions) and data relating to criminal convictions and offences.
    - → **engage your Data Protection Officer**. Consult appropriate guidance (**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**)

- *Privacy/transparency aspects of fully automated decision making systems*. There is currently a legislative gap on how to comprehensively handle fully automated decision making systems, including AI, robots and machine learning systems. The applicable legislation may have certain scattered requirements, which are set out below. Ethical aspects of these issues will be covered by the deliverable D1.5. (Ethical Principles for using Data-Driven Decision in the Cloud).
    - "*Label bot as a bot*." Data subjects must be informed upfront when they are dealing with an automated system instead of a human being.
    - Data subjects have a 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. The GDPR does allow certain exceptions, however, in case appropriate safeguards are in place (e.g., data subjects may appeal to a human to review the automated decision).

K. Collaboration and trust

Key aspects: standardisation; standard licenses; contextual controls; contracts for personal data processing; use- restriction marking; data retention periods.

For a transparent and accountable data driven decision making, an open and transparent communication with all stakeholders with no hidden agendas is recommended. There is a number of good practices (and legal requirements, mainly in the personal data field) that enhance collaboration and build trust:

- *Standardisation*. As recognised by the D8.3 (Data Management and Modelling Plan) standardisation is a prerequisite for enterprise-wide data-centric initiatives.
    - *Standard licenses.* Open Data legislation allows public authorities to share data subject to standard licenses. To foster collaboration and competition, however, the license conditions must be objective, proportionate, non-discriminatory and justified on grounds of a public interest objective. They should not unnecessarily restrict possibilities for re-use and should not be used to restrict competition" (e.g., by preferring certain economic operators over others without any objective justification).

- Organisations should try to improve the way in which information is presented, including in a given adopted decision or a data model. The way data is presented as the result of a decision-making process may influence future decision making processes.

- *Contextual controls* are legal, organisational and technical measures that help address the risks of re-identification of anonymised personal data. Interconnected or collaborating organisations handling personal data may wish to implement the following:
    - Legal and organisational controls such as obligations between collaborating parties and/or internal policies adopted within one organisation aimed at directly reducing re-identification risks, e.g., contractual obligation not to re-identify or not to link, data use purpose limitations clauses, etc.
    - Security measures such as data access monitoring and restriction measures, auditing requirements, monitoring of queries, aimed at ensuring the de facto enforcement of the first set of controls;
    - Legal, organisational and technical controls relating to the sharing of datasets aimed at ensuring that the first set of legal controls are transferred to recipients of datasets. They include obligations to share the datasets with the same set of obligations or an obligation not to share the datasets, as well as technical measures such as encryption to make sure confidentiality of the data is maintained during the transfer of the datasets. These measures are used to balance the strength of data sanitisation techniques with the degree of data utility.

- *Contractual arrangements regarding personal data***.**
    - *Processor agreements.* If your organisation outsources processing of personal data to third party organisations ("processors", 3.2 Definitions), written contracts that meet GDPR and other legal requirements must be put in place. → **engage your Data Protection Officer.** D8.3 (Data Management and Modelling Plan) envisages that data sharing with DUET partners will be arranged through processing agreements.
    - *Joint controllers agreements.* Where more than one organisation is considered a "controller" (3.2 Definitions) with regard to particular data processing, the GDPR requires them to put in place an arrangement which, in a transparent manner, determines their respective responsibilities for compliance with the GDPR. → **engage your Data Protection Officer.**

- *Restriction marking/tags*. Mark stored data (particularly data to which third parties have access) with any processing limitations that you wish to apply in the future. This ensures that the data can only be used in certain pre-defined circumstances.

- *Set clear data retention periods*. Being transparent about how long you intend to store data for specified purposes is a good trust builder. As regards personal data, it is a GDPR requirement not to store data for longer than necessary for their processing (*storage limitation principle*).

6. Conclusions and future work

This document is the first in a series of three planned deliverables concerning "Cities Guide to Legal Compliance for Data-Driven Decision Making". The ambition of this document was to bring to the table the knowledge of legal landscape and requirements extensively sketched by deliverable D1.1, in order to kick start the transition from abstract overview of legal frameworks to their concrete application in practice.

In that vein, this document expanded slightly the deliverable's scope not only to provide a "guide", but sought to furnish also a hopefully friendly-to use list of definitions, concepts and principles in order to facilitate understanding of the emerging guidance. This effort could also tap into other DUET partners' contributions in this area, in particular the emerging Data Management and Modelling Plan (WP8), and a DUET-wide glossary of terms. The guide provided in the last section of this document can be understood as a skeleton of an emerging structured step-by-step guidance to help smart cities comply with the most pressing legal necessities in data driven decision making processes.

This first version document may be found lacking in a variety of issues:

a) The list of issues, concepts and definitions may be under-inclusive as regards issues specific to advanced data analytics and data management used by the smart city teams. The list may also be over-inclusive as regards certain concepts not encountered by smart cities in practice.

b) The list of overarching principles may require further discussion with DUET partners to include or exclude guidance above the legal minimums that may have a steering effect on a number of DUET's planned activities.

c) The guide's current focus is the legal necessities of a "data-driven decision making process". It does not deal specifically with issues of legal (and ethical) data collection, processing, storing and sharing as such, even though these issues do necessarily overlap to an extent. It may turn out practical to expand the guidance's scope to cover all stages of the data management lifecycle and reduce the scope (or level of detail) of the emerging Data Management Plan, or vice versa.

d) DUET is gradually developing understanding of the pilot activities, particularly as regards practicalities and possibilities of data driven processes. The first version of the guide may be found too abstract to give helpful advice to piloting teams or other personnel in the field. It may also be unnecessarily detailed in discussing situations that may relatively rarely be encountered in real life.

e) The textual and bullet-point format of the guide may be found unsatisfactory and some more intuitive formats may need to be explored to amend the future versions.

Future work will include:

a) Reaching a common understanding among DUET partners on the scope and a consistent approach regarding this "guide" and the Data Management and Modelling Plan (WP8).

b) Continue facts-gathering efforts to develop our understanding of DUET data management processes, mainly by help of targeted data questionnaires to both the piloting as well as central architecture teams. This knowledge should also help us to focus down most practically pressing issues in daily data decision making and tweak the guide in this direction.

c) Seeking feedback from piloting teams on the “guide”, validation of its usefulness in practice.

d) We contemplate initiating consultations with national Data Protection Authorities in pilot jurisdictions on selected topics. In particular, should DUET piloting teams turn out hands on collecting data from sensors and other IoT devices and then integrating that data into DUET datasets, there may arise challenging questions as regards application of the ePrivacy legislation and processing of electronic communications data. The legislative framework is considered outdated and its overhaul a moving target in an unclear stalled legislative process.