Deliverable

D1.4 Cities guide to Legal Compliance for Data-Driven Decision Making It. 3

Project Acronym:	DUET	
Project title:	Digital Urban European Twins	
Grant Agreement No.	870697	
Website:	www.digitalurbantwins.eu	
Version:	1.0	
Date:	31 May 2022	
Responsible Partner:	GSL	
Contributing Partners:	AIV, IMEC, ISP	
Reviewers:	Lieven Raes (AIV) Laura Temmerman (IMEC) Andrew Stott Geert Mareels (AIV)	
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	30.03.2022	Tomas Pavelka, Annabel Pemberton, Marco Lauro	GSL	Initial structure
0.2	17.05.2022	Tomas Pavelka, Annabel Pemberton, Marco Lauro	GSL	First draft
0.4	19.5.2022	Tomas Pavelka, Annabel Pemberton, Lieven Raes, Geert Mareels, Ilıa Christantoni, Dimitra Tsakanika, Jana Jerabkova (and others from pilot partners)	GSL, AIV PLZ DAEM	Q&A + interviews
0.5	23.5.2022	Tomas Pavelka, Annabel Pemberton, Marco Lauro	GSL	Second draft + consolidated guide version + annexes
0.6	30.05.2022	Jiri Bouchal, Lieven Raes, Andrew Stott	ISP, AIV	Review
1.0	31.5.2022	Tomas Pavelka, Annabel Pemberton, Marco Lauro	GSL	Final version

Table of contents

Executive Summary

This deliverable seeks to provide an updated and final version of the "Cities Guide to Legal Compliance for Data-Driven Decision Making" (the Cities Guide) tailored to assist in any data-based decision-making process: the **selection of datasets and models.**

The deliverable also assesses and extends the previous guide provided by deliverable D1.2 and suggestions provided by deliverable D1.3, and updates the background of applicable laws as set out in deliverable D1.1.

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.

This document comprises of feedback on the early Cities Guide section:

- **Summary of the general use of the early Cities Guide** by the pilot cities Athens (DAEM), Flanders (Digital Flanders, formerly AIV) and Pilsen (SITMP). Feedback was collected through a questionnaire and online interviews. The questions posed to the pilots can be found in **Annex 1.**

- **Granular feedback on the usability of the early Cities Guide sections** including personal/non-personal data, selection of data bases, risks of certain data types, documenting and communicating the use.

- **Expert opinions and views on the guide** including that of DUET project ethics manager Geert Mareels. This section addresses interesting examples where risks are highlighted as the data collection practices by cities infringe on privacy of individuals, in addition to further feedback on the Cities Guide.

- **Areas of improvement that were identified** relating to the Cities Guide content, structure and user experience for the future presentation of the content. In particular, while the content was up to date and relevant three main improvement areas were highlighted including providing information concerning variations in model/data licensing/data use; additional use cases including automatic number-plate recognition data (ANPR), floating car data (FCD), noise/pollution sensor data and wifi/mobile/other terminal equipment data; and the transfer of personal data to third party countries.

Which is followed by a final version of the Cities Guide:

- **The final version** provides a logically organised set of areas to check for legal necessities of a smart city data-driven decision making process and can be found in **Annex 2.** Following an introduction (including a necessary legal disclaimer), the Cities Guide elaborates on the following areas outlined in D1.2 and D1.3 with the addition of:

    - **Different use cases** including simulation models and the data protection implications.

    - **Risks in further data processing** including changes arising as a result of the Schrems II judgement, affecting the transfer of personal data to a third-party country.

- **Definitions provided in deliverable D1.2 are included in Annex 3**. This addition aspires to improve the readability and ease of use of the Cities Guide.

1. Introduction

This deliverable is the third out of the three deliverables in this working stream aiming to build up an easy to understand Cities Guide to legal compliance for data-driven decision making.

Significant effort and resources were dedicated to the first draft Cities Guide (D1.2) in order to build up a strong knowledge base and guidance for DUET in matters legal and ethical as soon as possible in the Project lifecycle. D1.3 further contributed to expand the guidance in the area of database selection, a topic specifically requested by DUET Consortium partners as being highly relevant for Smart City operations and policy making.

The present deliverable is focused on the feedback from two principal sources: (i) DUET pilot city partners and (ii) experts on the Cities Guide. Based on that feedback, this final deliverable provides evaluation of the Cities Guide, and suggests improvements to be included in the final re-stated guidance. Several final adjustments in the guide have also been provided in light of emerging regulation in the technology used in smart city projects and feedback from DUET Consortium partners.

This final iteration of a Cities Guide deliverable therefore focuses on:

- Analysis of feedback from DUET partners and experts regarding format, content, usefulness and impact of the guidance in It. 1 and 2
- Identifying any gaps in the existing guidance. Asses from the content perspective, as well as format and user experience
- Publishing a final guidance version
- Steps to raising awareness (conclusion)

Accordingly, Chapter 1 summarises feedback collected and assessed from Pilot city partners and DUET expert team via written Q&A and interviews over video conference.

Chapter 2 looks at the identified or perceived gaps and items for improvement/further research.

Conclusion provides an overall evaluation plus summary of the steps to be taken to increase and spread awareness of the DUET project's achievements in the legal and ethical fields

Annex 1 sets out the questionnaires submitted to Pilot partners for purposes of this deliverable.

Annex 2 contains the final version of the "Cities Guide to legal compliance for data-driven decision making." This version integrates the general guidance provided in D1.2 and the specific guide on database selection and documenting set out in D1.3 deliverables. In addition, it supplements / plugs some of the gaps identified throughout the review, even though the adjustments had to be only minor. From a legal point of view, the guide can be considered as up-to-date.

However, we would advise any reader to use the Cities Guide in conjunction with deliverables D1.2 and D1.3, as each of those provide a more detailed analysis of the topics at hand (D1.2 regarding general data management and personal data issues, and D1.3 topics of selection of databases for use in a project, and suggestions for the database use documentation system).

For completeness, Annex 3 transposes the set of definitions provided in the introductory chapters of deliverable D1.2. Re-stating these in this final deliverable may help avoiding the need for guide readers/users to circle back to earlier deliverables for the explanation of some of the basic concepts.

This document and the final guideline version is complementary to, and does not replace, existing or future applicable 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. This document should be read in conjunction with the definitions and broader guide to legal requirements as set out in deliverable D1.2 (the first iteration in this series of deliverables on “easy to use guides”) and D1.3.

* * *

Legal notice: Even though it is an ambition of this document to provide a useful guidance to any interested smart cities and other stakeholders out there, it is important to note that this document or deliverables D1.1, D1.2 and D1.3 do not, and are not intended to, constitute legal advice to DUET partner organisations or any third parties. 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 finalised 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 (including their **Data Protection Officers** (DPOs)) or an attorney qualified in the concerned jurisdictions to obtain advice with respect to any particular legal matter

2. Feedback from DUET partners

For purposes of this deliverable, we created questionnaires on the use of D1.2 and D1.3 deliverables within pilot organisations for the DUET project purposes. The questionnaire is reproduced as Annex 1.

Based on analysis of the responses obtained, we conducted a series of interviews with each Pilot city partner organisation. The responses / feedback are summarised in this chapter. As previously for purposes of drafting deliverable D1.2, questionnaires were fashioned as relatively open-ended, hence the responses differed in focus and granularity of information.

2.1 General use of the early “Cities Guide” versions

Athens (DAEM)

DAEM as a private organisation supporting the City of Athens services, indicated in response to the questionnaire that most of their database sources are open source. This indicates no change from their response in D1.2 that the city and citizens data is not owned or processed by DAEM. Therefore, while they consulted the Cities Guide, they generally found that it was not directly applicable for much of these sources.

The pilot partner emphasised that they had reviewed and commented on the Cities Guide deliverables in the course of their drafting phase, and that the Guide was subsequently consulted by the DAEM legal department. Generally, during the interview the pilot partner highlighted that their legal department is closely involved in these types of projects (including a DPO).

Specifically regarding the use cases, DAEM discussed that the Guide would be more directly relevant if they would decide to use parking data for DUET purposes, however the data was not selected for implementation at the end of the day for prioritisation reasons mainly. While only informational, elements that allowed to increase awareness by the team included information on ANPR and FCD data since the team had less prior experience in this area.

As regards data availability, the partner indicated that while data is getting increasingly open and available in Greece/Athens, some public data holders are yet to open up fully (e.g., Ministry of Transport, police). The pilot partner initially considered purchasing / collecting original data by installing new sensors (pollution, noise), but with help of other DUET consortium partners, DAEM was able to source data of similar / sufficient quality from existing open sources instead.

Flanders (Digital Flanders, formerly AIV)

Digital Flanders (**DV**) is part of the Flemish Government, in Belgium. DV (formerly AIV) is a public body tasked with support in the areas of digitization of data in e-government, GIS and public information.

According to the partner, the Cities Guide was useful as a reference framework and a way to check what kind of actions are needed as an answer to the questions raised. However, DV considered (in line with other partners) that given the lack of personal data used for DUET purposes, the guidance was of less direct relevance for the Flanders use cases. In any event, the theoretical chapters of the Cities Guide could benefit from providing real-life examples to improve the understandability of readers not familiar with smart city data and legal data aspects.

In their role of DUET coordinators, the DV team also went through the Cities Guide step by step in order to make sure that nothing was missed that needed a legal arrangement/action.

DV also engaged in procurement of 'model as a service' data, where the data model is not owned/operated by the DUET partner but stays with the service provided instead. This feedback is discussed in Chapter 3 as one of the perceived content gaps with regard to the Cities Guide. Furthermore, the Flanders team commented that while the use cases described by the Cities Guide were challenging, they could not be stretched to their full potential given the lack of use cases related to personal data. However, as the a pilot utilising simulation results, DV did see value in the questions relating to simulation results including *"Check for inherent risks of a simulation model (e.g., possibility of unrealistic results, gaps in coverage of influencing factors, drift caused by combining models, information on the correct interpretation of model outcomes)."*

Pilsen (SITMP)

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 (Pilsen).

Pilsen highlighted that as they only used anonymised and non-personal data, they did not consult D1.2 / D1.3. specifically for the purpose of data selection. They also did not need to use the checklists due to the lack of personal data in their data selection. However, they did note that the document is useful for smart city use cases utilising personal data. The partner considered the Cities Guide a concise and well-written document.

In addition, the Pilsen team found the Cities Guide beneficial for further work with data and the prevention of possible risks when working with data modelling results (traffic modelling, air, noise pollution). In particular, the guide was used by SITMP's (geographic) data administrator and applications operator. This role is responsible for storing (geographical) data and is in charge of managing the applications that the city uses to analyse data and create outputs for decision-making processes.

Similar to Athens, Pilsen highlighted that while only informational for their own pilot, Section 2.4.5 of deliverable D1.3 highlighted to them the risks of reidentification of anonymised data in combination with other datasets (for example, information on individuals being used in a newspaper articles about a significant event, which can identify individuals in a previously anonymised data set (e.g. car accidents)).

2.2. Personal data vs. non-personal data

The overarching comment regarding sections on legal requirements around personal data was stemming from the fact that the pilot cities didn't have to process a significant amount of personal data/sets for DUET purposes at the end of the day. Where personal data was involved (for example, processing and storing of information related to the identity of respondents to surveys or testers), the necessary measures and safeguards were already taken at the stage of the data collection (see generally WP9). Equally, there were no additional measures that needed to be taken vis-a-vis datasets acquired from third parties or from city organisations connected to DUET partners. Such data has been sufficiently anonymised and sanitised outside the Digital Twin.

In response to whether there was sufficient information concerning the sourcing of personal data, and what other information would they like to see in this section, the pilots highlighted the following points:

- In general there was sufficient information concerning personal data, however there could be other practical examples for better understanding.
- Pilsen among the partners considered Section 2.4.5 particularly helpful in building awareness of the re-identification risks that comes with anonymized data (for example, by combining the anonymised data points used in a database with information published in a newspaper article).
- While addressed in other deliverables, ethical dimensions could be directly included in D1.4.
- Reference to D1.1. and D1.2 for further details on personal data (such as definitions and basic concepts) is sufficient.

The Flanders team considered that the Cities Guides provides a framework for extending the digital Twin as a platform for personal services provided by using personal data as a source to get personal services and advice. These kinds of cases are not part of DUET today, but are part of other Digital Twin projects where Flanders is active (e.g. the H2020 Urbanage project).

2.3 Selection of databases / risk of certain data types

Step-by-step guidance was created for the selection/use of datasets and simulation models to provide a structured checklist that can be applied for each data selection action. Success in using this section was determined by identifying legal risks rising from specific use cases.

In general, while deliverable D1.3 was found useful for checking legal requirements when selecting a database for use in DUET, the following was noted:

- While the criteria set out in section 2.2 of the deliverable (Original/collected data vs. third-party data (sourced data)) was used as a framework, also other criteria were used due to the fact that datasets were open source and the use of the models was under an agreement (and some datasets were official models to be used in official procedures).
- Majority of dataset selection was done at earlier stages of the project, i.e., before Deliverable D1.3 was scheduled to be prepared.
- In general, however, these chapters of the Cities Guide were considered still broadly in line/up to date with the developing DUET ambitions/use cases. Certain use examples were, however, not used by some pilot cities (for example, the Telraam system data was not used by other partners than Flanders).
- For both Athens and Pilsen, as no personal data was utilised (for Pilsen) or openly accessible data from open portals such as the governmental data.gov.gr (for Athens), the step-by-step guide was not strictly indispensable.

- For Flanders, the step-by-step guide was used partially by DUET coordinators, also with regard to the design of the DUET Data Management Plan (WP8). No legal issues were identified through using the checklist.
- The guide could benefit from extending the considerations of data hygiene and quality.
- As regards licensing requirements, Athens highlighted that the section appeared less relevant directly to the pilot city team, because their legal department would typically deal with all issues around licensing.
- The following licence types were found to be used/present most often among data used for DUET purposes[1]:
    - Open data (typically CC 4.0 BY)
    - Creative Commons BY-NC-ND (examples: Czech Hydrometeorological Institute air pollution data)
    - Licence subject to contract for works (e.g., the 3D model in Pilsen - use allowed for DUET purpose as a non-commercial use).

2.4 Documenting and communicating the use / selection of data in DUET use cases

Pilots were asked about the use of the Cities Guide in how they were storing and then communicating the use and selection of data in the DUET use cases. Deliverable D1.3 contains a set of suggestions for improving the documentation and communication of use/selection of databases and models in the DUET data management lifecycle.

- The section was considered overall useful.
- All pilot partners highlighted that the guidance could be stretched to its full potential in cases where a smart city is collecting and processing personal data.
- Pilsen and Flanders highlighted that the team used the DUET Dataset Inventory document for database selection, which also contained helpful guidance / information, and was used in the design of the DUET Data Management plan.
- Positives of the guidance were that it highlights that some metadata couldn't be stored in existing metadata standards like DCAT or ISO19115-19 (Geospatial data), leading to some pilots becoming aware of this risk and finding appropriate solutions. However, in terms of usage, pilots which used the guide used it in conjunction with the DUET Deliverable D8.3 Data Management Plan.
- Flanders flagged that they log issues regarding legal aspects of datasets internally in coordination with the DPO. Because there were no issues found, no specific logs related to DUET have been made. This would have been only relevant for specific datasets created by DUET, which is not the case.

2.5 Experts' views

For purposes of this deliverable, we also held an interview with the DUET project ethics manager Geert Mareels. The expert expressed a (naturally) more conservative view on how (much) data should be managed by cities. Even though documents such as the Cities Guide may be helpful, there is some scepticism about their practical impact on daily practice of city officials (which was somewhat confirmed by the questionnaires/interviews).

The expert identified interesting examples where the data collection practices by cities clearly infringed on privacy of individuals (such as handling of data collection by private companies on behalf of the public authorities. In the

[1] The databases utilised in DUET are referenced in the following internal working document DUET Dataset Inventory.

given example, a municipality outsourced installation and operation of ANPR cameras to a private company, which subsequently lobbied the municipality not to install road bumps, so as to maximise profit from issuing speeding tickets). Such situations may lead to unclear division of legal and ethical responsibilities, blurring of public and private interest or outright misuse of public mandate. These examples are related, however, to situations / use cases outside of the DUET perimeter.

The expert has provided useful contacts for the GSL team to follow with regard to the final ethical deliverable D1.6.

Overall, however, the expert highlighted the usefulness of the Cities Guide on a broader level. There seems to be a lack of documents about legal necessities of data management that are sufficiently user-friendly and addressed to individuals at managerial positions within cities and other stakeholders. Other documents tend to be on the one hand too technical (in the sense of data management technology) or on the other hand too “legalese”. In that light, the DUET Cities Guide could become a fairly unique document available in the public domain.

3. Identified gaps - improvement points

3.1 Content

As regards content, the Cities Guide was viewed as up to date and still relevant. It was suggested that the theoretical foundations should reflect the new literature published between the date of issue of the earlier DUET deliverables and to-date.[2] Such a complex review of theoretical foundations of this stream of deliverables (set out in deliverable D1.1 - Legal landscape) is beyond the scope of this final evaluatory deliverable. However, we hope that there may be scope for such a more fundamental review in potential DUET project continuation, or other follow-up smart city projects.

We identified the following main three gaps / improvement points (the first two thanks to the interview with DV(AIV) DUET partner):

A) variations in model / database licensing, ownership, and use

- **Model-as-a-service.** In this case, the model (and data processing) runs at the provider's servers (premises). The customer only receives the processed results/inputs for further use. There is no licence agreement, only contract to provide the service.[3] In this case, the model and its methodology/workings may be unknown to the user (a black box problem). This setup may be complemented by an agreement to provide feedback by the user to the model provider.
- **Service provider's data vs. user's data.** Various ownership scenarios may occur: the model provider may use its own (or its partners' sourced) data to provide the agreed service, or it may use the recipient (user)-provided data, or a mix. Complex legal issues may arise with regard to who owns the results of the data processing, for example, or who is liable for any defects or inaccuracies of the results.
- **Marketplace for data**. Marketplaces for data may offer tremendous opportunities to stakeholders, but also may pose a range of interesting legal questions related to privacy, (cyber)security, data processing agreements (division of responsibilities between data controllers and data processors), etc.

While we consider the above gaps/topics as at least partially covered by the respective Cities Guide chapters (see mainly D. Ownership issues (decision, data, IP)), they may nevertheless be interesting areas for further deep diving and research. Like other areas of technological innovation in the past, it may be decades until legislation and judicial decisions establish definitive norms for these issues, and so it would be prudent for parties to make their intent explicit and to ensure that required responsibilities (for instance as "data controller" and "data processor" under the GDPR) are explicitly assigned to parties who can effectively carry them out.

B) additional use cases

[2] For example, article Data Governance and Regulation for Sustainable Smart Cities, available here: https://www.frontiersin.org/articles/10.3389/frsc.2021.763788/full, and others.

[3] This contractual setup was used by the Flanders pilot partner in an agreement with VITO, a Flemish Institute for Technological Research, for Spatial model Flanders 2050 including optimisation models for land use and infrastructures. See https://vito.be/en.

While the Cities Guide aimed at covering the most interesting (higher-risk) use cases deployed in DUET, there may be other use cases which deserve further consideration. We have been alerted to other projects that may have considered interesting use cases (beyond DUET use cases) as a subject for further research: the **Pilot case Gent**, for example, involved interesting questions around data subject definition (how to define categories such as students, residents, commuters and irregular visitors?) and involved more difficult cases of data sourcing and processing (mobile location data, wifi-sniffing data, etc.).[4]

The Cities Guide has been conceived as a fairly complex document, providing the legal guidance or at least pointing the direction even in use cases beyond DUET ambitions. As such, we trust that most other (even hypothetical) use cases should be manageable / reviewable through the lens of the present Cities Guide or at least the guide may be used as a starting point for a more thorough legal analysis. For example, chapter H. Risks in further data processing contains guidance on use cases such as the Telraam cameras (privacy by design practice example), highlights the sensor fusion risks or discuss legal aspects of machine-to-machine communication, which may cover potential other use cases (beyond DUET ambitions), and, importantly, cover use cases that may use or result in personal data processing, which is a high-risk activity by definition.

In addition, Chapter 2.4 of deliverable D1.3 (Risks of selected database / model types) contains a discussion of various data types, including automatic number-plate recognition data (ANPR), floating car data (FCD), noise/pollution sensor data, wifi/mobile/other terminal equipment data, which should cover a range of use cases for a smart city data-based policy and decision making. References are provided to projects in which there was scope (and need) to further deep dive these database types (e.g., the **H2020 PoliVisu project**).

C) transfer of data to third countries

We have identified a gap in the early Cities Guide related to the transfer of personal data to third (non-EU) countries. The discussion has revived again after the July 2020 Schrems II judgement by the Court of Justice of the European Union[5] invalidating again the privacy shield for EU-US data transfers, and the emergence of the "Privacy Shield 2.0" (officially known as Trans-Atlantic Data Privacy Framework) in March 2022.[6] Around the same time, the European Data Protection Board (EDPB) published new guidelines to develop a code of conduct as a tool for transfer of data to third countries.[7] We have complemented chapter H. Risks in further data processing of the Cities Guide with a subsection discussing this guidance.

3.2 Structure and user experience

In general, the pilot cities were happy with the structure and navigation through the early version of the Cities Guide. Suggestions which were made included:

[4] PoliVisu deliverable D4.7 available at https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ced672f9&appId=PPGMS

[5] https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf

[6] European Commission and United States Joint Statement on Trans-Atlantic Data Privacy Framework of 25 March 2022 https://ec.europa.eu/commission/presscorner/detail/es/ip_22_2087

[7] Guidelines 04/2021 on Codes of Conduct as tools for transfers Version 2.0 Adopted on 22 February 2022 available at < https://edpb.europa.eu/system/files/2022-03/edpb_guidelines_codes_conduct_transfers_after_public_consultation_en_1.pdf >

- To upload all deliverables on the DUET webpage separately so each city that aims to consult the guidance can navigate through url links;
- To import the checklist into an online survey format for ease of use.

4. Conclusion and steps to increased awareness

This document is the third and final 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 complement D1.2 and D1.3 by highlighting the useability of the guidance and present areas of strengths and areas for improvement.

It may appear from the feedback collected from DUET partners that the WP1 has “missed the mark” by focusing too much on personal data management and its legal necessities, whereas personal data turned out not to be the central issue for DUET. We believe rather that the privacy considerations and personal data protection should be the backbone of every project with extensive data management ambitions, even if the risk crystallises as rather low in the course of the project’s life cycle. As some “failed” smart city projects show only all too well (the Toronto Sidewalk project, for example[8]), respect for privacy and due consideration of ethical aspects relating to local stakeholders as well as wider population, must be both the starting and the end point of any successful smart city project management.

By treating extensively about personal data and the challenges to privacy posed by new technologies, we believe that the WP1 deliverables (i) contributed to overall compliance and awareness among the DUET consortium, and - as the Flanders team highlighted during our interview - (ii) provided a framework for extending the Digital Twin as a platform for personal services provided by using personal data as a source to get personal services and advice. These kinds of cases are not part of DUET today, but are part of other Digital Twin projects where for example the Flanders team is active (e.g. the **H2020 Urbanage project**).

As confirmed by the invested DUET partners, the Cities Guide also played a positive role in shaping the DUET’s data management plan and policy (WP8).

The Cities Guide also contributed - we believe - to the opening of a debate about many developing niches related to data management, including ePrivacy, (cyber)security, big data and HPC large-scale data processing, AI, and cloud computing. Cloud especially will get a further special focus in the upcoming **D1.7 deliverable - Recommendations for European Cloud Infrastructure**.

In addition to technical data management questions, the Cities Guide also attempted to explore more theoretical issues and concepts behind decision- and policy-making processes, and tried to tie these concepts together into (ideally) legally-compliant and ethically-aware process flows. We believe that these conceptualising efforts are relatively unique among similar documents available today.

In addition to the feedback collected from DUET partners, this deliverable also presented a final version of the guidance to be published for use by other smart cities. While the final version (Annex 2) integrates the separate guides provided by deliverables D1.2 and D1.3, we advise any reader to use the Cities Guide in conjunction with these deliverables because each provides a more detailed analysis of the topics at hand and their theoretical foundations.

[8] https://www.sidewalklabs.com/toronto, see also Deliverable D1.1. for an extensive discussion on this project.

Finally, the GSL team remains committed to increasing and spreading awareness of topics that have been analysed or in some cases merely hinted at during our work on the Cities Guide. More specifically, we aim to:

- Sync with DUET coordinators to implement user-experience suggestions from DUET pilot cities:
    - Upload all legal/ethical deliverables on the DUET web page separately so that a city that aims to consult the guidance can navigate through url links;
    - Import the checklist into an online survey format for ease of use;
- Begin work towards the final iteration of the ethical deliverable (D1.6 - Ethical Principles for using Data-Driven Decision in the Cloud), including to connect with contacts suggested by DUET partners in the course of interviews for the present deliverable;
- Coordinate with 21c Consultancy partner to identify opportunities for presenting / publishing results (blogs, conferences);
- Allocate resources to contribute to the upcoming Springer Computing publication about the Digital Urban and Local Twins;
- Coordinate with other DUET consortium partners with the view to (i) continue engagement with the DUET project, if the European Commission decides to prolong it, and (ii) identify new opportunities in research projects in the field.

Annex 1 - Questionnaire

Questionnaire for Feedback on D1.2 and D1.3 Cities Guide to Legal Compliance for Data-Driven Decision Making

Opening questions:

1. Does the document (D1.3) meet the following goals of providing legal assistance in the selection of datasets and models? If so, how?
2. How have you used the document for the selection of datasets and models?
3. When implementing DUET ""ambitions / use cases"", which use case you have found most challenging from the perspective of privacy or ethical considerations?
4. Please describe how you utilised the checklists that were available in this guide. Were the questions sufficient enough to check the dataset alone?
5. Please describe the steps you took as a result of a negative result from the self-check.
6. Which elements do you now have a better understanding of as a result of using the Guide?
7. Please suggest any improvements in the format of the guidance - e.g embedded links, structure.

Section specific questions - D1.3

2.1 Personal data vs. non-personal data 8

Is there sufficient information concerning the sourcing of personal data? What other information would you like to see in this section?

2.2 Original/collected data vs. third-party data (sourced data) 9

How are you currently navigating this section between other chapters referred to in this section? How could the navigation be improved?

2.3 Licence requirements 10

Please provide an overview of licence types that are most-often used by your organisation in relation to third-party datasets (top 3 most often used).

2.4 Risks of selected database / model types 12

Do you consider this chapter still broadly in line/up to date with the developing DUET ambitions/use cases?

2.4.1 Automated Number-Plate Recognition (ANPR) data 12

Does the suggestion to use fully anonymised data practically work for your usecase? Or are there use cases that require a different safeguard?

2.4.2 Crowd sourced traffic counts (e.g. Telraam data) 13

1. (*applicable to Flanders*) Does the data your use case process also fall outside of the Telraam project usage outlined? If so, how? Please describe if a different process is used for the same process which stores data locally.
2. (*applicable to Athens and Pilsen*) If you source crowd-sourced traffic counts, to your knowledge, does the source provider implement similar privacy-by-design features such as those described for Telraam?

2.6 Step-by-step guidance to selection of datasets / simulation models 16

1. Has your organisation used the step-by-step guidance in selected datasets?
2. Please state who in your organisation used the checklist.
3. Please describe how you utilised the checklist.
4. Please describe the steps you took as a result of a negative result from the guidance. Did you reach any 'dead-ends' as a result of the guidance?

3 Documenting and communicating the use / selection of data in DUET use cases 19

1. How well does this section assist you in improving your data management?
2. Please describe how you used this guide - e.g in conjunction with DUET Data Management Plan (D8.3)
3. Have you used / created any centralised system where your team logged instances of using or changing the datasets, as a result of using the guide? Please describe how you used the guide in conjunction with this process.

Section specific questions - D1.2

“Cities Guide to Legal Compliance for Data-Driven Decision Making” (first version) 35

1. If you used the Cities Guide, how did you view it?
2. Please provide an overview of how you used D1.2. Which role in your team used it?

A. Introductory remarks 35

1. Do you consider this section still broadly in line/up to date with the data usages for the developing DUET ambitions/use cases? If no, please describe what also should be addressed?
2. To your knowledge, how could this section be improved with other content?

Annex 2 - “Cities Guide to Legal Compliance for Data-Driven Decision Making” (final and consolidated version)

This is a final and consolidated version of the “Cities Guide to Legal Compliance for Data-Driven Decision Making.”

A typical data-driven decision or policy-making consists of several steps, which usually follow in a 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 finalising 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; (vii) database selection issues, etc.

* * *

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 from making 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 finalised 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[9]. 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**

- **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 (Annex 3 Definitions), and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;

[9] See Section 3.2.3.

        - Processing on a large scale of special categories of data (Annex 3 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 determine 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" (Annex 3 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* (Annex 3 Definitions), which may produce legal or similarly significant effects concerning data subjects (Annex 3 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. (See also 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)

-> see also process guidance in **G. Selection of datasets / simulation models**

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 (Annex 3 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 L. 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 licence, open data licence, 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 (Annex 3 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 licence, open data licence, 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 (Annex 3 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 right holder) 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 practise in your jurisdiction with negotiating licensing. 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 licensing, depending on the rules of the pilots' jurisdictions, consider specific clauses in the context of licensing 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)

-> see also process guidance in **G. Selection of datasets / simulation models**

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

-> see also process guidance in **G. Selection of datasets / simulation models**

Key aspects: typical legal defects: GDPR breach; ePrivacy rules breach (state-level differences); licence 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 licence (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)**.**
    - *Existing own / proprietary data* (Annex 3 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* (Annex 3 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 **K. 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. Selection of datasets / simulation models (process guidance)

The following principles constitute a good practice at the dataset selection stage in a Smart City project / use case design and implementation:

- **Set the purpose for your action**. This is important to know what legal rules and limitations may be applicable to your action. For example:
  a. I am sourcing a dataset to create a simulation model
  b. I am sourcing a simulation model to integrate it with the system (combining with other datasets / models) in order to enable a system functionality
  c. I am sourcing a dataset to create a derivative dataset (by changing the data or by combining multiple datasets)
  d. Each such action should have an identified (meta) purpose, such as "*I want to create a simulation model in order to achieve objective X*". These may be typically derived from user cases - epics.

- **Data availability** (licencing conditions, format compatibility, APIs). These conditions are important to ensure that you meet applicable IP laws requirements for sourcing and use of a dataset / model.
  a. If you need to negotiate individual access to a closed / proprietary database or simulation model, engage your legal department in the contractual negotiations.

   b. Check restrictions on use of open licenced databases / models, such as the attribution rule (CC BY ⓘ), or share-alike (CC SA⟲) condition. Some licences may prohibit you from creating derivative works (e.g., CC-ND ⊜). Consult with your legal department if you are unfamiliar with the terms
   c. Check for any licence conflicts or restrictions on combining multiple datasets. Consult with your legal department if complex questions arise

- **Dataset / simulation model quality.** These steps are important to mitigate risks stemming from lack of data's legal or factual quality.
   a. Self-check the following:
      i. 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?
      ii. Doesn't the data suffer from defects in important properties? (accurateness, timeliness, consistency, etc.)
      iii. Isn't the data manifestly unfit for the required purpose?
      iv. Is the data shared under reasonable licensing conditions?
      v. Are there apparent issues with correct anonymization, aggregation or pseudonymization of the data? (e.g., use the Telraam example as a good practice to achieve a privacy-by-design solution that helps to minimise the risks to citizen's privacy).
   b. If sourcing third-party data, ask the provider for:
      i. declaration of conformity with privacy legislation
      ii. legal basis and permitted use purposes of any acquired personal data
      iii. if the data is anonymised/pseudonymised, ask what measures/techniques have been used to achieve this.
   c. Anonymised data preference (this may depend on particular project priorities):
      i. To minimise legal risk, prefer fully anonymised data. If, exceptionally, pseudonymised data must be used in order to meet the specific needs of a user epic, these should be selected and ingested only with the prior explicit approval by the DPO/other dedicated data management officer.
      ii. Beware the issue of anonymisation of the original dataset. Check with the third party data provider whether they have fully anonymised the original dataset as well so that the anonymized / aggregated dataset provided to you can be treated as non-personal data. If they have not done this, the data is treated merely as pseudonymised, as opposed to anonymised. Where such guarantees cannot be provided by the third party provider, treat the dataset as personal data (principle of precaution).
   d. Check for inherent risks of a simulation model (e.g., possibility of unrealistic results, gaps in coverage of influencing factors, drift caused by combining models, information on the correct interpretation of model outcomes).

- **Transparency as a general risk mitigator.** A clear record of your action will help to mitigate and correctly attribute any residual risks in using a dataset / model.
   a. Note and report any issues with data quality, e.g. accuracy, non-up to date data, incorrect data
   b. Note and report any issues with the selected data model, e.g.

    i. Risks of calculation of unrealistic results
    ii. Risk that not all of the influencing factors are covered by the model
    iii. Risk of drift caused by combining multiple models
    iv. Lack of information about the correct interpretation of model outcomes.

  c. Report and disclose these risks vis-a-vis third parties or the general public, if they are the intended addressees of your actions (they are the addressees of your decisions based on the data / models, they are the next users of datasets/models created by your organisation, they are testers, etc.)

  d. If collecting or processing personal data, you may be required under the GDPR to issue a Data Protection Impact Assessment (DPIA) where your actions are "likely to result in a high risk" to privacy. In such cases, work closely with your DPO to take the necessary and timely action. The following list of typical "high risk" activities is not exhaustive and you should request further guidance from your DPO.

    i. use of new technologies (e..g, IoT applications)
    ii. systematic monitoring of a publicly accessible area on a large scale (including with CCTV cameras or sensors able to collect personal data)
    iii. Systematic and extensive evaluation of personal aspects based on automated processing, including profiling of individual persons,
    iv. processing on a large scale of special categories of data and data relating to criminal convictions and offences

- **Meet intra-organisational requirements (partner-level approval process)**. A general good practice would be to check with your DPO each selection / use of personal data (including, for example, pseudonymized ANPR data and similar data with higher risk profile (e.g., wifi / cellular mobile "sniffing" data, app collected data)

- **Meet other project-specific requirements** - follow the Data Management Plan

H. 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), transfer of data to third countries.

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 I. 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 J. 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 I. 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 (Annex 3 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 (Annex 3 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 (Annex 3 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* (Annex 3 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 Annex 3 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 licences as well.
    - Comply with licence 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

- *Transfer of data to third countries:*
  - GDPR restricts transfers of personal data outside the European Economic Area (EEA), or the protection of the GDPR, unless the rights of the individuals in respect of their personal data is protected in another way, or one of a limited number of exceptions applies.
  - In February 2022, the EDPB published new guidelines to develop a code of conduct as tool for transfer[10].
    - First, the code of conduct should include a description of all the transfers which should be included in the code taking into consideration aspects such as: i) nature of data transferred; ii) categories of data subjects; iii) countries. The code should also provide for the description of the data protection principles to be complied with under the code. Principles must include: a) transparency, b) fairness c) lawfulness, d) purpose limitation, e) data minimization f) accuracy, g) limited storage of data. Moreover, there should be the setting up of an appropriate governance through DPOs or other privacy staff in charge of compliance with data protection obligations resulting from the code as well as the identification of accountability principle measures taken under the code. Paramount for a code of conduct concerning the transfer of data in third countries is the implementation of a data protection audit or other internal mechanism for monitoring compliance with the code, independently from the oversight to be performed by the monitoring body as for any code of conduct. The creation of third-party beneficiary rights for data subjects to enforce the rules of the code as third-party beneficiaries (as well as the possibility to lodge a complaint before the competent SA and before EEA Courts). In addition, it should be implemented an appropriate complaint handling process for data protection rules

[10] Guidelines 04/2021 on Codes of Conduct as tools for transfers Version 2.0 Adopted on 22 February 2022 available at < https://edpb.europa.eu/system/files/2022-03/edpb_guidelines_codes_conduct_transfers_after_public_consultation_en_1.pdf >

infringements maintained by the monitoring body which if deemed appropriate may be complemented with an internal procedure to the code member for managing complaints.

I. 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 K. 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 not possible only based on the GDPR purpose compatibility test. End users' additional consent 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*).

J. 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 I. Purpose limitation).
    - *Privacy-by-default* means that the controller (Annex 3 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* (Annex 3 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**.*

K. 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' rights* 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* (Annex 3 Definitions) also when controller (Annex 3 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 (Annex 3 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 L. Collaboration and trust).
  - *Maintain documentation of data processing activities* (Annex 3 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 (Annex 3 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 (Annex 3 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).

L. Collaboration and trust

Key aspects: standardisation; standard licences; 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 licences.* Open Data legislation allows public authorities to share data subject to standard licences. 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", Annex 3 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*).

M. Documenting and communicating the use / selection of data (process guidance)

The following principles may be advisable to complement an organisations / project's data management plan:

- **Create a centralised logging and communication system**
  a. Define which partner should set up and oversee the system
  b. Define "responsible persons" and "users" at the level of individual person - user, at the level of city or partner organisation +consortium level

- **Define steps and events that should be logged in,** for example:
  a. When (on what date) the data sourcing approval processes were met
  b. Effective database / model sourcing date (date of concluding the licence agreement, date of first access, date of downloading, date of last access)
  c. All data sanitization measures applied to the dataset (anonymisation, pseudonymisation, aggregation, statistical evaluation, metadata generation, etc.)
  d. All other defined "uses" of the database (what is a data use? See Annex 3 for definitions), e.g., integrating the data into the DUET system, updating the data, using the data to make a decision. If necessary to decrease the burden on users, define a level of "significance" for database use. Only a significant use would trigger then need to create a log.
  e. AI- or automated system-related decisions (including the use of data to "train" an AI), use of the "Digital Twin Data Broker" to process and aggregate data
  f. Data storage-related decisions (e.g., decision to upload to a Cloud-based service)
  g. Access to / use of a database or model by third parties (e.g., external testers, final users in the course of data management lifecycle)

- **Define and make mandatory reporting of other events and errors**, for example:
  a. Data breaches, security incidences, or similar adverse events
  b. Accidental data re-identification

c. Reporting of data getting outdated, data getting beyond the permitted time for their storage
d. Reporting of significant change of circumstance that may make a dataset / model obsolete or inaccurate

- **Define when a sharing of data / model with other DUET partners should be recorded**, transfer of responsibilities under the logging system and the data management plan.

- **Reporting of gaps and data needs.** Define ways to log into the system a particular issue with the data or the lack thereof: missing dataset type, missing dataset set or subset, missing updates.

- **Trust-building and transparency.** The logging system should allow seamless communication about dataset / model selection and use processes with
  a. partner organisations
  b. users and testers
  c. the general public (dissemination)
  d. Stakeholders and users should be allowed and motivated to conduct prior consultations with respective partners regarding the data needs and gaps for user cases

- **FAIR data principles: ensure full machine-actionability of the logs** (findable, accessible, interoperable, reusable)

Annex 3 - Definitions

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[11].	**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

[11] Article 4(1) GDPR.

		merits[12].
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[13].	“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[14].
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 data which were initially personal data, but were later made anonymous[15].
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.[16] GDPR must be observed for the personal data part of the set[17]. **Example**: data related to the Internet of Things, where some of the data allow assumptions to be

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

[13] Article 9(1) GDPR.

[14] Article 4 GDPR.

[15] 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.

[16] 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.

[17] *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.

		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[18].	**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"[19]. 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[20]. **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.
Confidential data	Data subject to an obligation of professional secrecy regulated by Union or Member State law, including a statutory obligation of secrecy.[21]	
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[22].	**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[23].
Business confidential data	For DUET purposes, intermediate versions of DUET consortium project data and datasets are deemed	It is important to distinguish this concept from the concept of "data confidentiality and integrity" (see

[18] 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.

[19] 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."

[20] 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.

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

[22] 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.

[23] Ibid, recital 14.

	business confidential, irrespective of the licence that the consortium establishes for final datasets[24].	**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[25].	**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.[26] **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 personal data anonymous in such a manner that the data subject is not or no longer identifiable[27].	**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. 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[28]. Sufficiently robustly anonymized data are not personal data.

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

[25] 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).

[26] Recital 162 GDPR.

[27] 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.

[28] Recital 26 GDPR.

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[29].	**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.

Data categories according to origin and purpose

Table 2: Data according to origin and purpose

Category	Definition
Original data	Data collected / produced by a Smart City / partner organisation (e.g., during a dissemination action or a pilot activity).
Proprietary / 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 Smart City / project 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[30].
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 visualise 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 visualise the present.
Location data	Data indicating the geographic position of a person or the terminal equipment of a person (user)[31].
Modelling data	Modelling data contains all data related do models and interactions[32].
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

[29] Article 4(5) GDPR.

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

[31] 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.

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

	users achieve meaningful results[33]. 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[34].
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[35].
Metadata	Metadata is data that provides information about other data[36]. 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”[37]. 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”[38].

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[39].

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

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

[34] 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.

[35] 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.

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

[37] 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.

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

[39] Article 4(2) GDPR.

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 licence access).
Obtaining data from a third party data provider/vendor (non-open data)	Purchasing data or otherwise individually negotiating access to data (individual access licence).
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.
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[40].
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.

[40] Article 4(3) GDPR.

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[41].
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[42].
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[43].
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[44].
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 and the interface of the network[45]. 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[46].

[41] Article 25(1) GDPR.

[42] Article 25(2) GDPR.

[43] Article 4(6) GDPR.

[44] Article 4(12) GDPR.

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

[46] Article 4(7) GDPR.

Processor	Natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller[47].
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[48].
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[49].
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[50]. These services may be considered an "electronic communication service"[51] and thus be subject to ePrivacy laws.

[47] Article 4(8) GDPR.

[48] Recital 84 GDPR.

[49] Article 4(4) GDPR.

[50] 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.

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