Deliverable

D8.3 Data Management and Modeling Plan

Project Acronym:	DUET	
Project title:	Digital Urban European Twins	
Grant Agreement No.	870697	
Website:	www.digitalurbantwins.eu	
Version:	1.0	
Date:	31 May 2020	
Responsible Partner:	AIV	
Contributing Partners:	IMEC, KUL, ATC, DAEM, VCS, TNO, P4ALL, PLZ, ISP, GFOSS	
Reviewers:	Andrew Stott, Yannis Charalabidis, Michiel Van Peteghem, Pieter Morlion	
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	13.05.2020	Lieven Raes	AIV	Initial structure
0.2	26.05.2020	Yannis Charalabidis & Pieter Morlion	AEGEAN + MoreLion	Comments
0.3	29.05.2020	Kletia Notti & Valeria Comegna	Grimaldi	Chapter 2 - PSI update
1.0	03.06.2020	Lieven Raes	AIV	Final version

Table of Contents

List of Figures

List of Tables

Executive Summary

DUET aims to establish the use of big data and data visualisation as an integral part of policy making, particularly, but not limited to, the local government level and the mobility and transport policy domain. The project's relation with data is therefore essential and connatural to its experimental research objectives and activities.

Additionally, the consortium has adhered to the H2020 ORDP (Open Research Data Pilot) convention with the EC, which explicitly caters for the delivery of a DMP (Data Management Plan).

According to the DUET DoA (2019), data management planning, monitoring and reporting is part of WP8 - the Project risk and quality management work package - and foresees the delivery of four consecutive editions of the DMP at months 6, 12, 24 and 36.

This first edition, however, is not a mere collection of principles, as it sets the stage for the ongoing and next activities handling with data, before and even after the project is completed. As per the DoA description: "*DMP describes the data management lifecycle for all data sets that will be collected, processed or generated by the research project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved".*

We basically envisage three main data usage scenarios, which jointly compose DUET data management lifecycle:

- Original data produced by the DUET consortium and/or individual members of it (e.g. during a dissemination action or a pilot activity)
- Existing data already in possession of the DUET consortium and/or individual members of it prior to the project's initiation
- Existing data sourced/procured by the DUET consortium and/or individual members of it during the project's timeline

The structure of this document is as follows:

- **Section 1** presents DUET data management lifecycle and frames the DMP within the EU H2020 Guidelines and FAIR data handling principles, thus setting the stage for the following parts.
- **Section 2** is a brief overview of the legal framework, including the EU regulation on personal data protection (GDPR), the H2020 provisions for open access to research data, the specific provisions of the DUET Grant Agreement and Consortium Agreement and some special provisions for big data management.
- The core of the DMP is **Section 3**, in which the data usage scenarios are presented and the key issues to be examined in relation to each scenario are discussed. These issues include decisions on e.g. data anonymization, privacy and security protection measures, licensing etc.
- **Section 4** concludes the document by anticipating the expected contents of future editions of the DMP.

For completeness of information, the reader interested in getting to know how the DUET consortium plans to deal with data may also refer, in addition to this DMP, to the following, already or soon to be published, deliverables: D1.1 Legal Landscape and Requirements Plan [February 2020], as well as D1.2 Cities Guide to Legal Compliance for Data-Driven Decision [September 2020], D4.2 Duet data integration (November 2020).

1. Introduction

Visualisation and management of (big) data in a user friendly way for public administration bodies is one of the primary goals of the DUET project. The intention is to support integration of (big) data into policy and decision making processes. The project's relation with data is therefore essential and connatural to its experimental research objectives and activities. Additionally, the consortium adhered to the H2020 ORDP (Open Research Data Pilot) convention with the EC, which explicitly caters for the delivery of a DMP (Data Management Plan).
According to the DUET DoA (2017), data management planning, monitoring and reporting is part of WP2 - the Project and Quality Management work package - and foresees the delivery of four consecutive editions of the DMP at months 6, 12, 24 and 36. This first edition, however, is not a mere collection of principles, as it sets the stage for the ongoing and next activities handling with data, before and even after the project is completed.

1.1. The DUET Data Management Lifecycle

As per the DoA description, the DUET DMP "*describes the data management lifecycle for all data sets that will be collected, processed or generated by the research project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved".*
This paragraph summarizes the management procedures that will be followed when dealing with the data of relevance for the DUET project, and which will be further described in Section 3 of this document.

We envisage **three main data usage scenarios**:

a) Original data produced by the DUET consortium and/or individual members of it (e.g. during a dissemination action or a pilot activity);
b) Existing data already in possession of the DUET consortium and/or individual members of it prior to the project's initiation;
c) Existing data sourced/procured by the DUET consortium and/or individual members of it during the project's timeline.

For each of the above scenarios, the key issues to be examined are displayed by the following logic tree:

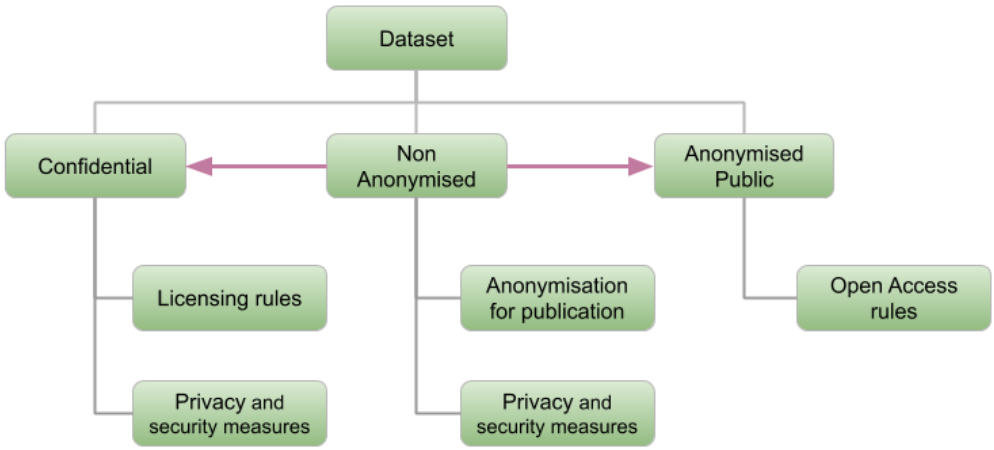

Figure 1: The Data Management Life Cycle

For each dataset (or even data point) handled in the project, the first level of control/decision making must deal with its **nature**, notably whether it has been (or will be) deemed Confidential, or Anonymised and Public (it cannot be that the two latter things diverge, apart from very special occasions, which are coped with in the third logical category displayed in the picture).

Depending on the assessment of nature, the resulting, mandatory **action lines** can then be summarized as follows:

- For any acknowledged **Confidential**[1] dataset (or data point), the Consortium and/or each Partner in charge of its handling shall control (if existing) or define (if not) the **Licensing rules** and the **Privacy and security measures** (to be) adopted in the process.
- For any acknowledged **Anonymised and Public** dataset (or data point), the only relevant discipline to be clarified is the set of **Open Access rules** that apply to the case. This set is little controversial for DUET, as the ODRP convention has been adopted, as specified above. Note that the use of open data across the DUET pilots falls in this category.
- Any dataset (or data point) that does not belong to any of the former two categories is subject to an additional level of action by the Consortium and/or Partner in charge, leading to its classification as either Confidential or Anonymised and Public. In that regard, the two, mutually exclusive action items belonging to this level are:
  - the **anonymisation for publication** action, leading to the migration to the second category of data, or
  - the adoption of appropriate **privacy and security measures** (very likely the same applied to the category of Confidential data) in case anonymisation is not carried out for whatever legitimate reason. Note that in this latter case, i.e. without anonymisation, **no licensing rules are applicable** (i.e. the DUET consortium rejects the commercialisation of the personal profiles of human beings as a non-ethical practice).

[1] Confidential in a meaning as defined by the GDPR (https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN)

1.2. Reference Framework and Perimeter of the DMP

The following picture – borrowed from the official EU H2020 information portal[2] - clearly identifies the positioning of the DMP in the context of projects that – like DUET – have voluntarily adhered to the Pilot on Open Research Data in Horizon 2020[3].

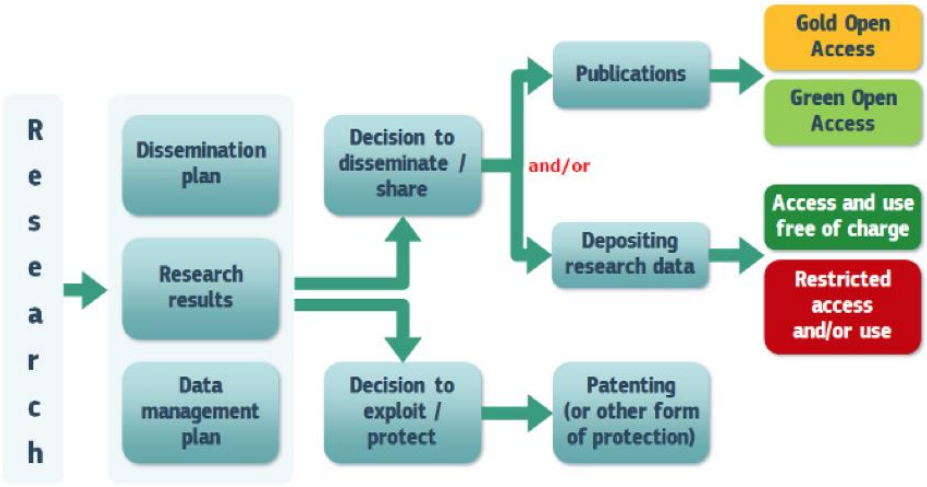

Figure 2: Open access to scientific publications and research data in the wider context of a project's dissemination and exploitation (source: European Commission, 2017)

As can be seen, a DMP holds the same status and relevance as the project's Impact realisation roadmap[4] and Business and Exploitation scenarios[5]. More specifically, in the former document, one should retrieve the full list of research data and publications that the project will deliver, use or reuse, as well as the indication of whether some data will be directly exploited by the Consortium, having been patented or protected in any other possible form. In the latter document, one should retrieve the Consortium's detailed provisions for all data and publications that can be shared with interested third parties, with or without the payment of a fee[6]. In particular, the following definitions – all taken from the aforementioned EU H2020 portal – shall apply to our discourse:

- **Access**: "*the right to read, download and print – but also the right to copy, distribute, search, link, crawl and mine*";
- **Research Data**: "*[any] information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form*";
- **Scientific Publications**: "*journal article[s], … monographs, books, conference proceedings, [and] grey literature (informally published written material not controlled by scientific publishers)*", such as reports, white papers, policy/position papers, etc.;

[2] http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

[3] Extended to all H2020 thematic areas since the start of the 2017 Work Programme, the focus of the Pilot is on "*encouraging good data management as an essential element of research best practice*". See European Commission (2017).

[4] Duet D7.1 Impact Realisation roadmap (M3)

[5] Duet D7.6 & D7.7 Business and Exploitation scenarios (M18, M30)

[6] This means that, at least in principle, some research data might also remain undisclosed, without undermining the Consortium's participation in the ORDP.

- **Open Access Mandate**: "*comprises 2 steps: depositing publications in repositories [and] providing open access to them*". Very importantly, these steps "*may or may not occur simultaneously*", depending on conditions that will be explained below:
    - **"Green" Open Access (aka Self-Archiving)**: it is granted when the final, peer-reviewed manuscript is deposited by its authors in a repository of their choice. Then open access must be ensured within at most 6 months (12 months for publications in the social sciences and humanities). Thus, open access may actually follow with some delay (due to the so-called "embargo period");
    - **"Gold" Open Access (aka Open Access Publishing)**: it is granted when the final, peer-reviewed manuscript is immediately available on the repository where it has been deposited by its authors (without any delay or "embargo period"). Researchers can also decide to publish their work in open access journals, or in hybrid journals that both sell subscriptions and offer the option of making individual articles openly accessible. In the latter case, the so-called "article processing charges" are eligible for reimbursement during the whole duration of the project (but not after the end of it).

In the DUET DoA (2019), the following provisions for Open Access were defined, which have become part of the **Grant Agreement (GA)** itself. The GA covers open access in Article 29.2 and 29.3 Open access to scientific publications and Open access to research data. Art 29.2 states: ensure open access to the deposited publication — via the repository — at the latest:
(i) on publication, if an electronic version is available for free via the publisher, or
(ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.
Also in part B, section 2.2.1.6, there is the following on open access: DUET will follow the Open Access mandate for its publications and will participate in the Open Research Data pilot, so publications must be published in Open Access (free online access). Following the list of deliverables, the consortium will determine the appropriate digital objects that will apply to the Data Management Plan. Each digital object, including associated metadata, will be deposited in a designated institutional repository, with an objective to increase visibility and make it accessible and preservable. This reflects the commitment of the Consortium within the framework of the Initiative for the Budapest Open Access, with the movement of open access to knowledge gained from joining the Berlin Declaration and Institutional Policy on Open Access. **Duet will work according to the "Green" Open Access**.

As far as patenting or other forms of protection of research results is concerned (the bottom part of Figure 2), the ground for this has been paved by the **DUET Consortium Agreement (2019)** in section 8. Following the DoA, which recognises that "If and when any joint Result is eligible for patent protection or any other intellectual property right protection, the joint owners shall, within a reasonable period following creation of any such joint Result, enter into good faith discussions in order to agree on an appropriate course of action for filing applications for patent protection or other protection".

Further steps towards a clarification of the licensing mechanisms will be taken in the context of the 2 foreseen editions of the Business and Exploitation Scenarios in the context of WP7 (deliverables D7.6 due at month 18 and D7.7 due at month 20.

As a general principle, the GA article 26.1 is faithfully adopted in the DUET Consortium Agreement (CA), according to which "*Results are owned by the Party that generates them*".
We take the above provisions also as a **guideline for the attribution of responsibilities of data management**, as far as DUET research results are concerned. Very shortly, we posit that **ownership goes hand in hand with the responsibility for data management**. The latter involves the same project partner(s) who generate new data, individually or jointly. 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** (as specified in Figure 1 above) and **undertake the consequent actions** as will be further described also in Section 3 below.

1.3. Alignment to the Principles of FAIR Data Handling

Generally speaking, a good DMP under H2020 should comply with the FAIR Data Handling Principles. FAIR stands for Findable, Accessible, Interoperable and Re-usable, as referred to a project's research outputs – notably those made available in digital form.
The FAIR principles, however, do not belong to H2020 or the EC but have emerged in January 2014, as the result of an informal working group convened by the Netherlands eScience Center and the Dutch Techcentre for the Life Sciences at the Lorentz Center in Leiden, The Netherlands[7].
Very pragmatically, the European Commission (2016) considers the FAIR principles fulfilled if a DMP includes the following information:

A. *"The handling of research data during and after the end of the project"*
B. *"What data will be collected, processed and/or generated"*
C. *"Which methodology and standards will be applied"*
D. *"Whether data will be shared/made open access", and*
E. *"How data will be curated and preserved (including after the end of the project)".*

In the case of DUET, the above information is provided in Section 3 of this document, which consists of five paragraphs, respectively:

3.1 Data summary (*typologies and contents of data collected and produced*)
3.2 Data collection (*which procedures for collecting which data*)
3.3 Data processing (*which procedures for processing which data*)
3.4 Data storage (*data preservation and archiving during and after the project*)
3.5 Data sharing (*including provisions for open access*)

Information about data curation and preservation can be found in section 3.9.
The following table matches the aforementioned EC requirements with the contents dealt with in Section 3 paragraphs.

[7] For more information, one can visit: https://www.force11.org/fairprinciples

This document's Section 3 TOC EC requirements	3.1 Data Summary	3.2 Data Collection	3.3 Data Processing	3.4 Data Storage	3.5 Data Sharing
A. "The handling of research data during and after the end of the project"					
B. "What data will be collected, processed and/or generated"					
C. "Which methodology and standards will be applied"					
D. "Whether data will be shared/made open access"					
E. "How data will be curated and preserved (including after the end of the project)"					

Table 1: Alignment principles of fair data handling overview

This Introduction has presented DUET data management lifecycle and frames the DMP within the EU H2020 Guidelines and FAIR data handling principles. The remaining structure of this document comes as follows:

- **Section 2** is a brief overview of the legal framework, including the EU regulation on personal data protection (GDPR), the H2020 provisions for open access to research data, the specific provisions of the DUET grant agreement and consortium agreement and some special provisions for big data.
- **Section 3** presents and discusses the data usage scenarios in the framework outlined in the above Table and examines the key issues in relation to each scenario. These issues include decisions on e.g. data anonymization, privacy and security protection measures, licensing etc.
- **Section 4** concludes the document by anticipating the expected contents of future editions of the DMP.
- In **Annex I** the interested reader can find a running list of utilized / relevant data sources, which will be further updated over the course of the project.

2. Legal framework

This section briefly overviews the key normative references making up the DMP external context. The next paragraphs respectively deal with:

1. The PSI Directive and its recent modifications and revisions proposals (dated April 2018);
2. The General Data Protection Regulation;
3. The terms of the H2020 Open Research Data Pilot (ORDP) the Duet consortium has adhered to;
4. The resulting, relevant provisions of both the Grant and the Consortium Agreements;
5. A general outline of Duet licensing policy.

2.1. The PSI Directive

The Directive 2003/98/EC on the re-use of Public Sector Information (PSI) entered into force on 31 December 2003. It was revised by the Directive 2013/37/EU, which entered into force on 17 July 2013. The consolidated text resulting from the merge of these two legislative documents is familiarly known as the PSI Directive, and can be consulted on the Eur-Lex website[8].

On 25 April 2018, the EC adopted a proposal for a revision of the PSI Directive, which was presented as part of a package of measures aiming to facilitate the creation of a common data space in the EU. This review also fulfils the revision obligation set out in Article 13 of the PSI Directive. The proposal has received a positive opinion from the Regulatory Scrutiny Board and is now being discussed with the European Parliament and the Council. It comes as the result of an extensive public consultation process, an evaluation of the current legislative text and an impact assessment study done by an independent contractor[9].

This approach gave rise to a new PSI 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[10].

It entered into force on 16 July 2019 and the Member States will have to transpose it by 16 July 2021.

The new Directive lays down an actual obligation upon Member States to make all existing documents held by 'public sector bodies' as well as public undertakings re-usable, unless access is restricted or excluded under national rules on access to documents or subject to the other exceptions. Public authorities can limit the making available of public data by imposing conditions in the standard licenses as regards the re-use by the licensee dealing with issues such as liability, the protection of personal data, the proper use of documents, guaranteeing non-alteration and the acknowledgement of source.

The new rules will stimulate the publishing of dynamic data and the uptake of Application Programme Interfaces (APIs). They will also limit the exceptions which currently allow public bodies to charge more than the marginal costs of dissemination for the re-use of their data. Moreover, the scope of the previous PSI

[8] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02003L0098-20130717

[9] Available online at: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51491

[10] https://eur-lex.europa.eu/eli/dir/2019/1024/oj

Directive is wider as it expands to:

- data held by public undertakings, which the undertakings make available for re-use. Charges for the re-use of such data can be above marginal costs for dissemination;
- research data resulting from public funding – Member States will be asked to develop policies for open access to publicly funded research data. New rules will also facilitate the re-usability of research data that is already contained in open repositories.
- strengthen the transparency requirements for public–private agreements involving public sector information, avoiding exclusive arrangements.

The current PSI Directive and its expected evolution is noteworthy and useful to define the context of the DUET project in general and of this DMP in particular. Thanks to the PSI Directive and its modifications and implementations[11], the goal of making government data and Information reusable has become shared at a broad European level. In addition, the awareness has been remarkably growing that as a general principle, the datasets where PSI is stored must be set free by default. However, fifteen years after the publication of the original PSI Directive, there are still barriers to overcome (better described in the aforementioned impact assessment study) that prevent the full reuse of government data and information, including data generated by the public utilities and transport sectors as well as the results from public funded R&D projects, two key areas of attention for Duet and this DMP.

More information can be found in the DUET Legal Landscape Analysis D1.1.

2.2. The EU Personal Data Protection Regulation (GDPR)

Regulation (EU) 2016/679 sets out the new General Data Protection Regulation (GDPR) framework in the EU, notably concerning the processing of personal data belonging to EU citizens by individuals, companies or public sector/non government organisations, irrespective of their localization. It is therefore a primary matter of concern for the DUET consortium.

The GDPR was adopted on 27 April 2016, and became enforceable on 25 May 2018, after a two-year transition period. By then, it will replace the current Data Protection Directive (95/46/EC) and its national implementations. Being a regulation, not a directive, GDPR does not require Member States to pass any enabling legislation and is directly binding and applicable.

The GDPR provisions do not apply to the processing of personal data of deceased persons or of legal entities. They do not apply either to data processed by an individual for purely personal reasons or activities carried out at home, provided there is no connection to a professional or commercial activity. When an individual uses personal data outside the personal sphere, for socio-cultural or financial activities, for example, then the data protection law has to be respected.

On the other hand, the legislative definition of personal data is quite broad, as it includes any information relating to an identifiable individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a home address, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address.

While the specific requirements of GDPR for privacy and security are separately dealt with in other DUET Deliverables - such as D1.1 Legal Landscape and Requirements Plan [February 2020], as well as D1.2 Cities Guide to Legal Compliance for Data-Driven Decision [September 2020], D1.5 & D1.6 Ethical Principles for using

[11] For instance, the INSPIRE Directive (2007/2/EC) builds mechanisms for data and corresponding Web services discoverability on top of the PSI Directive. See: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32007L0002&from=en

Data-Driven Decision in the cloud [February 2021 & November 2022], D1.3 & D1.4 Cities Guide to Legal Compliance for Data Driven Decision [July 2021 & May 2022] - it is worth noting here that the Duet consortium has formed a working group composed of the partner organisations Data Protection Officers (DPOs). The DPO function and role has been introduced by the GDPR and better defined by a set of EC guidelines, given on 13 December 2016 and revised on 5 April 2017[12].
The GDPR text is available on the Eur-Lex website[13].

2.3. Open Access in Horizon 2020

As partly anticipated in Section 1, the EC has launched in H2020 a flexible pilot for open access to research data (ORDP), aiming to improve and maximise access to and reuse of research data generated by funded R&D projects, while at the same time taking into account the need to balance openness with privacy and security concerns, protection of scientific information, commercialisation and IPR. This latter need is crystallised into an opt-out rule, according to which it is possible at any stage - before or after the GA signature - to withdraw from the pilot, but legitimate reasons must be given, such as IPR/privacy/data protection or national security concerns.
With the Work Programme 2017 the ORDP has been extended to cover all H2020 thematic areas by default. This has particularly generated the obligation for all consortia to deliver a Data Management Plan (DMP), in which they specify what data the project will generate, if it will not be freely disclosed for e.g. exploitation related purposes or how it will be made accessible for verification and reuse, and how it will be curated and preserved.
The ORDP applies primarily to the data needed to validate the results presented in scientific publications. Other data can however be provided by the beneficiaries of H2020 projects on a voluntary basis.
The costs associated with the Gold Open Access rule, as well as the creation of the DMP, can be claimed as eligible in any H2020 grant.
As already mentioned, the DUET consortium has adhered to the **Green Open Access** rule.

2.4. Grant Agreement and Consortium Agreement Provisions

The key GA and CA provisions worth mentioning in relation to our discourse on data management can be found in the Grant Agreement (art 24.1, 26.1 & 26.2, 29.1 - 29.3 & 39.2) and in the Consortium Agreement attachment 1. The articles itself can be found in Annex 3 and 4.

2.5. The DUET licensing policy

There is at the moment no single licensing policy within the DUET consortium, either for the software (DUET Digital City Twin solution) or their individual components, some of which belong to the Background as mentioned in the previous subparagraph. However, a few building blocks can already be identified, based on the discussion done in this document, the GA provisions quoted above as well as others not quoted yet, and the individual partners declarations in the CA. These provisions have been implicitly accepted by the DUET consortium members upon their signature of the aforementioned documents and are therefore totally

[12] See: http://ec.europa.eu/newsroom/document.cfm?doc_id=44100
[13] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679

enforceable. They are summarized in the table below, which also mentions in a footnote the examples currently dealt with in the DUET pilot cities.

Typology of data	Licensees	During the project period	After the project period	Legal references
Pre-existing (e.g. part of the Background knowledge of DUET, as listed in the CA Attachment 1)	Other members of the DUET consortium	Royalty free usage No right to sublicense	Under fair and reasonable conditions	GA Art. 25.2 GA Art. 25.3
	Any interested third party	As per the Background commercial licence	As per the Background commercial licence	CA Attachment 1
Sourced from third parties for the execution of project activities (e.g. portions of large datasets)	Other members of the DUET consortium	Royalty free usage No right to sublicense	Within the scope of the third party's license	General rules on IPR and license details
	Any interested third party	No right to sublicense	No right to sublicense	General rules on IPR and license details
Freely available in the state of art (e.g. Open Data)	Other members of the DUET consortium	Royalty free usage	Royalty free usage	Within the scope of the data owner's license
	Any interested third party	Royalty free usage	Royalty free usage	Within the scope of the data owner's license
Newly produced[14] during the project (i.e. part of the Foreground knowledge of DUET)	Other members of the DUET consortium	Royalty free usage No right to sublicense	Under fair and reasonable conditions	GA Art. 26.2
	Any interested third party	Open access at flexible conditions	Open access at flexible conditions	GA Art. 29.3

Table 2: DUET licensing policy overview

3. Duet data management plan

In this Section, the data usage scenarios presented in the Introduction are used as a basis for discussing the key issues to be examined in relation to each distinct paragraph of the DUET DMP. As a reminder, the three scenarios, which jointly compose the DUET data management lifecycle, are:

- Original data produced by the DUET consortium and/or individual members of it (e.g. during a dissemination action or a pilot activity);
- Existing data already in possession of the DUET consortium and/or individual members of it prior to the project's initiation;
- Existing data sourced/procured by the DUET consortium and/or individual members of it during the project's timeline.

[14] Special provisions must apply to the case, disciplined by the GA Art. 26.3, of a third party producing Foreground knowledge during the project activities.

On the other hand, the datasets handled within the three above scenarios can belong to either of these three categories:

- Confidential data (for business and/or privacy protection);
- Anonymised and Public data (as explained in the Introduction, these two aspects go hand in hand);
- Non anonymised data (the residual category).

The last category “Non anonymised data” is a temporary status as mentioned in the DUET Data Management Lifecycle. This temporary status must be as limited as possible since it creates a kind of a “grey zone” in terms of data management.

Specific attention goes to provisions for big datasets and provisions for models.

3.1. Data types

There are different possible data classifications. For the Duet Digital Twin, it is important to distinguish the data owner (government, private sector, citizen initiatives), the time dimension (real-time, historical/static) and openness of the data (anonymized/public, confidential).

	(near) Real-time data		Historical/static data	
Government data	(1) Anonymized/public	(4) Confidential	(7) Anonymized/public	(10) Confidential
Business data	(2) Anonymized/public	(5) Confidential	(8) Anonymized/public	(11) Confidential
Citizen data	(3) Anonymized/public	(6) Confidential	(9) Anonymized/public	(12) Confidential

Table 3: DUET Data types overview

Concerning data management, the essential difference is the ownership and the end-state of the data (anonymized/public) on the one hand and confidential on the other side. Both categories determine if the data can be used in a digital Twin and how the data needs to be processed.
Below you can find some examples of potential useful data to be used in a digital Twin:

- (1) Open IoT data from government managed sensors (air quality, water levels)
- (2) Test datasets (often with limited use) as social media data, mobile device data
- (3) Citizen science data as Luftdaten open air quality API
- (4) Closed government owned (IoT) sensor data like (raw) ANPR data
- (5) Live floating traffic data, social media data, mobile device data
- (6) -
- (7) Inspire datasets, city boundaries
- (8) Historical floating traffic data
- (9) Telraam citizen science traffic count data
- (10) Personal data like the register of persons
- (11) Telecom data based people flow data
- (12) –

3.2. Data summary

The following table summarizes the typologies and contents of data collected and produced. For each distinct category, a detailed list will be provided in the next edition of the DMP, due by month 12.

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary status)
Original data produced by the DUET consortium	Raw survey/interview/sensor data Evidence from project pilots Personal data of end users New contacts established	Summaries of surveys/interviews Data in reports of pilot activities End user data on public display Contact data within deliverables	Photos/videos shot during public events Audio recordings (e.g. Skype) Data in internal repositories
Existing data already in possession of the DUET consortium and/or partners	Data embedded in some of the Background solutions (see par. 2.4.2 above) Contact databases	Data embedded in some of the Background solutions (see par. 2.4.2 above) Website logs and similar metrics	N/A
Existing data sourced/procured by the DUET consortium and/or partners	Raw data in possession of the Cities or of any third party involved in the pilots	Free and open data (including from scientific and statistical publications)	N/A

Table 4: DUET Data usage scenarios - data summary overview

The main implications of the above table for the three usage scenarios are the following, in **decreasing order of urgency** for the related action lines as well as **increasing order of gravity** for the consequences of any inadvertent behaviour by the members of the consortium:

- The organisation of Living Lab experimentations (as foreseen by the project's work plan) implies that personal data handling of the end users acting as volunteers must be carefully considered, also for their ethical implications.
- For any photos/videos shot during public events, it is crucial to collect an informed consent note15 from all the participants, with an explicit disclaimer in case of intended publication of those personal images on e.g. newspapers, internet sites, or social media groups. This will bring the data back into the Confidential category, where it is legitimate to store and/or process it for legitimate reasons.
- For any audio recordings stored, e.g. in the project's official repository (currently Google Drive) or in individual partners' repositories, care must be taken of the risk of involuntary disclosure and/or the consequences of misuse for any unauthorized purpose. Same goes for the personal data of each partner in the consortium.
- Informed consent forms must be signed (also electronically) by all participants in surveys, interviews and/or pilot activities. As an alternative option, the partner in charge will commit to anonymisation and other related measures as a way to protect the identity of the respondents/pilot users.

[15] Informed means that consent cannot be given before reading and understanding the provisions of extant legislation on personal data protection and the way the partner asking for signature commits to enforcing / abiding by it.

- Informed consent forms are also required when using available contacts (be they preexisting to the project or created through it) to disseminate information via e.g. newsletters or dedicated emails. In this respect, the GDPR provisions are particularly binding and must be carefully considered, at least in any doubtful case.
- As a general rule, access conferred to Background knowledge on a royalty free basis during a project execution does not involve the right to sublicense. Therefore, attention must be paid by each partner of DUET to ensure the respect of licensing conditions at any time and by any member of the team.
- This also applies to any dataset sourced or procured from third parties during the DUET project's lifetime.

3.3. Data collection

The following table summarizes the procedures for collecting project related data. For each distinct case, some concrete examples will be provided in the next edition of the DMP, due by month 12.

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary status)
Original data produced by the DUET consortium	Surveys Interviews Pilot activities F2F / distant interaction	Newsletters Publications Personal Emails Open Access repositories	Events coverage - directly or via specialised agencies A/V conferencing systems Internal repositories
Existing data already in possession of the DUET consortium and/or partners	Seamless access and use during project execution	Seamless access and use during project execution	N/A
Existing data sourced/procured by the DUET consortium and/or partners	Licensed access and use during project execution	Free and open access and use during project execution	N/A

Table 5: DUET data usage scenarios - data collection overview

An implication of the above table that may not have been evident in the previous one, is that **every partner is responsible for the behaviour of all team members**, which may also include subcontracted organisations (e.g. specialised press agencies) or even volunteers. The latter circumstance does not exempt the delegate of a certain job in case of improper application of extant norms and rules.

All data will be collected in a digital form – therefore CSV, PDF, (Geo)JSON, XML, Shape, spreadsheets and textual documents will be the prevalent formats. In case of audio/video recordings and images, the most appropriate standards will be chosen and adopted (such as .gif, .jpg, .png, .mp3, .mp4, .mov and .flv). Ontologies will be created in Protégé file format (.pont and .pins) or .xml/.owl can also be used. Website pages can be created in .html and/or .xml formats.

Individually, each research output will be of manageable size to be easily transferred by email. However, it is important to note that email transfer can become a violation of confidentiality under certain circumstances.

3.4. Data processing

The following table summarizes the procedures for processing DUET related data that can be envisaged at this project's stage. As one can see, most of them make reference to the contents of paragraph 2.6 above. In this sense, more can probably be added to the cells of the table. Annex 1 to this document ("Running list of data sources") provides some meaningful case descriptions.

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary status)
Original data produced by the DUET consortium	Anonymisation Statistical evaluation Metadata generation Visualisation	Statistical evaluation Visualisation Analytics	Selection/destruction Blurring of identities
Existing data already in possession of the DUET consortium and/or partners	Anonymisation Statistical evaluation Metadata generation	Statistical evaluation Visualisation Analytics Publication as map services	N/A
Existing data sourced/procured by the DUET consortium and/or partners	Anonymisation Statistical evaluation Metadata generation	Statistical evaluation Visualisation Analytics Publication as map services	N/A

Table 6: DUET data usage scenarios - data processing overview

Apart from the specific software listed in paragraph 2.6 above, state of the art productivity tools will be used to process/visualize the data used or generated during the project. Typically, the partners are left free to adopt their preferred suite (such as Microsoft Office™ for PC or Mac, Apple's iWork™ and OpenOffice™ or equivalent). However, the following tools are the ones mainly used by the consortium[16]:

- Google's shared productivity tools (so-called G-Suite™) are used for the co-creation of outputs by multiple, not co-located authors.
- Jira™ is a proprietary issue tracking product developed by Atlassian that allows bug tracking and agile project management.
- Adobe Acrobat™ or equivalent software is used to visualise/create the PDF files.
- Protégé™ or equivalent software is used to generate the ontologies.
- Photoshop™ or equivalent software are used to manipulate images.
- State of the art browsers (such as Mozilla Firefox™, Google Chrome™, Apple Safari™ and Microsoft EDGE™) are used to navigate and modify the Internet pages, including the management and maintenance of social media groups.
- Teams™, Zoom™ or Skype™ (depending on the number of participants) are the selected tools for audio/video conferencing, which may also serve to manage public webinars.

[16] This list is non-exhaustive and can be updated at all times without prior advice.

- Tools like Google Forms™, and optionally SurveyMonkey™ and LimeSurvey™, are used for the administration of online surveys with remotely located participants.
- Dedicated Vimeo™ or YouTube™ channels can help broadcast the video clips produced by the consortium to a wider international audience, in addition to the project website.
- Mailchimp™ or equivalent software is helpful to create, distribute and administer project newsletters and the underlying mailing lists.

3.5. Data storage

The following table summarizes the procedures for storing project related data, during and after the DUET lifetime, and the most frequently used repositories. As for the previous paragraphs, we limit ourselves now to listing the headlines. Annex 1 to this document ("Running list of data sources") provides some contents for the pilot descriptions.

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary status)
Original data produced by the DUET consortium	Individual partner repositories Common project repository	Project website Open access repository	Individual partner repositories Common project repository
Existing data already in possession of the DUET consortium and/or partners	Specific software repositories	Digital Twin data broker Dashboard services	N/A
Existing data sourced/procured by the DUET consortium and/or partners	Individual partner repositories Third party repositories Cloud repositories	Digital Twin data broker Dashboard services Cloud repositories	N/A

Table 7: DUET data usage scenarios - data storage overview

Data storage and management of the use DUET datasources

The main principle is that DUET will use as much as possible data sources stored on external data platforms. Those can be hosted in the cloud or on-site by the data providers. Only in case, there is a strict need to host the data ourselves, DUET (pilots) will host the data itself.

Several technical components will process and align the data for use in the Digital Twin. In the current architecture, the access & authorisation services, the knowledge graph API, the data & model catalogue API and the context graph API will take up that role. Other data related messages will come from the IoT event services, the IoT time-series service and the geo-data service.

All those services will deliver information to a data broker combining message broker, message mapping & validation and message streaming functionalities.

The data itself will follow standardisation principles based on ISA² and W3C. The OGC/Inspire standards will be used for managing geospatial data. W3C Linked open data principles will be used for linking data semantically.

Data management of the DUET consortium project data

Google Drive™ is the selected tool as DUET's data and information repository. This includes both the project deliverables (including relevant references utilised for their production or generated from them as project publications, e.g. journal articles, conference papers, e-books, manuals, guidelines, policy briefs etc.) and any other related information, including relevant datasets. This implies that the privacy and security measures of Google Drive™ must be GDPR compliant. The verification of such circumstances is the responsibility of the coordinator.

Additionally, the coordinator will make sure that the official project repository periodically generates back-up files of all data, in case anything may get lost, corrupted or become unusable at a later stage (including after the project's end). The same responsibility goes to each partner for the local repositories utilised by them (in some cases, these are handled by large organisations such as Universities or Municipalities; in others, by SME or even personal servers or laptops).

Collectively, we expect the whole set of outputs to reach the size of 1 TB all along the project duration. This range will particularly depend on the number and size of the received datasets to be utilised for the execution of DUET pilots.

Whatever the license that the consortium establishes for final datasets, their intermediate versions will be deemed as **business confidential**, and restricted to circulating only within the consortium.

Finally and as stipulated in the DoA, each digital object identified as R&D result, including their associated metadata, will be stored in a dedicated open access repository. In addition to the above, other datasets may be stored on the following repositories:

- Cordis, through the EU Sygma portal
- The DUET website (with links on/to the Social Media groups)
- Individual Partner websites and the social media groups they are part of
- The portals of the academic publishers where scientific publications will be accepted
- Other official sources such as EUDAT[17]
- Consortium's and Partners' press agencies and blogs
- DUET official newsletters.

3.6. Data sharing

Last but not least, the following table summarizes the procedures for sharing DUET related data in a useful and legitimate manner. When sharing, it is of utmost importance to keep in mind, not only the prescriptions and recommendations of extant rules and norms (including this DMP), as far as confidentiality and personal data protection are concerned, but also the risk of voluntary or involuntary transfer of data from the inside to the outside of the European Economic Area (EEA).

In fact, while the GDPR applies also to the management of EU citizens personal data (for business or research purposes) outside the EU, not all the countries worldwide are subject to bilateral agreements with the EU as far as personal data protection is concerned. For instance, the US based organisations are bound by the so-called EU-U.S. Privacy Shield Framework, which concerns the collection, use, and retention of personal information transferred from the EEA to the US. This makes the transfer of data from the partners to any US based organisation relatively exempt from legal risks. This may not be the same in other countries worldwide, however, and the risk in question is less hypothetical than one may think, if we consider the case of personal

[17] https://eudat.eu/what-eudat

sharing of raw data with e.g. academic colleagues being abroad for the purpose of attending a conference. It is also for this reason that the sharing of non-anonymised data is discouraged for whatever reason, as shown in the table.

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary status)
Original data produced by the DUET consortium	Personal email communication Shared repositories	Project website Open access repository	N/A
Existing data already in possession of the DUET consortium and/or partners	Personal email communication Shared access to software repositories	Shared access to Playbox components Map services	N/A
Existing data sourced/procured by the DUET consortium and/or partners	Personal email communication Shared repositories	Shard access to Playbox components Map services	N/A

Table 8: DUET data usage scenarios - data sharing overview

As for the previous tables, Annex 1 to this document ("Running list of data sources") provides some relevant case descriptions.

3.7. Special provisions for big datasets

The DUET DoA describes how big data from different sources – notably available at city level, in relation to the nature of the identified project pilots, dealing with mobility, traffic flows, air quality – can distinctively contribute to a digital city twin and the processes of policy experimentation and policy making belonging to its Framework: design, implementation and (real time) evaluation of policy solutions.

Big data, as defined in ISO/IEC CD 2046, is data stored in "extensive datasets − primarily in the characteristics of volume, variety, velocity, and/or variability − that require a scalable architecture for efficient storage, manipulation, and analysis". This may include 'smart data', i.e. coming from sensors, social media, and other human related sources. "Smart Data" can be described as data or datasets that have been extracted from larger amounts of data (see big data)using algorithms according to certain structures and receive meaningful information. This data has already been collected, sorted and analysed and prepared for the end-user. The data must also be understandable by the user in order to achieve meaningful results. The semantics of the data, the data quality, the data security and the data protection are to be observed[18]. Smart data can be used both to gain new insights using raw data and to create models that can be used to analyze data." This obviously raises questions about ethics, privacy and legal requirements, which are explicitly and extensively dealt with in a dedicated WP (1), security that is part of WP (3) and will ultimately become part of a Digital Twins for Policy Making Starter Kit with book as Deliverables D7.8 (due July 2022).

In another WP (3), the DUET DoA extensively deals with the DUET DIGITAL CITY TWIN data infrastructure for cities that is now going to be developed within the project.

[18] Translated from: https://de.wikipedia.org/wiki/Smart_Data

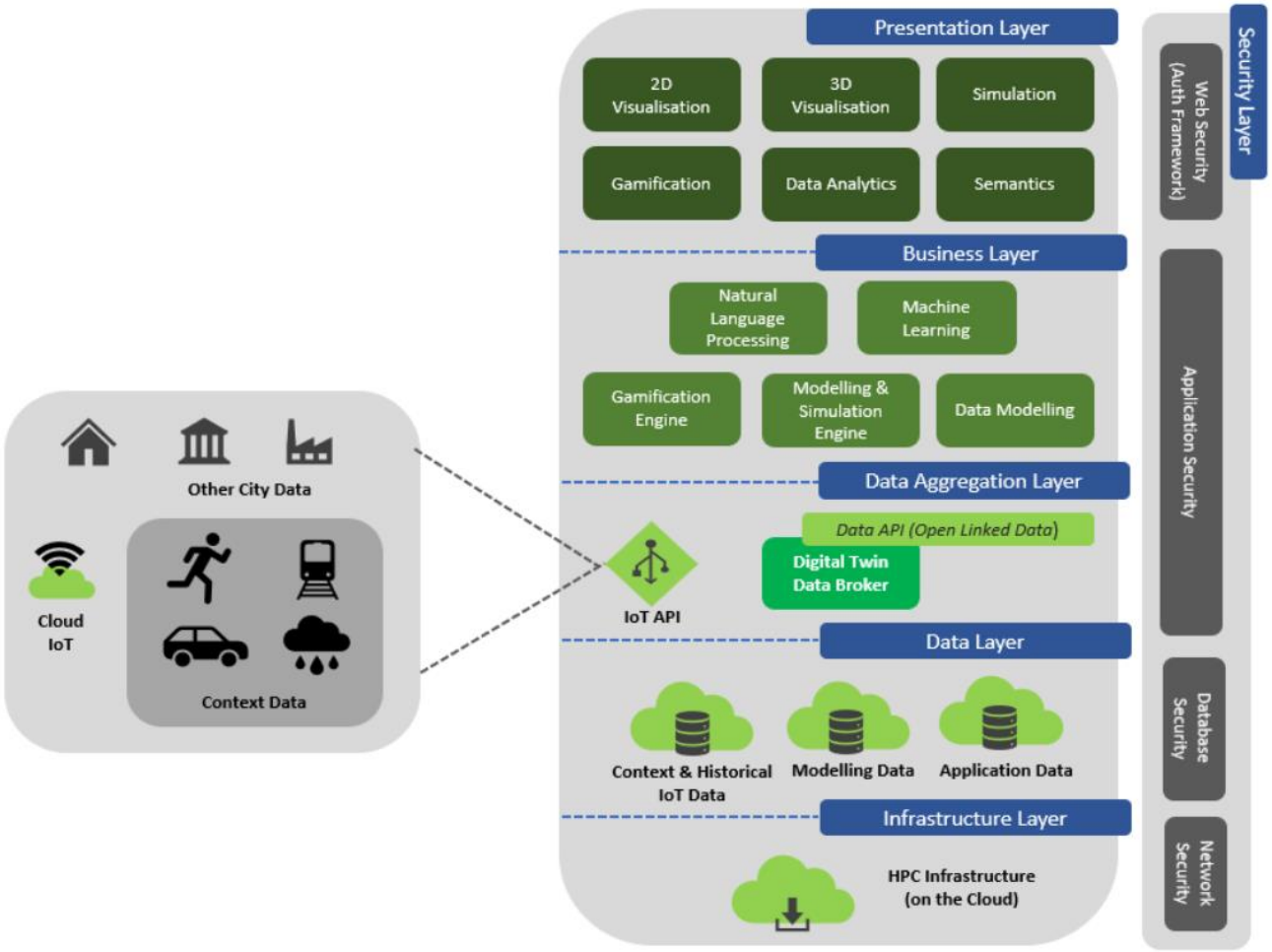

Figure 3: DUET layered architecture overview

Duet Digital Twin Layered architecture

The different layers of the DUET solution deal with big data. The most important layers concerning storing and processing granular data are the Infrastructure layer (1), the data layer (2) and the data aggregation layer (3). The Security layer (6) has also a big influence on the way data will be processed.

Layer overview

1. The **Infrastructure layer** corresponds to a cloud-based infrastructure with support for the High Performance Computing (HPC) workloads of the platform. The platform will be designed to be agnostic of the backed cloud provider.
2. The **Data Layer** refers to the repositories/databases of the platform where each database is deployed in a distributed mode that spans multiple nodes in the cloud cluster for enhanced scalability, availability and performance.
3. The **Digital Twin Data Broker** is in the heart of the Data Aggregation/Access layer that serves 2 purposes: it is responsible to not only aggregate data from the various external data sources but also to expose a unified data API for all open linked data in the repository. The IoT API component facilitates the ingestion of new data pipelines from sensors and other sources.
4. The **Business layer** corresponds to the processing components that implement the business requirements of the platform. All components in that layer will expose a REST API to be consumed by the visualization/UI components of:

5. The **Presentation layer**, relying thus on a service-oriented architecture. This layer provides the interfaces between the systems and the user.
6. The **Security layer** is applied to the whole architecture as a crosscutting concept and it affects different aspects of the architecture (web, database, network etc).

Tools

An overview of the tools that are foreseen to process the data can be found in Annex 2. The databroker, part of the data aggregation layer will play a central role in the data processing. The databroker can be seen as: "The data broker is the heart of the digital twin. It aggregates data for modelling, simulation and visualization purposes and exposes a secure API for modules to use as they see fit. As the name suggests, the data broker itself does not contain any specific modelling, simulation or visualization logic. An authentication and authorization framework enables secure access to the data and facilitates collaboration through data sharing."

Standards

DUET will make use of a number of (data) standards supported by International organisations.

- EU Inspire standards on geospatial data and metadata
- ESSC standards (European Sensor Systems Cluster)
- OGC standards on Geospatial data
- BDVA standards on big data (BDVA Reference model)
- OSLO Isa² standards
- LOD (Linked Open Data) standards
- W3C Web and metadata standards
- OASC smart city standards

More details about the precise standards and versions will be part of the diverse technical analysis work of DUET.

3.8. Special provisions for models

The data aggregation layer (3) accumulates data from different sources. Either by accessing them directly upon request or by storing it in a uniform way inside the digital twin broker. The latter can be done in several ways, for instance (a) by importing raw (historical) data, (b) by subscribing to external events and recording the data from these events. Event data can be stored as historical data or to update the state of the context. An overview of the existing models that DUET will initially integrate for environmental, health and mobility policy include the following. All the models rely heavily on the availability of GPU hardware and/or cloud infrastructure because of the compute intensive nature of the calculations.

- Air quality model visualisation (IMEC/TNO): The Air quality model will calculate the dispersion of air pollution caused by traffic for a grid of geospatial placed calculation points. The results will be converted to map images using interpolation or heatmap technology and placed on top of a map. Calculations are done for several compounds (NOx, PM10, PM2.5, EC, etc.), based on weather information (wind direction, wind speed) and spatial conditions (street canyons, shielding, etc.).
- Noise model visualisation (IMEC/TNO, P4All): The Noise model will calculate the propagation of sound caused by traffic for a grid of geospatial placed calculation points. Given the distribution of type of vehicles, noise levels per road link are calculated in an eight band frequency spectrum. The results will

be converted to map images using interpolation or heatmap technology and placed on top of a map. Calculations are done for noise levels in an eight band frequency spectrum.

- Traffic model visualization (IMEC,TNO, P4All): The Traffic model uses OD-matrices (origin, destination matrices) to assign executed trips to routes in the road network. The Traffic model assigns the amount of vehicles on roads using a 'route cost function' to determine the route having the lowest cost. Information from real-time traffic counts will be used to calibrate the traffic model predictions. Visualization on the map is done by drawing colored bands beside the road. The width of the band is modulated by the amount of vehicles, the color of the bands can have various meanings. By calculating the ratio between intensity and capacity, the color can represent the change for congestion

Depending on the pilot needs and the existing data new models will be created or existing models will work together towards integrated results and outcomes.

3.9. Data usage after DUET

The data usage after DUET will depend on different elements related to:

- **The dataset type**: Anonymized/public (open) data is rather easy to keep. For business data with limited use or confidentiality agreements will be made during the project. The nature of these agreements is subject to tailored data agreements and negotiations.
- **The maturity level of the implemented use cases**: The maturity level of the implemented use cases will differ and depend on the usability of the outcomes. It is expected that for some of the use cases, qualitative data and qualitative models will be available. In this case, the maturity of the case can lead to a useful smart city service with a broad user base and a willingness to pay. For other pilot cases, an innovative proof of concept will be realized with less need to keep the data and models after the end of the project.
- **The hosting, processing costs and access of the datasets**: The hosting costs is a third factor that determines the access and use of the data. Performant hosting and access to HPC power can be costly and only economically interesting when there are sufficient (different) customers who are prepared to pay a part of the bill. (Long Term) Agreements between various buyers and the hosting and data providers are needed.

4. Conclusions and future work

This document is the first of a series of four planned deliverables concerning the DUET Data Management Plan (DMP) in fulfillment of the requirements of the project's work plan. The main reason for planning four versions of the DMP (at months 6, 12, 24 and 36) and particularly two of them during the first project year, is evidently related to the need to hold on until the development as well as piloting activities of DUET gain further momentum, in order to:

- Secure the current, proposed structure of contents against any changes suggested by the gradual and incremental start up of the core project activities, and
- Colour the already existing contents with important add-ons based on the learning process that the DUET partners will activate throughout the project's lifetime, considering also that most of project work will be oriented to operationalizing the connection between data handling (including data integration, processing, analyzing and visualizing).

This edition of the DMP has, fulfilled the immediate goals of such a stepwise approach to data management, by:

- Presenting the legislative and regulatory framework, shaping the external context of this DMP in a relatively immutable manner, at least within the timeframe of the DUET project;
- Identifying the fundamental principles of FAIR data handling according to the EC requirements and that the DUET consortium and individual partners are bound to respect;
- Proposing a unitary description of the DUET data management lifecycle, a precise requirement of the DoA and that has been the leitmotif and conceptual architrave of the whole document;
- Summarizing the key aspects of data collection, processing, storage and sharing (the typical contents of a DMP) within the proposed lifecycle elements and particularly highlighting - first and foremost, to the attention of the partners - some key aspects of data management that go beyond the operational link with open access policy and interfere with privacy and security policies (an ethical topic falling under the competence of WP1) as well as with the way background knowledge and tools will be developed, deployed and customised to serve the needs of the city pilots (topics entirely covered by the WP3, 4 and 5).

As for now, it would be a great result if this first edition of the DUET DMP could enable all partners to understand the different action items that handling with data of different nature, origin and "size" imply for anyone wanting to stay in a "safe harbour" while actively contributing to the successful achievement of pilot and project outcomes.
Indeed, this document can be found lacking in a variety of respects, which will be gradually covered within the forthcoming editions of it. Some of the contents left unattended or only partly covered by this edition of the DMP include:

a) A timeline of partners contributions. Until now, the contents have been provided mainly by the responsible author (AIV) with the other partners acting as reviewers. In the future, and especially from now until month 12, further collaboration will be designed amongst the partners covering most of the aspects associated with small "signposts" here and there along the preceding text.

b) A clearer connection with data handling in other deliverables. In fact, due to the tight connection between project activities and data management, the reader interested in getting full information on how the DUET project deals with data should also refer, in addition to this DMP, to the following, already published, deliverables: D1.1 Legal Landscape and Requirements Plan [February 2020]. Upcoming delverables until month 12 related to data are: D1.2 Cities Guide to Legal Compliance for Data-Driven Decision [September 2020] and D4.2 Duet data integration (November 2020). It then makes sense to coordinate better and more explicitly the contents of these in order not to miss precious information while at the same time avoid duplications and inconsistencies in the framing and reporting of this crucial theme.

c) The Data management and modelling plan also takes the following topics into account: open standards, open data licensing, and consortium level policies. The latter aspect has been partly dealt with by reconstructing ex post some provisions of the GA and CA that are already binding for all partners. However, it is certainly worthwhile to make a more explicit and (to some extent) forward looking plan of e.g. what kind of licenses should be part of all the output categories making up the project results. It is also in that context that the issues of open standards and open data licenses (other than those belonging to the open access scheme) may be more extensively dealt with.

d) Another missing indication is certainly that of the partners responsible for the various steps of data management. At the moment, the crucial question of “who is in charge of” collecting, processing and storing data for each partner or deciding to limit or allow full access to some datasets, is subject of future decision making and will also depend on the maturity level of the pilot partners involved and strategic decisions when designing the DUET Digital Twin solution and platform. This question is not trivial (the answer equating the members of each partner team, or the heads of the teams, with the “people in charge” is by no means acceptable, giving too many things for granted, including the lack of hierarchies and other sorts of complexity within each partner’s organisation). In fact, some internal work is ongoing within the consortium at the level of creating a working group of the Data Protection Officers of each participant organisation. However there is more in between, and it will be the task of the next DMP edition to dig into the issue, thus contributing to the specialisation and clarification of the use cases now presented very superficially, in table form, within the preceding Section 3.

e) A final, indispensable aspect to be covered by a DMP is obviously the post-project scenario. What is the consortium’s and individual partners’ foresight of the management of pilot related datasets and more generally, of all the datasets created during the project’s lifetime that - for legitimate reasons, first and foremost exploitation related - are not subject to immediate publicity and may nonetheless require considerable attention and care to be maintained and preserved? Arguably the DUET work plan is at a too early stage to enable a firm definition of these aspects. However with the progress of activities (and time), we expect that the operational links created at pilot level between (big) data handling and predictive data driven models the behaviours of people involved in the Digital Twin experimentations, and the stages of the described Digital Twin evidence based policy cycle will start generating insights and enable the collection of evidence in view of the broader dissemination and exploitation phases of the project.

5. References

European Commission, Directorate-General for Research & Innovation (2016). *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, July 26th.

European Commission, Directorate-General for Research & Innovation (2017). *H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*, Version 3.2, March 21st.

DUET Consortium Agreement (2019). Innovation action No. 870697. Project Title: Digital Urban European Twin for smarter decision making.

DUET DoA (2019). Grant agreement No. 870697. Digital Urban European Twin for smarter decision making.

DUET Deliverable D1.1. - Legal Landscape and Requirements Plan, February 2020 (contractual delivery date).

6. Annex 1: Datasources & models overview

The following running list of datasources will be used during the project to keep track. Further updates will appear in the forthcoming editions of this Deliverable.

Nature of datasets	Used in the Flanders pilot	Used in Athens pilot	Used in Pilsen pilot
Confidential	New: ● - Existing: ● - Third party's: ● ANPR (anonymized/ pseudonymized) ● Kortrijk parking data ● Kortrijk empty and neglected business premises ● Kortrijk register of installed cameras ● Federal policy road accidents	New: ● DUET end-users data from Athens pilot Existing: ● - Third party's: ● -	New: ● - Existing: ● - Third party's: ● -
Anonymised and Public	New: ● - Existing: ● - Third party's: ● Telraam traffic counts ● Luftdaten air quality	New: ● - Existing: ● DAEM Controlled parking system data Third party's: ● -	New: ● - Existing: ● - Third party's: ● -
Non anonymised	New: ● - Existing: ● GRB 2D&3D ● Aerial photo ● Central address database ● Traffic loops ● Flanders road signs register Third party's: ● Cities 3D models ● KBO Cross-road DB of companies ● Antwerp bike sharing stations ● Antwerp car charging stations ● Antwerp hospitals ● Antwerp doctors	New: ● - Existing: ● Municipality of Athens environmental data Third party's: ● Atmospheric measurements on pollution ● Air-quality measurements ● Air quality Map (DRAXIS) ● Ministry of Environment and Energy - historic open data on air pollution ● Ministry of Environment and Energy - daily report on air pollution ● Athens air quality and air pollution data ● Urban transport - Locations of stations and timetables	New: ● - Existing: ● Traffic detectors ● Digital technical map ● Pilsen GIS data ● Pilsen Urban Plan (land use) ● public transport routes ● public transport stops ● - Third party's: ● Air quality sensor data ● Cadastre (national register) ● RUIAN (national register) ● -

A second table lists the used models.

Nature of datasets	Used in the Flanders pilot	Used in Athens pilot	Used in Pilsen pilot
Confidential	New: ● - Existing: ● Air quality models (used by Irceline / VMM) - OVL, Smogstop, Chimere ● Flemish multi modal traffic model ● Spatial model flanders 2050 including optimisation models for land use and infrastructures ● Land use optimisation model ● MatLabTrafficToolbox Third party's: ● -	New: ● - Existing: ● - Third party's: ● -	New: ● 3D model of the buildings ● Digital surface model ● Digital terrain model ● - Existing: ● 3D City centre and ZOO ● Traffic model ● - Third party's: ● 3D noise data ● Noise map (model) 2017 ● -

7. Annex 2: Duet Product overview

DUET Product overview:

Component	Product Description/Purpose	Partner	TRL
Presentation Layer			
2D and 3D Visualisation	The 2D and 3D visualisations provide the user interface to the Digital Twin for policy makers and their stakeholders. Components include: **VirtualcitySUITE**[19] comprising different software components: ● *virtualcityDATABASE* – A data storage and maintenance solution for CityGML-based 3D city models. ● *virtualcityWAREHOUSE* – A data distribution solution for extract/load/transformation processes that allows for extracting CityGML data from the virtualcityDATABASE into various industry GIS/CAD formats (e.g., ESRI mulitpatch, Sketchup, DWG, KML/COLLADA). ● *virtualcityPUBLISHER* – A publishing component for the web-based 3D visualization of CityGML- based 3D city models. All components use open standards for 3D visualisation and support advanced OGC CityGML models in a web browser environment. The suite allows combining WM(T)S services to link with standardised GIS layers to add additional data on the map. The solution has the potential as a map solution for an entire 2D and 3D digital twin and was already used for climate and energy visualisation in the city of Helsinki.	VCS	TRL9
	WebGLayer[20] is an open source JavaScript library focused on fast interactive visualization of big multidimensional spatial data through linked views. The technology provides interactive visual analytics that allows to identify patterns and risky areas from the large data (up to hundreds of thousands of features with several attributes) through heatmap, points or shapes on top of the map. WebGLayer is based on WebGL and uses GPU on the client device for high-speed rendering and filtering. The WebGLayer module provides components for advanced visualisations in the form of multiple linked views, filters through interactive graphs, parallel coordinates relationship analysis, map-screen extent filters, and area selection.	P4All	TRL9
Simulation	Simulations are predictions of policy impacts provided from the modelling and simulation engine in the Business Layer. In DUET the Air and Noise Pollution calculations will be driven by Traffic Volume on the road network. The Traffic Volumes are calculated using OD-matrices (Origin-Destination) describing 'Trips' going from A to B. With help of a global cost function, routes with lowest cost will be selected. Realtime data will be used to calibrate the OD-matrices, resulting in better prediction of the Traffic Volumes in the (near) future. For Air and Noise Pollution, emissions are calculated based on the Traffic Volume. Other (real time) data such as wind speed & wind direction will also be input for these environmental models. For an average city size 500.000 till 1.000.000 calculation points will be placed on the map for the area of interest. With DUET these calculations will benefit from the HPC cloud, because results will change based on the real time data feed. The results will be interpolated in real time producing bitmap (.png) tiles for display on 2D and 3D interfaces.	TNO	TRL7

[19] https://www.virtualcitysystems.de/en/
[20] http://webglayer.org/

Data Analytics	DUETs Big Data Analytics system will combine and analyze data from a variety of sources (databases, log files, web services, data streams, IoT networks). Some of the sources need to be processed in real-time, whereas for some other an ETL process is required. This process will bring all data into a common destination system and in a format suitable for further analysis. Big Data will be stored in special-purpose database systems that are able to store large amounts of data and can quickly and efficiently execute queries against them. Both proprietary (Redshift, Snowflake, BigQuery, Azure) and open source (Presto) solutions are available. For time-series data specialized tools (Prometheus) can be used in order to process the data and generate graphs, tables and alerts. Visualization tools, like Grafana, allow the creation of advanced and beautiful dashboards for analyzing and monitoring time-series data. In some domains the results of the analysis and selected visualizations need to be make publicly available. A policy maker would need to publish analysis results, tables and graphs in order to inform the public. In this case specialized development and custom tools are necessary.	ATC	TRL7
Semantics	DUET will leverage the OSLO[21] interoperability framework responsible to initiate and govern data standards applicable to the complete Flemish Government. OSLO provides in the tools, processes, guidelines and means to create and publish data standards. The OSLO standards are based on a strict interpretation of the ISA, Inspire and W3C/OGC standards All this happens in full transparency, accessible to all potential stakeholders. To further enhance the interoperability OSLO is almost entirely based on a Tim Berners Lee 5 stars Linked open Data approach. This approach is essential to ensure interoperability between cross border Digital Twins.	AIV	TRL8
Gamification	Gamification is the process of improving systems and people's experience using lessons, techniques and elements taken from games. Done well it increases engagement, motivation, retention and activity. In DUET gamification techniques will be explored like interactive voting and storytelling with the pilots to see if they can bolster participation of stakeholders in the policy design process. Selected approaches will be coded into the digital twin interface to be integrated in the 3D and 2D interfaces.	ATC	TRL3
Business Layer			
Natural Language Processing	A language detection module will be configured for a number of supported languages and a topics generator (based on LDA algorithm or its variants) that will take the number of topics as a parameter. More complicated machine learning techniques can be applied based on the output of these modules.	ATC	TRL7
Machine Learning	The social media content will be grouped into sentiment categories and the policy makers can focus on certain factors (named entities) and assess their importance. For each detected entity, the positive and negative references will be available. The implementation will be based on a classifier that will use a feature vector to categorize the text. If sufficient training data is available, deep learning techniques could be also applied (instead of the SVM approach).	ATC	TRL7
Gamification Engine	Coding based on co-creation eg. voting mechanisms and storytelling with data. The storytelling component will guide users step by step through a data visualisation to get a better insight. A voting mechanism will allow users to compare different policy measures, options and strategies.	ATC	TRL7
Modelling & Simulation Engine	In order to connect (real-time) data to models and analytics a messaging framework (TNO IMB) is used to interconnect them on a 'publish/subscribe' basis. (Real-time) data is published to the framework using gateway modules. The gateway modules are able to connect to various data sources (IoT, CoT, etc.) and publish their data to the model connection framework (publishing). The models use the framework to acquire relevant data elements (subscribe) which trigger (re)calculation of the results	KUL	TRL4

[21] https://joinup.ec.europa.eu/solution/oslo-open-standards-local-administrations-flanders

Data Modelling	Initial modelling components include: ● **Traffic Modeller**[22] is an open source server based solution for city and regional traffic modeling that allows city and transport planners to visualise traffic flow in near real time based on several parameters. The tool combining IT and GIS expertise can be fully implemented in a server environment with an application programming interface (API) for mobile and web applications. This creates an opportunity for a city or a region to test various traffic scenarios (e.g. road or lane closures, traffic accidents, planned roads) within seconds and without a need to install and learn how to use desktop traffic modelling software. ● **Urban Strategy Tile generator:**Environmental calculations mostly generate there results on top of the map. The Urban Strategy Tile generator is capable of (re)generating tiles at requested zoom levels for changing data. When environmental calculations are driven by real-time data, the changing results should be reflected on top of the map when available. The generator makes use of a multi-threaded, multi-core environment being able to process different data-layer in parallel. Results are standard .png images for easy inclusion in web based front ends. ● **MatlabTrafficToolbox**[23]This Matlab package provides access to open source library of codes to run and visualise dynamic traffic models. More specifically it presents tutorials and insights into the Link Transmission Model (LTM) a state-of-the-art dynamic network loading procedure and related dynamic equilibrium route assignment procedures.	P4All TNO KUL	TRL5 TRL7 TRL4
Data Aggregation/Access Layer			
Digital Twin Data Broker	The data broker is the heart of the digital twin. It aggregates data for modelling, simulation and visualization purposes and exposes a secure API for modules to use as they see fit. As the name suggests, the data broker itself does not contain any specific modelling, simulation or visualization logic. An authentication and authorization framework enables secure access to the data and facilitates collaboration through data sharing.	IMEC	TRL5
Data API	The data API facilitates and regulates access to the different data sources. It's main purpose is to provide a unified way of working with the different kinds of data. It shields the data from unauthorized access and presents the users with efficient ways of finding and retrieving the data.	IMEC	TRL5
IoT API	***The IoT API is very important in ensuring that the system built is a Digital Twin and reflects the real-world, and is not just a process of digitisation***. 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. The DUET architecture relies on the ability of these IoT stacks to deliver sensor data in an open format. DUET uses the Fiware open source platform and extended concepts that are found in the Synchronicity reference architecture because they use open standards and because they embrace the concept of linked open data in order to deliver true open smart city platforms. The city of things competence center of IMEC has realized several IoT solutions for smart cities and digital twins. DUET will leverage some of the existing IoT solutions and new ones as well to collect smart city data. The IoT stacks will be designed to deliver the necessary data to DUET's data aggregator. **Senslog**[24] is a web-based sensor data management system suitable for static in-situ and mobile devices with live tracking ability. It enables reception of sensor data directly from sensor devices or indirectly from any front-end elements, storing sensor data in the SensLog data model implemented in RDBMS, pre-processing of data for easy querying and analysing, and publishing data through the system of web-services to other front-end	IMEC P4All/ UWB	TRL9 TRL9

[22] https://github.com/kolovsky/spark-traffic-modeler/wiki
[23] www.itscrealab.be
[24] http://www.senslog.org/

	elements or other applications. SensLog provides a system of web-services with the JSON encoding or provides standardised services using core methods of the OGC SOS version 1.0.0. The latest version of the REST API is following the CRUD schema. SensLog is available as a cloud solution, written in JAVA 8 with utilisation of the Spring framework.		
Data Layer			
Context & Historical IoT Data	Sensor data typically comes in two forms. The historical data set is a large volume of data typically indexed according to time and geographicall dimensions. The historical data is mainly used to train digital twin models and visualize the past. The context data contains values as currently measured by the different devices. The context data is used as input for simulations and to visualize the present.	IMEC	TRL9
Modelling Data	The modelling data contains all data related to models and interactions (incl the (1) Structured data; (2) Time series data; (3) Geospatial data; (4) Media, Image, Video and Audio data; (5) Text data, including Natural Language Processing data and Genomics representations; and (6) Graph data, Network/Web data and Metadata.). A number of existing data models will be integrated into the Digital Twin, these are described below this table. New data models will also be created based upon city/policy need.	TNO	TRL7
Application Data	Application data is needed to regulate access to and management of the digital twin data. This includes user data, user rights management, model management data, data catalog management (metadata catalogue), subscriptions to data sources, etc. This repository contains data related to the user management needs as well as data related to the role-based access control mechanisms of the platform. Moreover, it enables integration layer to store data related to the synchronization and integration needs of the various workflows of the platform.	ATC	TRL7
Infrastructure Layer			
HPC	The Flemish Supercomputer Centre (VSC) is a virtual centre making supercomputer infrastructure available for both the academic and industrial world. The KUL has premium access and reduced fees for the usage of HPC within European projects. VSC also has support in setting up and converting software for HPC usage.	KUL	TRL9
Cloud	The Duet Digital Twin solutions will be cloud provider independent and agnostic to ensure re-usability but will be based on the Duet principles regarding ethics, security, functional and non-functional requirements (digital Twin Blueprint). One of the important needs will be the needless cooperation between the cloud-based solution and the use of HPC modelling.	IMEC	TRL4
Security Layer			
Web	The security layer is a collection of mechanisms that will enforce privacy-preserving security policies on all the platform layers. A major part of it will be the Access Control and Identity Management mechanism to retain authorized access across all components. Moreover, the use of AEGIS' Advanced Visualisation Toolkit will offer analysis and situational awareness to the end users of the platform, facilitating data process auditing and investigation of security alerts and actions that might result to data breaches or other security incidents.	AEG	TRL6

8. Annex 3: Legal data related aspects of the Grant Agreement

24.1 Agreement on background

The beneficiaries must identify and agree (in writing) on the background for the action ('agreement on background').

'Background' means any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights — that:

(a) is held by the beneficiaries before they acceded to the Agreement, and

(b) is needed to implement the action or exploit the results.

26.1 Ownership by the beneficiary that generates the results

Results are owned by the beneficiary that generates them.

'Results' means any (tangible or intangible) output of the action such as data, knowledge or information — whatever its form or nature, whether it can be protected or not — that is generated in the action, as well as any rights attached to it, including intellectual property rights.

26.2 Joint ownership by several beneficiaries

Two or more beneficiaries own results jointly if:

(a) they have jointly generated them 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.

29.1 Obligation to disseminate results

Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — 'disseminate' its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium).

29.2 Open access to scientific publications

Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.

29.3 Open access to research data

Regarding the digital research data generated in the action ('data'), the beneficiaries must:

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

(ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan');

(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

(...)

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

39.2 Processing of personal data by the beneficiaries

The beneficiaries must process personal data under the Agreement in compliance with applicable EU and national law on data protection (including authorisations or notification requirements).

The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement.

9. Annex 4: Legal data related aspects of the Consortium Agreement

Attachment 1: Background included

According to the Grant Agreement (Article 24) Background is defined as "data, know-how or information (…) that is needed to implement the action or exploit the results". Because of this need, Access Rights have to be granted in principle, but Parties must identify and agree amongst them on the Background for the project. This is the purpose of this attachment[25]. (...)

1. VLAAMSE GEWEST

As to VLAAMSE GEWEST, it is agreed between the Parties that, to the best of their knowledge no data, know-how or information of VLAAMSE GEWEST shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

2. INTERUNIVERSITAIR MICRO- ELECTRONICA CENTRUM

As to INTERUNIVERSITAIR MICRO- ELECTRONICA CENTRUM, it is agreed between the Parties that, to the best of their knowledge, the following background is hereby identified and agreed upon for the Project. Specific limitations and/or conditions, shall be as mentioned hereunder:

[25] Note: the partners not listed in this summary of Attachment 1 have declared not to have any background knowledge to share. While this can be true for software, models and algorithms, a comparable assessment is missing for the datasets that esp. the Cities intend to put at consortium's disposal for the execution of project pilots. Reasonably, that information was not available by the time the CA was prepared and signed. More details will be made available within the next editions of this DMP.

Describe Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for Exploitation (Article 25.3 Grant Agreement)
All methodologies, architecture, processes, technologies and their respective optimizations and implementations as they are used in the current set up of imec's Digital Twin. Specific for the IT-related pre-existing know-how, all architecture and it's implementations present in the current Digital Twin platform including all of their modules and sub-systems as well as all individual solutions and extensions of that platform are considered to be background IP. In the case where existing imec software modules are used, such modules will also be considered as background assets.	No access rights are needed for implementation.	No access rights are needed for exploitation.
Obelisk is a time series storage platform, designed for high throughput with the following properties: ● Storage of values (such as sensor measurements, event data, ...), support for various datatypes ● Handles high throughput, hundreds of thousands of requests per second ● Scales horizontally across all layers ● Ingrained support for user authentication and data separation ● HTTP REST API, allowing for easy integration with client libraries, applications, ...	Imec will grant access rights on Obelisk to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.
DYAMAND AGENT is an interoperability software platform that allows the automatic discovery of and communication with diverse sensors, actuators and user interfaces. This platform allows an easy integration of new IoT prototypes and the combination with off-the-shelf (e.g., EnOcean, Z-Wave, Zigbee, Amazon Alexa, Philips Hue, social robots) and experimental technologies from diverse vendors. The DYAMAND platform also offers remote management tools for upscaling prototype testing from a lab environment to remote test environments such as field trials in urban areas	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.
DYAMNAND CONTEXT BROKER Is an NGSI (V2/LD) compliant context broker that is a central component in the architecture and guarantees scalable processing of IoT data streams from and to devices and dispatching data to subscribers and time series storage.	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.
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. The DUET architecture relies on the ability of these IoT stacks to deliver sensor data in an open format. The city of things competence center of IMEC has realized several IoT solutions for smart cities and digital twins. These IoT stacks will be used in this project to deliver the necessary data and are to be seen as background IP.	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.

The Digital Twin Data broker: The digital Twin data broker aggregates data for modelling, simulation and visualization purposes and exposes a secure API for modules to use as they see fit. It does not contain any specific modelling, simulation or visualization logic. An authentication and authorization framework enables secure access to the data and facilitates collaboration through data sharing.	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.
The data API facilitates and regulates access to different data sources. It's main purpose is to provide a unified way of working with different kinds of data. It shields the data from unauthorized access and presents the users with efficient ways of finding and retrieving the data.	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.
Automated deployment process: Automatic deployment of Digital Twin components is done on Azure through Azure Devops multi-stage pipelines. The actual environment the components run in is a Kubernetes cluster, an open-source system for deploying, scaling and management of containerized applications.	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.
Onboarding process for shapefile and xml data: Connectors exist to parse and format data sources in Shapefile or XML format. Specific mappings to Digital Twin Entity models should be provided but the connectors can be scaled to handle any size of XML or Shapefile source.	Imec will grant access rights to some of the Partners by request for implementation of the Project.	No access rights are needed for exploitation.

This represents the status at the time of signature of this Consortium Agreement.

3. KATHOLIEKE UNIVERSITEIT LEUVEN

As to KATHOLIEKE UNIVERSITEIT LEUVEN, it is agreed between the Parties that, to the best of their knowledge, the following background is hereby identified and agreed upon for the Project. Specific limitations and/or conditions, shall be as mentioned hereunder:

Describe Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for Exploitation (Article 25.3 Grant Agreement)
Traffic flow model in Matlab code: Link Transmission Model as available and described at https://www.mech.kuleuven.be/en/cib/traffic/downloads	The traffic flow model code is available as open-source for non-commercial use under a GNU General Public License For commercial use, license conditions will need to be negotiated with KUL-LRD	The traffic flow model code is available as open-source for non-commercial use under a GNU General Public License For commercial use, license conditions will need to be negotiated with KUL-LRD
Traffic flow model in C++: Link Transmission Model as described in https://doi.org/10.1016/j.procs.2019.04.056	The traffic flow model code is available as open-source for non-commercial use under a GNU General Public License For commercial use, license conditions will need to be negotiated with KUL-LRD	The traffic flow model code is available as open-source for non-commercial use under a GNU General Public License For commercial use, license conditions will need to be negotiated with KUL-LRD

This represents the status at the time of signature of this Consortium Agreement.

4. ATHENS TECHNOLOGY CENTER ANONYMI BIOMICHANIKI EMPORIKI KAI TECHNIKI ETAIREIA EFARMOGON YPSILIS TECHNOLOGIAS

As to ATHENS TECHNOLOGY CENTER ANONYMI BIOMICHANIKI EMPORIKI KAI TECHNIKI ETAIREIA EFARMOGON YPSILIS TECHNOLOGIAS, it is agreed between the Parties that, to the best of their knowledge, the following background is hereby identified and agreed upon for the Project. Specific limitations and/or conditions, shall be as mentioned hereunder:

Describe Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for Exploitation (Article 25.3 Grant Agreement)
All methodologies, algorithms, processes, technologies and their respective optimizations and implementations as they are used in the way of working at ATC. Specific for the IT-related pre-existing know-how, all algorithms and their implementations present in the current software products of ATC including all of their modules and sub-systems as well as all individual customer solutions and extensions of that products are considered to be excluded pre-existing know-how. In the case where existing ATC workflow software modules are used, such models will also be considered as background assets	n/a	n/a

This represents the status at the time of signature of this Consortium Agreement.

5. 21C CONSULTANCY LIMITED

As to 21C CONSULTANCY LIMITED, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information of 21C CONSULTANCY LIMITED shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

6. AEGIS IT RESEARCH LTD

As to AEGIS IT RESEARCH LTD, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information of EGIS IT RESEARCH LTD shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

7. OPEN & AGILE SMART CITIES

As to OPEN & AGILE SMART CITIES, it is agreed between the Parties that, to the best of their knowledge, the following background is hereby identified and agreed upon for the Project. Specific limitations and/or conditions, shall be as mentioned hereunder:

Describe Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for Exploitation (Article 25.3 Grant Agreement)
OASC Minimal Interoperability Mechanisms[26]: Minimal Interoperability Mechanisms (MIMs) are universal tools for achieving interoperability of data, systems, and services between cities and suppliers around the world. As they are based on an inclusive list of baselines and references, MIMs take into account the different backgrounds of cities and communities and allow cities to achieve interoperability based on a minimal common ground. Implementation can be different, as long as crucial interoperability points in any given technical architecture use the same interoperability mechanisms.	N/A	N/A

This represents the status at the time of signature of this Consortium Agreement.

[26] https://oascities.org/wp-content/uploads/2019/06/OASC-MIMs.pdf

8. GRIMALDI STUDIO LEGALE SPRL

As to GRIMALDI STUDIO LEGALE SPRL, it is agreed between the Parties that, to the best of their knowledge,no data, know-how or information of GRIMALDI STUDIO LEGALE SPRL shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

9. DIMOS ATHINAION EPICHEIRISI MICHANOGRAFISIS

As to DIMOS ATHINAION EPICHEIRISI MICHANOGRAFISIS, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information of DIMOS ATHINAION EPICHEIRISI MICHANOGRAFISIS shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

10. virtualcitySYSTEMS GmbH

As to virtualcitySYSTEMS GmbH, it is agreed between the Parties that, to the best of their knowledge, the following background is hereby identified and agreed upon for the Project. Specific limitations and/or conditions, shall be as mentioned hereunder:

Describe Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for Exploitation (Article 25.3 Grant Agreement)
virtualcitySUITE: The virtualcitySUITE is a bundle of software tools and services for realizing a 3D spatial data infrastructure for 3D city models. It consists of different software modules and add-ons for the data management, distribution and web-based visualization of massive 3D GIS data such as OGC CityGML models, mesh models, point clouds, oblique imagery or legacy 2D map data based on open formats, services and APIs. 3D web maps produced with the virtualcitySUITE provide map tools (widgets) such as 3D measurement, viewshed and shadow analysis that can be directly used within this project. Moreover, there is a well-defined API and plugin mechanism that allows for programmatically adding new functionalities required for the DUET project. Our virtualcityPLANNER add-on allows for easily integrating additional 3D models such as architectural designs or drawings as well as georeferenced images into the 3D city model in the browser, and to interactively work with the existing urban objects. Planning scenarios can be easily shared with other users.	N/A	Based on license (depending on required modules, add-ons and plugins).
Urban Simulation Framework: This R&D framework is a bundle of software tools and services for the integration of simulation solutions into the virtualcitySUITE.	N/A	N/A

This represents the status at the time of signature of this Consortium Agreement.

11. NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO

As to NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

12. PLAN4ALL ZS

As to PLAN4ALL ZS, it is agreed between the Parties that, to the best of their knowledge,

no data, know-how or information of PLAN4ALL ZS shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

13. SPRAVA INFORMACNICH TECHNOLOGII MESTA PLZNE, PRISPEVKOVA ORGANIZACE

As to SPRAVA INFORMACNICH TECHNOLOGII MESTA PLZNE, PRISPEVKOVA ORGANIZACE, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

14. IS-practice

As to IS-practice, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information of IS-practice shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.

15. ETAIREIA ELEYTHEROY LOGISMIKOY LOGISMIKOY ANOIKTOY KODIKA

As to ETAIREIA ELEYTHEROY LOGISMIKOY LOGISMIKOY ANOIKTOY KODIKA, it is agreed between the Parties that, to the best of their knowledge, no data, know-how or information of ETAIREIA ELEYTHEROY LOGISMIKOY LOGISMIKOY ANOIKTOY KODIKA shall be Needed by another Party for implementation of the Project (Article 25.2 Grant Agreement) or Exploitation of that other Party's Results (Article 25.3 Grant Agreement).

This represents the status at the time of signature of this Consortium Agreement.