Deliverable

D1.7 Recommendations for European Cloud Infrastructure

Project Acronym:	DUET	
Project title:	Digital Urban European Twins	
Grant Agreement No.	870697	
Website:	www.digitalurbantwins.eu	
Version:	1.0	
Date:	30 November 2022	
Responsible Partner:	GSL	
Contributing Partners:	AIV, IMEC	
Reviewers:	Gert Vervaet (AIV), Dimitra Tsakanika (DAEM), Tomáš Krblich (PLZ), Andrew Stott, Yannis Charalabidis	
Dissemination Level:	Public	X
	Confidential – only consortium members and European Commission	

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

Revision History

Revision	Date	Author	Organization	Description
0.1	5.9.2022	Tomas Pavelka, Annabel Pemberton, Marco Lauro	GSL	Initial structure
0.2	6.11.2022	Lefever, Stefan Michiels, Philippe	Imec	Imec learnings
0.3	9.11.2022	Lieven Raes	AIV/DV	Initial structure
0.4	15.11.2022	Tomas Pavelka, Annabel Pemberton, Simona Uhrinová, Lieven Raes, Gert Vervaet, Leonidas Kallipolitis, Stefan Lefever, Karel Jedlicka, Jiri Bouchal, Tomas Mildorf, Koen Triangle	GSL, IMEC, DV, P4All, AIV	First draft
0.5	18.11.2022	Lieven Raes	AIV/DV	Review
0.6	24.11.2022	Dimitra Tsakanika	DAEM	Review
0.7	24.11.2022	Andrew Stott		Review
0.8	28.11.2022	Tomas Pavelka, Annabel Pemberton, Simona Uhrinová	GSL	Second draft
1.0	30.11.2022	Lieven Raes, Tomas Pavelka, Annabel Pemberton, Simona Uhrinová	AIV/DV, GSL	Final version

Table of Contents

Executive Summary

Smart cities should ensure that the processing of data in the cloud, whether through a cloud platform or within a data space, meets both common industry standards and conforms to the applicable law. Failure in either area could result in penalties based on security issues and reputational damage, in addition to unauthorised access and harm to individuals.

This deliverable seeks to provide an outline of the technical and legal recommendations for projects operating in the European Cloud Infrastructure and Data Spaces, tailored to assist Smart Cities when processing data. In particular, the document **addresses the emerging shift from cloud infrastructures, focused on the surrounding environment, to data spaces, focused on the output and contained data**. A shift strongly supported based on the DUET experiences and learnings.

The following therefore addresses:

- Lessons learned and recommendations to provide Policy-Ready-Data as a Service - an insight into the requirements to transform raw data into smart data.
- Acknowledgment of previous work - acknowledgement of the previous DUET deliverables and frameworks that are influential to this document.
- Utilising cloud computing - legal requirements and recommended practice - suggestions of a framework to follow when processing personal and non-personal data in the cloud, including the location of data, usage of cloud add on services and the contractual provisions in agreements between a Smart City and the cloud provider. This section also addresses the legal considerations of working with data spaces.
- Risks, benefits and opportunities for cities in utilising the cloud infrastructure - an extension on previous guidelines and suggestions concerning the usage of data spaces and simulation models by Smart Cities.
- Conclusion and steps to increased awareness - technical and legal conclusions including recommendations for future development.

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

2. Introduction - Lessons learned and recommendations to provide Policy-Ready-Data as a Service (PRD-a-a-S)

2.1. Technical considerations

Digital twins can largely facilitate data-driven decision taking in a public and societal context. However, this does not come for free. During the DUET project, we have seen that multiple partners are contributing data and intelligence on data (producing itself new data) to this process in a chained way. This process connects to front-end tools that assist by giving understandable insights into the data, as is illustrated in the DUET T-cell architecture (Figure 1) depicted below.

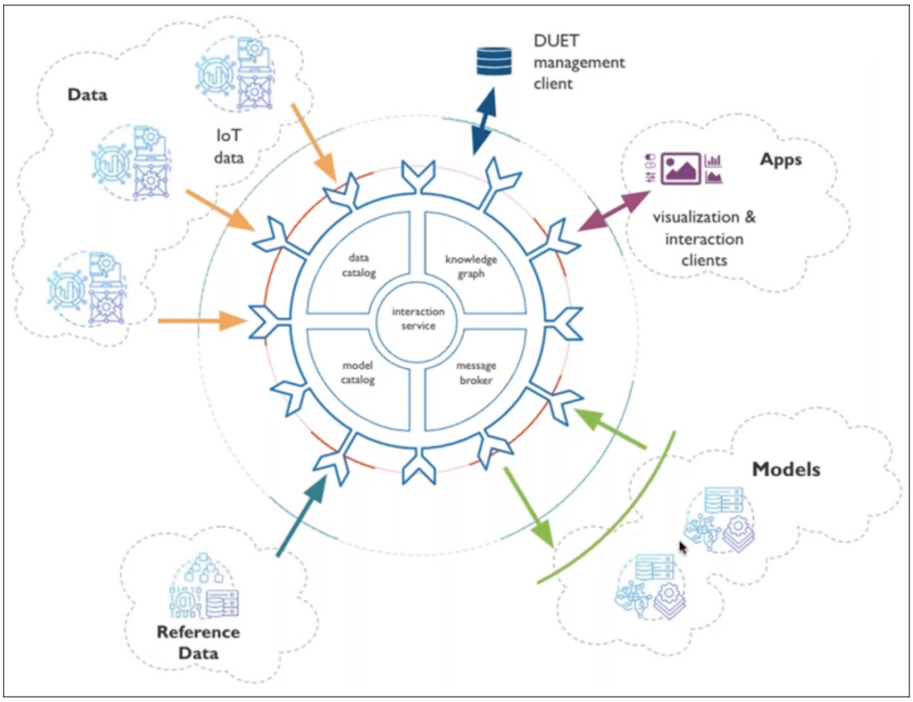

Figure 1: The DUET T-Cell architecture

To make data usable for decision making, trust into the data sources and the data chain is needed. Moreover, the data needs to be presented in a way that it is easy to consume, understood and relevant. We therefore recommend an approach to create and manage **smart data** at all levels in the T-Cell by participants, ensuring no weak or difficult to interpret or understand links in the eventual chain leading to insights for complex societal challenges, like shown in DUET pilots.

Smart Data refers to smaller sets of valuable and actionable information, extracted from big sets of raw data. The focus is to extract from raw data value, meaning, veracity for a certain purpose or outcome (see figure 1). Raw data is mostly available in big quantities from various sources in structured and unstructured formats, boosted from Society 4.0 (Information Society/ Industry 4.0) enablers[1]. Current state-of-the-art (cloud) technologies enable to capture, stream and store the data continuously.

[1] https://news.cgtn.com/news/2019-06-28/What-is-Society-5-0-at-the-G20-summit--HT4YQ8BXIC/index.html

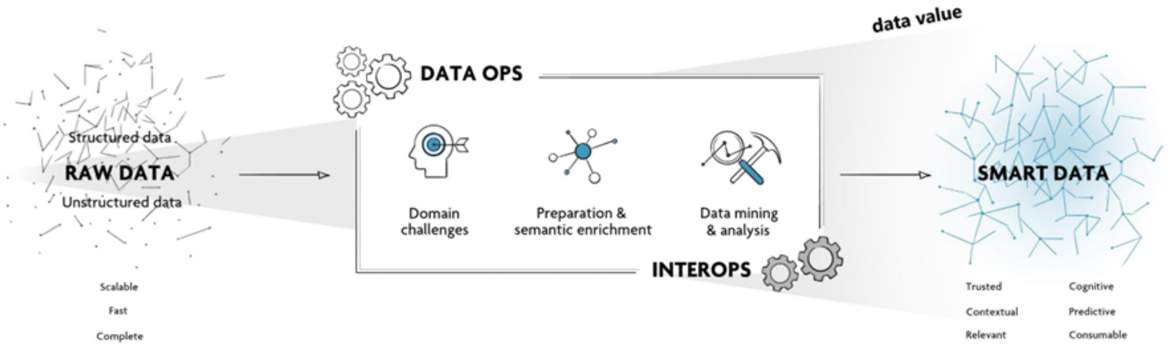

Figure 2: The smart data creation process

DUET recommends the use of (cloud) technology to transform big sets of raw data into policy-ready smart data by a process-vision that uses “data-ops” and “inter-ops” to prepare the data for the domain challenges, enrich it semantically and apply the right data mining and analysis to finally create smart data. The six main properties of smart data should be that it can be (1) trusted (e.g. by implementing lineage and provenance), is (2) contextual (e.g. by adding metadata), (3) relevant (e.g. by tuning it to the domain needs), (4) cognitive (e.g. by focusing on understandable patterns in the data), (5) predictive (e.g. by adding context on changes or evolutions) and (6) consumable (e.g. by making it easy to digest by computers and humans). These properties are needed to make data policy-ready and assure that the data can be shared within the digital twin framework to guide trusted decision-taking.

Creating policy ready, smart data, requires common practices, templates and agreed upon standards used in the technology to address the main properties, especially in a federated landscape that is so typical for digital twins. We therefore recommend:

- **to work on “data-ops” and “inter-ops” standardisation** for inter and intra cloud providers to realise as much as possible the same quality levels. Examples are using existing standards to realise the properties (e.g. linked data vocabularies or smart data models for context), or investigating new standards (e.g. to track lineage and provenance of data across multiple organisations) to set the next steps into uniform preparation of policy-ready data.
- **for standards, best practices and agreed upon frameworks** to enable governance, which is an essential step towards more trustworthy decision making based on data, as policy-ready data really aims in the end at addressing the right level of accountability.
- **for the data protection considerations** discussed in 4. Utilising cloud computing - legal requirements and recommended practice to be woven into policy ready data **to ensure privacy-by-design** in the setup, organisation and data protection measures of data processing by Smart Cities.

Today (End of 2022), it is evident that the European Commission has supported data spaces and offers the central framework and building block towards smart and re-usable data for policy making. There is no need today to stress the shift from cloud infrastructure to a more smart data-driven data and even simulation modelling-centred approach. The focus has therefore shifted from the more technical and solution-oriented cloud infrastructure towards a more result-driven data spaces approach, supported by the DUET consortium.

3. Acknowledgment of previous work

During the DUET project, effort was made to convert the requirements of users of a Digital Twin into practical policy issues (translated into pilot cases) by linking appropriate data and simulation models.
In Section 3 of this report, we summarise the relevant results based on a chronological overview of the various reports related to cloud computing and data space technologies.

3.1. Reports covering Legal and Ethical aspects

Previous DUET deliverables focus on the legal and ethical aspects of the European cloud infrastructure and data spaces. D1.1 (Legal Landscape and Requirements Plan) focuses on the usage of the cloud and personal data processing, providing pilots with guidance on the legal framework of existing laws and regulations under which the collection and processing of data should operate. The deliverable also identifies certain gaps and differences at the EU Member States level (applicable at the time of writing).

D1.2 follows and expands on deliverable D1.1, dealing with legal requirements that are considered as the cornerstone of digital “ethics”, i.e., privacy and cybersecurity. The deliverable further expands upon the legal requirements and policies, for the purpose of creating a ‘Cities Guide to Legal Compliance for Data-Driven Decision Making’. In particular the guide relates to issues specific to the use of disruptive technologies, mainly AI and HPC that are used in the context of data aggregation and the automated and large-scale data processing for decision-making. D1.2 also discussed issues of disruptive technologies, such as the legal requirements for automated data processing, or data storage and security issues. D1.3 and D1.4 both iterated the Guide with D1.4 containing the final and most complete version. While the guide did touch upon the implications of using cloud computing as a pilot, the guide focuses on the surrounding issues of collecting and processing data in a Smart City context.

D1.5 (Ethical Principles for using Data-Driven Decisions in the Cloud) built further on both D1.1 and D1.2 with a specific sub section focused on Cloud Infrastructure (2.4), dedicated to an overview of current trends and initiatives in the area. At the time of writing, cloud-based services were not yet codified in any significant way so the deliverable is based on legal requirements in D1.1 and D1.2. The final iteration of the document (D1.6) also focused on the location of cloud data centers, which is further developed in this document, in addition to the usage of public/private data sources.

We also acknowledge efforts made outside of DUET including the Guidelines on the use of cloud computing services by the European institutions and bodies (2018) from the European Data Protection Supervisor[2]. The Ethics and data protection document produced by the European Commission (2021)[3] further addressed the issues arising from the use of European Cloud infrastructures, European Commission Advanced Technologies for Industry – AT WATCH - Technology Focus on Data sharing (April 2021)[4] and on Cloud Computing (February 2021)[5], and European Commission Staff Working Document on Common European Data Spaces (February

[2] https://edps.europa.eu/sites/default/files/publication/18-03-16_cloud_computing_guidelines_en.pdf

[3] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

[4] https://ati.ec.europa.eu/sites/default/files/2021-04/Technology%20Focus%20on%20Data%20sharing.pdf

[5] https://ati.ec.europa.eu/sites/default/files/2021-02/AT%20Watch%20Cloud.pdf

2022)[6]. The content and the structure of these documents have been influential to the approach taken in D1.7.

3.2. Reports covering Smart City domains, interaction frameworks, cloud computing and simulation modelling

Previous deliverables have also addressed and outlined technicalities when working in a cloud environment.

D3.3 outlines the different DUET utilised models, covering several Smart City domains and interaction frameworks. The report clarifies that the DUET modelling partners (TNO, KUL, P4ALL) and involved 3rd parties (VITO) have taken different approaches to describe their models. Each should contain a conceptual description of the model, more detailed technical requirements/inputs, some notions on validity and its dependency on data and a link to user-specified pilot cases. The introduction of this deliverable addresses concerns regarding matching user requirements on the simulation capabilities on a more conceptual level and sketches how models operate at different geographical scales (from regional to local neighbourhood level). The outcomes of a model are also closely linked to the effort involved in data acquisition - finer-grained models typically need data at a higher spatial resolution - and on the other hand, it affects model validity; the functional relationships and parameters that go into a descriptive model are typically developed for a range of geographical scales and may not be valid when 'zooming' in or out too far. D3.3 concluded that for a successful study within the Digital Twin environment, the user needs to be aware of such pitfalls and choose the appropriate tools to answer their policy inquiries.

The DUET report regarding the IoT stack and API specifications was released after the D.3.1 report and presented a conceptual architecture focused on onboarding IoT data sources in a scalable way into the DUET broker. The DUET data and message broker can be seen as a direct user of the outcomes of the future EU data spaces and cloud computing initiatives and takes into account existing standards like NGSI-v2 and NGSI-LD. Three data types were discussed: IoT data used in the early alpha version of the DUET Digital Twin using a Fireware NGSI-LD adapter, IoT historical data and geospatial data receptors. A data catalogue containing simulation models has been designed and later implemented to link the data spaces.

The D3.5 Cloud design for model calibration and simulation focuses on the simulation model integration. However, every model can use its cloud and High-Performance Computing infrastructure to do the calculations. An interaction API as a DUET core component is needed to take care of the cooperation between simulation models. The DUET report refines the T-Cell architecture and its modular approach enabling dynamic and on-demand attachment of models to the DUET system. Using a message streaming platform also allows the DUET system NOT to set a fixed sequenced chain of models. Models will be signalled to run when changed data or event messages instruct a model to perform a 'run'.

The D3.2 report discusses how Smart Data Management principles are essential for realising and maintaining Digital Twins. This touches directly on the importance of constant interaction between data spaces and clouds to allow Digital Twins to use the most recent and updated qualitative data; for example, data standards for geospatial data and IoT combined with interoperability frameworks like Oslo and Interoperable Europe. A specific element touching on the legal and even ethical aspects is how data spaces, cloud

[6] https://ec.europa.eu/newsroom/dae/redirection/document/83562

infrastructures and digital twins themselves deal with personal and other GDPR-related data.

The second version of the Smart City domains, models and interaction frameworks, D3.4, looks into future developments and enrichments of Local Digital Twin solutions. Many of these future developments are directly related to the conception of data spaces and cloud services. The first element is the level of detail needed to run different simulation models. For instance, the current traffic models in DUET focus on the operational level, where the impact of route choice on loads and externalities (air quality, noise) can be investigated. Other decisions like modal choice, departure choice and activity location choice, as well as other activity and spatial decisions at the demand side of passenger transport, require complementary model sets. Awareness is needed about the trade-offs between broader scope and higher resolution. Moreover, even at high resolution, any traffic model will inevitably be more valid in predicting aggregate flows (e.g. at major roads like arterials and motorways) as compared to local traffic.

A second element influencing data spaces and cloud services is the synchronisation frequency with the physical world: the models require different data and calibration techniques, depending on their use is intended for second-to-second real-time tracking for short-term prediction, day-to-day tracking for next-day predictions, or tracking of slow-moving changes for strategic impact of changes to infrastructure or activities.

To conclude, next to the need for slow and fast-moving (IoT) data, the evolution towards more fine-grained models delivering higher validity, multi-period assessment, and a wider variety of outputs also influence the cloud and data space demands. These approaches will also focus on more complex HPC models and asynchronous computations instead of the synchronous less fine-grained model computations today. Also, the inclusion of complementary aspects within the traffic and mobility domain (e.g. parking, transit, pedestrians, cyclists, mobility as a service) and in connected Smart City domains like population modelling, city logistics, housing/land-use, energy system, not covered yet are examples of how domain expert knowledge is to be expected as part of future Digital Twin solutions. Such endeavours require the development of a broader urban digital twin ontology formally as part of future (thematic) dataspaces describing all relevant entities and their properties that models and data should quantify in existing and what-if scenarios.

4. Utilising cloud computing - legal requirements and recommended practice

4.1. Model and data catalogues for the commercialised product

Section 4.1. Integrates conclusions of DUET Deliverable 7.7 (Business and exploitation scenarios) and D5.2 (Digital Twin Prototype) from the perspective of solution providers and (end-)users.

4.1.1. View of the solution providers

The DUET solution providers and third parties are still focused on offering policy silo-oriented solutions. There is a noticeable shift towards AI simulation models (e.g. new generation of traffic models). TNO was the only DUET partner offering policy domain overarching solutions where a wide range of city data from different silos can be used with different simulation models (traffic models, air quality and noise). Interoperability with other model providers (outside their organisations) was a broadly unknown terrain before DUET.

DUET shifted the vision of all technical DUET partners towards a more standardised way of working, combining domain-specific standards with their models (e.g. OpenLR as part of traffic modelling) and domain overarching standards to allow communication between simulation models. Recent discussions at, e.g. the Open Geospatial Community (OGC)[7] (related to their work on OpenMI) made clear that only a few standardisation experts have independent multi-simulation model interaction on their radar. DUET has brought new, very innovative insights supported by all technical partners that need further adoption.

4.1.2. View of the users

DUET users are not directly concerned with the use of cloud computing or model interaction standards, as these do not have a direct functional goal, but serve the scalability of digital twins. However, the speed of their implementations is hampered by easy access to and use of data and simulation models from different vendors or origin. DUET made clear that a one-does-it-all solution is not preferable and probably impossible. Especially the Flanders and Pilsen cases make it very clear that different cloud computing-based simulation models have to work together and use a combination of official datasets and simulation models (e.g. city or regional traffic models). Interaction with other cities using a Digital Twin like Rotterdam, Helsinki, and Sofia made clear that others share the same concerns and visions to avoid vendor lock-in. Data spaces, catalogues, extended to Simulation spaces and catalogues integrating new business models like simulation on-demand services are seen as worth exploring.

This section further builds on the preliminary considerations which were laid out as the ethical framework for ethics based decision making in the cloud in D1.5. The following is supplementary to Section 2, **Preliminary considerations on ethics and the Cloud,** as published in D1.5. Furthermore, readers should review the below in conjunction with the **DUET ethical principles** as published in D1.5.

[7] https://www.sciencedirect.com/science/article/abs/pii/S1364815219303391?via%3Dihub

4.2. Managing the cloud and dataspaces

4.2.1. Location of the cloud and dataspaces

While location data as a data type has been addressed in D1.5, the topic of location of data storage is also important for Smart Cities to consider. It should be recalled that the location of data processing is important under GDPR:

1. personal data of European citizens must be stored in the EU or,
2. personal data may only be transferred to a third party country based on an adequacy decision or subject to appropriate additional safeguards.[8]

Compliance with this rule can be problematic in practice:

1. Service providers may not provide transparency into where their data processing and storage occurs;
2. While a cloud may utilise a data centre located in the EU in normal operation, the distributed nature of cloud infrastructure means it may be difficult to guarantee that the data remains in the EU in all circumstances (for instance in the case of backups or contingency provision after the failure of the normal data centre);
3. Service providers may be seated within the EU, however, the data may be processed outside the EU (whether by the services provider itself or by its subcontractors).

That having been said, some cloud operators and service providers are now providing more information about how their services can be used for EU personal data in compliance with the GDPR[9] and it is not unreasonable to expect prospective suppliers to be able to address GDPR issues in their proposals.

Therefore, in a Smart City context, it is important for Smart Cities to:

1. duly **assess if the GDPR applies to the service provider.** GDPR applies if (i) the service provider is based within the EU, regardless of whether the processing takes place in the EU or not[10], or if (ii) the service provider is not based in the EU but personal data of data subjects who are in the EU are processed in certain cases[11]. In general, the data center to be utilised for data processing and storage by the service provider should be located in the EU (for security reasons, only a general location may be provided before entering into the contract in order for the controller to identify the applicable laws[12]), otherwise GDPR requirements relating to transfer of personal data to third countries have to be fulfilled. Therefore, if the result of such assessment is that a country outside the EU may be involved as a country of processing and/or storage of the data, then,

[8] Article 45 and Article 46 GDPR.

[9] See for instance https://aws.amazon.com/compliance/gdpr-center/.

[10] Article 3 (1) GDPR.

[11] Article 3 (2) GDPR.

[12] Opinion 02/2015 on C-SIG Code of Conduct on Cloud Computing - wp232. Available at https://ec.europa.eu/newsroom/article29/items/640601/en.

2. identify whether such **country is covered by a current EU Adequacy Decision**[13]. The Adequacy Decision is issued by the European Commission if the third country, a territory or one or more specified sectors within that third country (or the international organisation) in question ensures an adequate level of protection to GDPR protection. The data transfer in such case then may be carried out. If no EU Adequacy Decision is issued, then,
3. if the service processor is situated in a country not subject to GDPR and is not covered by an EU Adequacy Decision, avoid such service providers in general or such locations. If those services must be used, make sure to implement additional safeguards of data protection as stipulated by GDPR, e.g. SCCs (please, see below).

In practice, the following situations may arise when Smart Cities consider the above steps.

GDPR requirements for the transfer of data outside the EU

One example is the use of a data center of a third-country outside of the European Union. While already addressed in D1.6 in the Ethical Code of Conduct, this section clarifies the considerations Smart Cities should make when storing data in the cloud outside the EU.

In the USA, the Clarifying Lawful Overseas Use of Data Act (the **CLOUD Act**)[14] allows for the United States law enforcement to request data stored in the United States and overseas. Under Article 48 of the GDPR, such a request from the US authorities may be enforceable on the controller of the personal data and therefore require the sharing of the data to the authorities, however, only if based on an international agreement, such as a mutual legal assistance treaty, in force between the requesting third country and the EU or a EU Member State. Moreover, the CLOUD Act provides the right for service providers to challenge such a request.

To maintain the same standards of protection as in the EU, on 16 July 2020, the Court of Justice of the European Union (ECJ) in its Case C-311/18 (the **Schrems II case**)[15] invalidated the US-EU Privacy Shield, a mechanism that had previously been in place to meet appropriate safeguards of data transfers from the EU to the US.

As a result, if a Smart City wishes to store and transfer personal data to the US or any other non-EU country, the following is recommended:

- Carry out a Transfer Risk Assessment addressing the flow of data and the receiving location's data regulations.
- Put in place with the service provider Standard data protection Contractual Clauses (SCCs) adopted or approved by the European Commission, Binding Corporate Rules (BCRs), codes of conduct to

[13] You will find the list of all current EU Adequacy Decisions on the EC's website: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en/.

[14] Clarifying Lawful Overseas Use of Data Act or the CLOUD Act (H.R.4943 — 115th Congress (2017-2018) found here: https://www.congress.gov/bill/115th-congress/house-bill/4943/text.

[15] Case C-311/18, Data Protection Commissioner v Facebook Ireland and Maximillian Schrems.

govern the processing of data in the third country or an approved certification mechanism together with binding and enforceable commitments of the processor in the third country to apply the appropriate safeguards, which do not require any specific authorization from a supervisory authority.

- If SCCs are opted for:
    - carry out a transfer impact assessment and identify cross-border transfers under the Smart City’s responsibility,
    - assess if an alternative location is available,
    - use the latest version of SCCs released by the European Commission.
- If BCRs are opted for, the BCRs have to:
    - be legally binding and applied to and enforced by every member concerned of the group of undertakings of the controller,
    - expressly confer enforceable rights on data subjects with regard to the processing of their personal data; and
    - fulfill certain content requirements regarding the BCRs as laid down in GDPR.

Apart from the above, the appropriate safeguards may also be provided for by contractual clauses between the controller or processor and the controller, processor or the recipient of the personal data in the third country. However, such clauses require a prior authorisation from the competent supervisory authority.

In the absence of an adequacy decision or of appropriate safeguards as described above, a transfer or a set of transfers of personal data to a third country may take place only on one of the specific conditions under Article 49 GDPR (derogations for specific situations), for instance, if:

- the data subject has explicitly consented to the proposed transfer, after having been informed of the possible risks of such transfers for the data subject due to the absence of an adequacy decision and appropriate safeguards;
- the transfer is necessary for the performance of a contract between the data subject (not e.g. their employer) and the controller or the implementation of pre-contractual measures taken at the data subject's request;
- the transfer is necessary for the conclusion or performance of a contract concluded in the interest of the data subject between the controller and another natural or legal person (e.g. data subject's employer);
- the transfer is necessary for the establishment, exercise or defence of legal claims;
- or in other cases as set out in GDPR.

If none of the above cases applies, a transfer to a third country may take place only if the transfer

- is not repetitive,
- concerns only a limited number of data subjects,
- is necessary for the purposes of compelling legitimate interests pursued by the controller which are not overridden by the interests or rights and freedoms of the data subject, and if
- the controller has assessed all the circumstances surrounding the data transfer and has - on the basis of that assessment - provided suitable safeguards with regard to the protection of personal data.

However, in such a case the controller has to inform the supervisory authority of such transfer. Based on the above requirements, it is clear that the controller seeking to store the data in the cloud outside the EU may not rely on this specific situation and will have a more onerous task of showing the protection of personal data.

Even though the above principles are hard law only when it comes to personal data, **they may serve as helpful guidance for state-of-the-art data management and principled engagement** with cloud infrastructures more generally.

Working with service providers and vendor review process

Based on the above, storing and processing of Smart City's data can also occur through the use of service providers.

When working with service providers it is recommended to assess the service provider's processing of personal data through a vendor review process as outlined under standards such as ISO 27001.

In the process the following components should be considered within the vendor review process:

- The organisation and technical security measures of the vendor considering an end-to-end risk assessment and continuous monitoring workflow
- Compliance with data protection laws and regulations
- Mechanisms and protocols in the event of a data breach or incident
- Transparency to users of data location, processing purpose, storage and how to exercise their rights.

In summary, Smart Cities should seek to maintain citizens' data within the EU where this is practically possible and, in circumstances where it is not, ensure an equivalent level of protection is maintained. Furthermore, cities should aim to maximise the transparency on how and where they process citizen data together with reference to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.

Protecting personal data in data spaces

As usage of cloud infrastructures evolve, so do the methods of organising data within these structures. While data platforms pose one location for storing Smart City's data, data spaces can pose another challenge for data protection through the shared nature of the space. Using data spaces would allow for data to be shared across nine industries: Health, Industrial, Agriculture, Finance, Mobility, Green Deal, Energy, Public Administration, and Skills. When utilising data spaces:

- The above guidance regarding the location of data centres should be maintained.
- Smart Cities should remain cautious about sharing personal data even within a data space. While there is no special legal guidance on the use of data spaces, similar precautions should be followed as per sharing personal data with other service providers as listed above, as enforcing and auditing compliance with data protection principles may be difficult in practice. This is to not expose the original owner to liability. If the utility of the data is not significantly reduced, Smart Cities should normally anonymise personal data to allow for a free flow of data within the data space, without the risk of harm to data subjects.

- Smart Cities may share personal data outside of the data space if an individual agreement is in place which meets the standard of data protection under GDPR.
- Clear sovereignty over data should also be imposed. This ensures there is a responsible owner who can respond to data subject requests if the data is not anonymised. Ownership of data is further discussed in 4.2.2 of this deliverable and should be considered in light of the Data Governance Act (the DGA) principle of data alturism.

4.2.2. The Cloud and Data Space Governance

In addition to the country and location where data is collected and processed in, Smart Cities should be aware of stakeholder's connection and relationship to data within the cloud and data spaces. This section therefore addresses (i) the current models under the GDPR including the role of data controllers and processors, (ii) the blurring of roles and stakeholder responsibility in hybrid cloud environments, and iii) new models of data ownership and responsibility as proposed under the Data Governance Act.[16]

Data controller-data processor relationship

As mentioned in D1.5, legal requirements applicable to the cloud are mostly uncodified, except for the GDPR which governs personal data processing in general, i.e. without specific reference to cloud service providers.

Furthermore, Deliverables D1.1. (Legal Landscape and Requirements Plan), and D1.2 ((Cities Guide to Legal Compliance for Data-Driven Decision Making It. 1) describe extensively the principal legal requirements in this area and provide guidance on the legal necessities that a Smart City should take into account in its decision and policy making processes. To ensure clarification of governance of the cloud and data space, however, Smart Cities should take into account several specific legal rules when operating within the cloud including:

- the role of a Controller/Processor
- processing under GDPR; and
- the role of Subcontractors.

Pursuant to the GDPR, processing of personal data means *any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.* Mere storing the personal data in the cloud is thus perceived as personal data processing.

Based on the data processing character:

- **cities are in a position of a data controller** being defined as *the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are*

[16] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)

determined by the EU or the EU Member State law, the controller or the specific criteria for its nomination may be provided for by the EU or the EU Member State law.

- **A cloud service provider is considered a data processor**, being *a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.*
- **all service subproviders, who provide cloud (or other data processing) services to the main cloud service provider, are subprocessors,** i.e. subjects processing personal data on behalf and under the responsibility of the data processor. Subprocessors are themselves liable to only process data on the instructions of the controller as shared through the processor, and may be liable for the failure to follow the instructions and safeguards they state to have in place. However, the data controller should be aware of all such subprocessors (and consent to them beforehand, if agreed so with the data processor) at any moment of processing of the personal data, as the data controller is liable for the manner by which the personal data is processed.

Cities, being a data controller, should be aware that they do not cease to be the data owner only by the mere fact of using services of the cloud service provider and storing the data on someone else's servers (data storage facilities). Pursuant to the GDPR, if the data processor infringes the GDPR by determining the purposes and means of processing, the processor shall be considered to be a controller in respect of that processing. However, this infringement of the GDPR does not relieve the data controller of their position as the data owner. Therefore, by virtue of storing the data in the cloud the data processor processes the data owned by other data owners and is liable to that data owner as well as to supervisory authorities for fulfilment of its duties the whole period of processing.

To mitigate risk, Smart Cities should therefore:

- use only data processors (cloud service providers) providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of the GDPR and ensure the protection of the rights of data subjects, i.e. subjects whose personal data are processed (stored) in the cloud.

- when possible, test technical and organisational measures. However, it is understood that, in practice, some characteristics may be specific to cloud infrastructures that may raise difficulties when applying the GDPR requirements. For instance, as a rule of principle, the data controller's right to audit or inspect the (cloud) data processor should be generally guaranteed and not limited to the case where the cloud service provider has not been certified by an independent body. The data controller may not always have rights to audit the processor due to various reasons. For example, the cloud provider having thousands of clients will not be willing to enable their clients to carry out such audits for practical reasons.

- carefully assess the cloud provider's Terms & Conditions with regard to audits as well, among other requirements.

Hybrid cloud services

Generally, in a public cloud environment, each data controller shares the same infrastructure with other controllers (cloud customers) using the same public cloud, provided by a service provider. In comparison, a private cloud may (i) allow for the Smart City's data to be segregated from other users and may not share the service provider's IT infrastructure and capacity with any third parties, but still facilitated off site or (ii) segregated from other users and facilitated on site. In the case that a private cloud equates to an on-premises cloud, while the controller enjoys the enhanced control over the data compared to the cloud solutions outside the data controller's premises, the data controller retains the responsibility to manage their cloud environment which often requires additional resources and added costs. Therefore, In the case of Smart Cities, a cloud operated by a service provider outside the data controller's premises and on behalf of the Smart City will be highly likely the most common instance due to the scalability, maintenance and cost.

A hybrid cloud can also be opted for. This combined solution may be divided into two types. The first is a "multi-cloud" solution where the data controller uses two separate clouds and chooses which data will be stored and which processing activities will be carried out on each of them. This approach may diminish the risk of compromising all the data as the data is stored on different, separate clouds and may also help the data controller avoid vendor lock-in by removing the dependency on one cloud provider.

The second is a "hybrid cloud", which is a single cloud environment consisting of a combination of dedicated and shared cloud infrastructures (public and private cloud environments) operating as one standardised technology. Thus, a hybrid cloud facilitates a variety of benefits including flexibility and scalability of resource capacity and cost efficiency.

In comparison to a public or a private cloud, one of the key benefits of a hybrid cloud solution is enhanced control over the data stored in the cloud. Rather than entrusting all aspects of the data controller's IT infrastructure to one cloud provider or one type of cloud infrastructure, within the hybrid cloud the data controller can distribute the risk. Therefore, the hybrid cloud may be a good choice for Smart Cities to consider if the data controller's own solution is not the option (see Section 4.2.3 below).

Competition and the effect on cloud governance: Ofcom's UK market study, Digital Markets Act (DMA)

Smart cities should also be aware of the evolving environment which cloud providers operate within. One example is the Ofcom market study. Ofcom is the UK regulator of communication services. Pursuant to the information published on its official website, Ofcom will launch a market study under the Enterprise Act 2002 into the UK's cloud sector which aims to formally assess how well the cloud market in the UK works, especially the status of competition in the market, including limitations to innovation and growth, and the position of the three hyperscalers, i.e. Amazon Web Services, Microsoft and Google which cloud providers together generate around 81% of revenues in the UK public cloud infrastructure services market.

Ofcom plans to consult their interim findings and publish the final report after the consultation with interested parties. The call for input was closed in November, 2022. Based on its findings, Ofcom can take one or more of the following steps:

- make recommendations to government to change regulations or policy;
- take competition or consumer enforcement action;
- make a market investigation reference to the Competition and Markets Authority (CMA);
- accept undertakings in lieu of making a market investigation reference.

In practice the output of the Ofcom market study could influence cloud regulation also within the EU in the future.

Similar valuable lessons can be drawn from other regulatory investigations into the cloud services environment, including the issue of dominance in certain types of services and the effect of so-called vendor lock-in (artificial degradation of interoperability between services of different service providers). An example of a regulatory scrutiny in this area is the unfolding antitrust probe into Microsoft's cloud services practices by the European Commission. On October 1, 2022, Microsoft implemented previously announced changes to its cloud services licensing agreements. These were supposed to make good grievances voiced earlier in 2022 by a range of complainants (Aruba, Nextcloud or Slack), according to which Microsoft is limiting consumers' choice in the cloud market through a set of abusive practices such as bundling and tying distinct services together, and making it harder to use rival services. Already on November 9, 2022, however, an industry group which includes Amazon, a rival cloud services provider, filed a fresh complaint to the Commission alleging that the changes were insufficient and moreover, added "new unfair practices to the list". Smart cities and other stakeholders in the data spaces industry should watch these investigations closely and actively participate in the process, because their views can influence the far-reaching behavioural or structural remedies that can be imposed by regulators within their sphere of competence.

Newly adopted EU legislation will also play an increasing role in shaping the cloud industry. Even though inclusion of cloud computing services in the scope of the Digital Markets Act (**DMA**)[17] is not without controversy,[18] Smart Cities are advised to note two main consequences it may have on tackling several persistent issues. First, Article 5 of the DMA prohibits designated gatekeepers to require users to subscribe to any core platform services as a precondition for using the gatekeeper's cloud services, which should mitigate the concerns about alleged tying/bundling practices of the leading cloud firms. Second, Article 6 of the DMA provides, among others, an obligation of the gatekeepers to provide end users with effective data portability and real-time access to data provided for or generated in the context of the use of the relevant core platform service (such as cloud service). This may further help enable Smart Cities to maximise value and flexibility to run their systems and services, if they are using services of companies that can be possibly designated as gatekeepers (Amazon Web Services, Google, Microsoft Azure, etc.). The proposed Data Act[19] may have similar implications for switching between cloud services and promotion of interoperability and data transfers.

[17] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance).

[18] For example, Damien Geradin; Dr. Konstantina Bania; Dimitrios Katsifis; Alexandru Circiumaru: The regulation of cloud computing: Getting it right, available at https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4285731_code362710.pdf?abstractid=4285731&mirid=1.

[19] Commission proposal for Regulation on harmonised rules on fair access to and use of data (Data Act) (February 2022).

Data Ownership

Firstly, the ownership of data collected implicating individuals should be considered. The movement from cloud infrastructures to data spaces while presenting a free flow of data, could potentially give rise to data protection implications for individuals - e.g from data being used for purposes and by stakeholders which the data subject was not aware of. If the data stored is anonymised, data ownership becomes less of an issue from the data protection perspective. This is due to the lowered risk to data subjects from data mishandling or processing for undisclosed purposes. However, once anonymised, the utility of such data can be reduced.

The concept of Data Alturism under the Data Governance Act (the 'DGA')[20], where entities can agree to their data being used for the public good, therefore may provide an alternative for Smart Cities while respecting data subject rights. To be in compliance with the DGA, Smart Cities should:

1. Comply with the requirements of the DGA, Article 15 to qualify for data alturism including:
   a. be a legal entity meeting general interest objectives;
   b. operate on a not-for-profit basis and be independent from any entity that operates on a for-profit basis;
   c. perform the activities related to data altruism through a legally independent structure, separate from other activities undertaken.
2. Follow the data alturism entity requirements under DGA, Article 19 by:
   a. ensuring appropriate disclosure is made to data subjects including that data collected will be used for the general interest;
   b. obtain consent from data subjects for the processing;
   c. and ensure the data is only used for general interest purposes.

Furthermore, the above concept may reduce friction between data protection and the free flow of data in data spaces.

Secondly, the ownership of the data processed within the cloud infrastructure should remain with the controller of the data, even when the cloud infrastructure has been influential in creating a new product through the processing of stored data. However, with the usage of cloud platforms or data spaces as locations for the processing of data, the accessibility to collected data can be wide and other stakeholders may also hold an interest in content generated.

Smart Cities, when acting as controllers of personal data, should take the follow into consideration:

1. Is the chosen cloud environment providing any additional services? If so, does the Smart City own all the outputs from this service?
   a. An example here includes using Cloud-provided AI tools. These tools can be used to improve data management, data hygiene, or any part of processing data in the cloud. In the process, Cloud AI tools potentially recognise, ingest and also classify data, therefore having access to

[20] The Data Governance Act defines Data Alturism in Article 2 as "the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services…".

a large quantity of a Smart City's data and potentially leveraging the data to create new outputs.

   b. **Action: If cloud provided tools have access to personal data, additional clauses should be agreed that processing shall only occur on the Smart City's instructions and no rights in the outputs shall be transferred to the cloud provider.**

2. Does the cloud provider/data space have access to data stored? Will the cloud provider use the data to improve their own services and products?
   a. **Action: If cloud provided tools have access to personal data, additional clauses should be agreed that processing shall only occur on the Smart City's instructions and the outputs will only be used by the Smart City.**
3. If data is shared in a data space, who has ownership over the imputed data and takes the role as a controller?
   a. **Action: Clear delineation of roles and responsibilities should be defined within the data space before the Smart City shares data.**
4. If data is shared in a data space, are there mechanisms in place to ensure the security and protection of personal data whether through technical or organisational security measures?
   a. **Action: The technical and organisational security measures and policies should be shared with the Smart City before the Smart City shares data.**

4.2.3. Contractual provisions

In addition to the above, the use of clauses in contracts can also allow for the protection of personal data when using the cloud.

However, it is important to remember due to the nature of cloud operators that many cloud contracts are not negotiable. Clickwrap contracts which are accepted upon signing up to the provider for example, can result in a Smart City having no discretion over additional clauses in the contract. In these circumstances it is important for Smart Cities to check the cloud agreement to ensure they have the answers to the following questions:

1. Where is your data being stored?
2. What rights does the cloud provider have to your data whilst it is stored in their hosted cloud environment?
3. What is the data breach mitigation, notification, and handling policy?
4. What happens to your data when the cloud contract is terminated? What is the process to remove your data from this cloud and/or transfer your data to another cloud?
5. How long will the cloud provider keep your data available for retrieval after termination?

More discretion over the agreement however can be granted if a Smart City chooses a private over a public cloud environment (multi and single tenancy agreements). In practice, private cloud contracts allow the customer of the cloud environment discretion to request changes, heightened security and negotiation in the contract provisions.

The below is therefore most applicable in the circumstances that a contract can be negotiated such as in private cloud agreements.

Service Level Agreements

The availability of a Smart City organisation's data may be integral to the functionality of the Smart City. Several considerations should be made when entering into Service Level Agreements (**SLA**s), including from the data protection perspective. In 2014 the Cloud Select Industry Group released Service Level Agreement guidelines[21] to standardise the provisions and terminology used. While other safeguards can and should be utilised to protect personal data, the guidance highlighted that in order to protect personal data, in SLAs the cloud provider should:

1. Outline the data processing purposes
2. Outline the data processing means including location and organisational and technical security measures
3. Provide clarity in data retention, deletion and disclosure
4. And make all the relevant information relating to processing available to ensure transparency

For the purpose of maintaining data protection SLAs should be used to ensure service availability especially in relation to security controls. For example, Recovery Time Objective (RTO) determining how long the environment can be offline and Recovery Point Objective (RPO) determining the acceptable amount of data loss can be used as indicators to determine if data protection is still met.

Furthermore, Smart Cities should be aware of the legal obligations of SLAs on cloud providers while appearing to allow room for a claim if there is downtime, may be drafted in such a way to prevent liability. For example, if all the failures are restricted to availability zones within a single region this could exclude even a significant downtime. Smart Cities should therefore read the fine print of SLAs carefully to ensure a breach of the SLA is what they expect it to reflect in practice.

Third-party Access Requests

Contractual provisions should be negotiated to ensure that the Smart City is aware of a data access request made to the cloud provider concerning the Smart City's data. As the Smart City remains the controller of the data when utilising a cloud provider as explained in The Cloud and Data Space Governance, they remain responsible for responding to requests of the individuals whose data is processed as a result of the Smart City's data collection.

For data spaces, third party access requests also solidify the importance of data sovereignty and ownership, to ensure it is clear which party is responsible to reply to such requests.

[21] https://digital-strategy.ec.europa.eu/en/news/cloud-service-level-agreement-standardisation-guidelines.

4.2.4. Technical security controls

Cloud deployment of tools and services to foster the adoption of digital twins and Smart Cities in general should follow cybersecurity best practices and identify security and data protection aspects that need to be taken into account when selecting a Cloud provider. As mentioned in D1.5, cybersecurity regulatory frameworks like the Cybersecurity Act and NIS have been in place to provide the foundation for cybersecurity assurance, while updated version are close to come in action, i.e. the NIS2 Directive has been provisionally adopted by the European Parliament (May 2022).

Based on these frameworks as well as on recommendations made by ENISA[22],[23] for cloud services, the following list briefly presents the main cybersecurity-related considerations for cloud providers that should be taken into account by organisations for a transition to the cloud[24]:

- **Identification of security and data protection requirements**, so that proper regulatory and technical compliance can be requested by the cloud provider.
- **Conduct of risk assessment and data protection impact,** so as to prepare relevant mitigation policies.
- **Establishment of processes for security incident management**, that will define responsibilities between the cloud customers and the cloud provider.
- **Endpoint protection,** so that client devices connected to cloud services stay protected from cyber threats.
- **Authentication and access control,** by defining policies, roles and permissions to resources and services.
- **Auditing, logging and monitoring** to enhance trustworthiness and accountability.
- **Network Security protection,** via measures like Intrusion Detection Systems, Firewalls, etc. that prevent external attacks and identify unauthorised or malicious traffic.
- **Vulnerability assessment, software updates and patch management,** that have to be periodically conducted by the provider to eliminate emerging securing threats and mitigate malfunctions or discovered errors.
- **Asset management,** including asset inventories, safe handling policies, decommissioning of hardware and usage policies by the cloud provider.
- **Classification of information,** so that an organisation can better manage data protection.
- **Data encryption at rest and in transit,** to ensure confidentiality and integrity.
- **Data backup and recovery policies,** to increase reliability.
- **Security of encryption keys** must be well defined and implemented.
- **Data portability and interoperability,** by following industry-standard data formats and communication APIs, thus avoiding vendor lock-in and ensuring continuity.
- **Data and network segregation within the Cloud,** to enhance confidentiality and integrity of data.
- **Secure data deletion upon request,** or contract termination with a cloud provider.
- **Software Development Lifecycle Security,** to ensure confidentiality, integrity, availability, accountability and authenticity of developed software and services.
- **Cybersecurity awareness, education and training,** so that all involved users stay alert of the potential threats and know how to protect themselves and their data.

[22] https://www.enisa.europa.eu/publications/cloud-security-for-healthcare-services/

[23] https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme

[24] *with regards to encryption, data loss prevention and digital rights management, together with on premise integration of logging and monitoring and Identity and Access Management facilities, form a third level of risk mitigation measures.*

- **Ensuring business continuity and disaster recovery,** by having relevant procedures in place for critical services in case of downtimes due to technical, environmental or other dangers.

In the context of DUET, a detailed presentation of the technical controls in place to address the above considerations has been provided in D3.11. Accompanied by a set of well-defined procedures and industry-validated standards to develop the offered services, DUET sets an example of a secure, sustainable and data-privacy preserving cloud deployment of a city's digital twin.

5. Risks, benefits and opportunities for cities in utilising the cloud infrastructure

The Risks, benefits and opportunities for cities reflect the reasons broadly described in the EU EDPS Guidelines on the use of cloud computing services by the EU institutions and bodies report[25]. The report considers the use of cloud computing services because of advantages such as cost savings in up-front and management resources, and partial or complete outsourcing of software applications, IT infrastructure and data storage. This would allow to reduce or avoid internal IT management tasks and efforts, as well as for new capabilities offered and, under some circumstances, a number of possible advantages such as a higher level of IT security assurance. They are nevertheless faced with the specific risks that the cloud computing paradigm involves and remain fully responsible regarding their data protection obligations. Chapter five covers two elements where the DUET project and the experience of its consortium members stand out: The use of data models and licences related to the data spaces experience and the extensive use and cooperation between simulation models. Based on the DUET experience on Digital Urban Twins and Local Digital Twins, a number of cloud based business models have been outlined, which some of them tested.

5.1. Use of data models and licences

5.1.1. Data & data spaces experience

Creating and sharing smart (meta)data should be an important objective for any digital twin data suppliers. That makes the data consumable and trusted. However, more is needed to solve the problems of sharing the data in practice. At the start of DUET, the T-Cell architecture was framed to indicate the sharing challenges and decompose a digital twin as a system of systems. At that time, data spaces as governed federated ecosystems were not yet a hot topic. Since then, organisations such as Gaia-X[26], IDSA[27] and, more recently, the DSSC have leveraged the EU data-space strategy.

In September 2021, BDVA[28], FIWARE[29], Gaia-X and IDSA launched the Data Spaces Business Alliance (DSBA) with a common objective to accelerate business transformation in the data economy. One of the joint working areas of the DSBA is supporting the existing organisations and data spaces by pooling their tools, resources, and expertise in a focused way. In this context, the four European associations have developed – as part of their missions, strategies, and operations – international networks of national or regional 'Hubs': the BDVA i-Spaces, FIWARE iHubs, Gaia-X Hubs and the IDSA Hubs. Together a network of almost 90 Hubs (and growing) distributed over 34 countries becomes a key asset for the engagement of multiple stakeholders in public and private sectors (in particular SMEs) and the development and deployment of data spaces in Europe and beyond.

Plan4all became one of the BDVA i-Spaces in 2021 and is now part of this DSBA hubs initiative aiming to contribute to the European data spaces. I-Spaces are cross-sectorial and cross-organisational innovation hubs

[25] https://edps.europa.eu/sites/default/files/publication/18-03-16_cloud_computing_guidelines_en.pdf.
[26] https://gaia-x.eu/.
[27] https://internationaldataspaces.org/.
[28] https://www.bdva.eu/.
[29] https://www.fiware.org/.

that combine data sources and AI technologies. It also integrates competencies and all the technical and non-technical aspects needed to allow SMEs and start-ups to get their data-driven and AI-related services, products and applications quickly tested, piloted, and exploited. I-Spaces aim at accelerating the take up of data-driven innovation in commercial sectors such as Manufacturing 4.0, Logistics, e-Commerce, Media, Aerospace, Automobile, Energy, Agriculture and Agroindustry, Pharmacy, as well as in non-profit sectors such as e-Government, Environment, Public Health, Smart Cities).

We strongly recommend using data spaces as the main linking pin to connect the systems of systems, as pitched in Lawton 2022 and recently pitched and repeated in Usländer et al. 2022. Data spaces address interoperability, trust, and data value and put a layer of governance on top of that. So they are ideal and instrumental instruments to fulfil the T-Cell main connection fabric and realise a real trusted and sovereign connection between the different contributors of a system-of-systems approach. The main building blocks of data spaces are shown in the figure below, as defined by OpenDei.

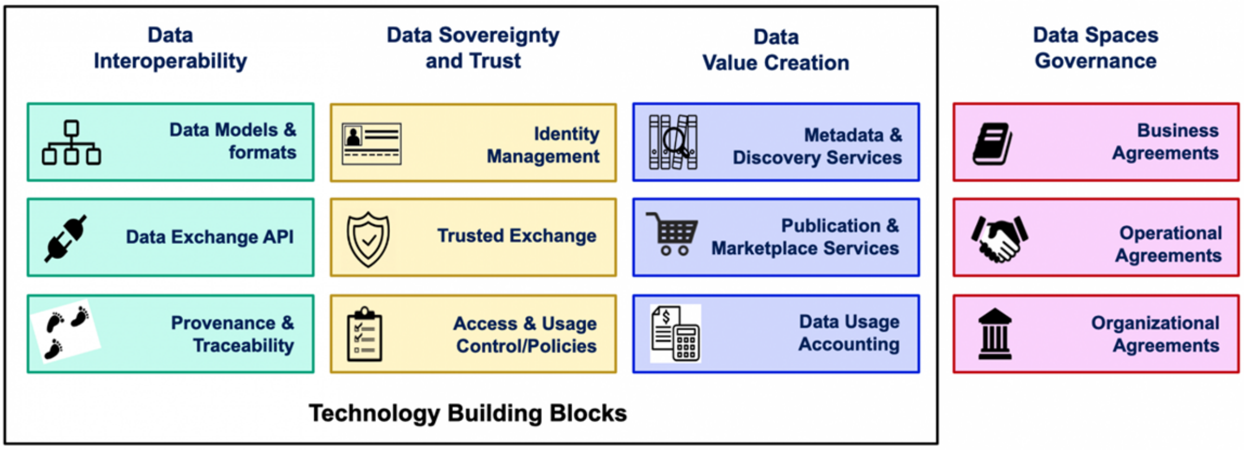

Figure 3: The main building blocks of data spaces are shown in the figure below, as defined by OpenDei.

In DUET, we used data and message brokers to simulate the federated aspect of the system-of-systems approach, leveraging the unique value of different players in the ecosystem, from IoT data suppliers towards intelligence providers using complex simulation algorithms. It helped a lot to identify the challenges. It still needs to replace a standardised way to address cataloguing (finding the data), using the right ontologies and vocabularies, accounting for data usage, allow provenance and traceability. Much broader cloud and digitisation activities are needed to address the challenges. Our strong recommendation is thus to build further on societal and public domain digital twins on top of the concrete outputs and outcomes from the different data space initiatives, with concrete attention on open standards, governance, vocabulary providers and certification.

As an example, the Plan4all datasets can be mentioned: Open Land Use - a dataset covering European countries, with an ambition to expand to other continents[30]; the EU-wide Open Transport Map dataset[31] is shared in OGC standards WMS and WFS, accessible via industry standard shapefile, and selected regions were converted into semantic RDF triples; Smart Points of Interests[32] - worldwide dataset maintained in RDF triples natively.

[30] https://hub.plan4all.eu/cs/olu.
[31] https://opentransportmap.info.
[32] https://sdi4apps.eu/spoi.

What's common for all these datasets is that they are stored in an interoperable way using a unified data model. There are multiple datasets in the world on points of interest. Each dataset has a different data model and stores different types of attributes for each point. This makes it difficult to use in applications. A common data model is a key for data reuse. In the case of Smart Points of Interest, a common data model was created in the SDI4Apps project[33]. Multiple heterogeneous databases of points of interest were harmonised into a single data model. In this way, Plan4all created the largest harmonised database of points of interest worldwide. Currently, there are activities of the Open Geospatial Consortium and, in particular, the Points of Interest Standards Working Group[34] to make a standardised data model for points of interest.

What's also common is that many initiatives are trying to standardise how data are handled and shared. In addition to the INSPIRE initiative of the European Union, there are efforts of the International Organization for Standardization, Open Geospatial Consortium, Research Data Alliance and open data initiatives. Data are and will be shared using different data specifications. That's also why initiatives aim to harmonise data across the different data specifications. This is not limited to using a common data model but also includes how data are distributed, in what quality and under which restrictions.

With restrictions on how data can be reused are connected intellectual property rights and data licenses. Most intellectual property rights can be sold or transferred. In some but not all jurisdictions, intellectual property rights can also be waived, placing the data in the public domain. In this case, the term waiver is used instead of data license. Data published as public domain can be reused without any restrictions, and no attribution is needed. The holder of intellectual property rights can license others to use the data while still retaining ownership. The licence is the legally binding terms and conditions of use and may limit the types of use, the duration of use or other ways of using the data. An open licence is a licence that conforms to the Open Definition: "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."[35] Licences conforming to this definition must permit any user not only to access the data but also to reuse it, change it, combine it with other data and re-distribute it, and to do so both for commercial and non-commercial purposes.

The Open Land Use dataset combines several data sources with different licences. In the case of the Czech Republic, four data sources, including CORINE Land Cover and Urban Atlas (hereinafter referred to as European datasets) with European coverage, and two national datasets, including cadastral parcels and Land Parcel Identification System. The Czech datasets are in the public domain (no restrictions), and access to the European datasets is governed by the Copernicus data and information policy Regulation (EU) No 1159/2013 of 12 July 2013. Free, full and open access to this data set is made with some conditions, including attribution. In this case, the public and commercial sectors can easily combine, share and reuse data. Certain problems can arise when combining data with licences that are not interoperable. In that case, combining data and their further reuse might be limited or impossible.

5.1.2. Simulation models

In the previous chapter, we identified data space technology as an important recommendation to create scalable digital twins at the heart of the T-Cell fabric, addressing data & model catalogues, interaction services, data exchange mechanisms and vocabulary sets. However, in DUET the concept of simulation

[33] https://sdi4apps.eu/.
[34] https://www.ogc.org/projects/groups/poiswg.
[35] http://opendefinition.org/.

models or computers that executes algorithms on datasets and produces new data is an essential part. There is no specific attention for managing chained models and data and making them interact together.
Managing chains of related data-compute-data cycles is an essential part of a system-of-systems digital twin. The orchestration needs to enable control and has been identified in the past as a challenge with attempts from OGC for standardisation: OGC® Open Modelling Interface (OpenMI)[36] Interface Standard.

The starting points at that time (2014) were really valuable, and still relevant within a digital twin ambition:

- Combination of models brings real value
- Interactions between models should be facilitated to keep model autonomy
- Freedom of each model algorithmization is important, but data in/out types need to be stable
- Data sharing semantics should be unified

However, at that time specific choices were made such as:

- Focusing on workflow management to answer a specific “ question “ needing chained models
- Using local chaining (not using different cloud assets) with master of workflow
- Models have to know each other and use interaction libraries directly
- And it was a data-pull driven architecture

Within DUET, some next steps were set to let models interact from within their cloud setups. However, the work is not finished, and more research and standardisation are needed to bring this up to a scalable level, just as data sharing aims for in data spaces. Therefore, we recommend investing in research on a clear separation of concerns in the split between models and the data they use on the one hand and explicit model integrations on the other. Model orchestration across different agents needs clear agreements, messaging and standardisation to move into models-as-a-service first and model marketplaces next. Having this in place within a dataspace is the next step for digital twins to access the intelligence and trust that intelligence to address complex cross-domain challenges very easily and in a policy-ready way.

5.2. Cloud based business models

The outcomes of Cloud-Based Business Models utilised in DUET are interesting in two distinct ways, first from the perspective of the Smart City and Urban Digital Twins and secondary from the perspective of Cloud Solution providers.

From a city perspective, the Business Model Analysis concluded in D2.4 gives cities a choice between different types of Digital Twins (Inside-In, Inside-Out, Outside-Out and Outside-In), which all have a different (data) business model.

[36] https://www.ogc.org/standards/openmi.

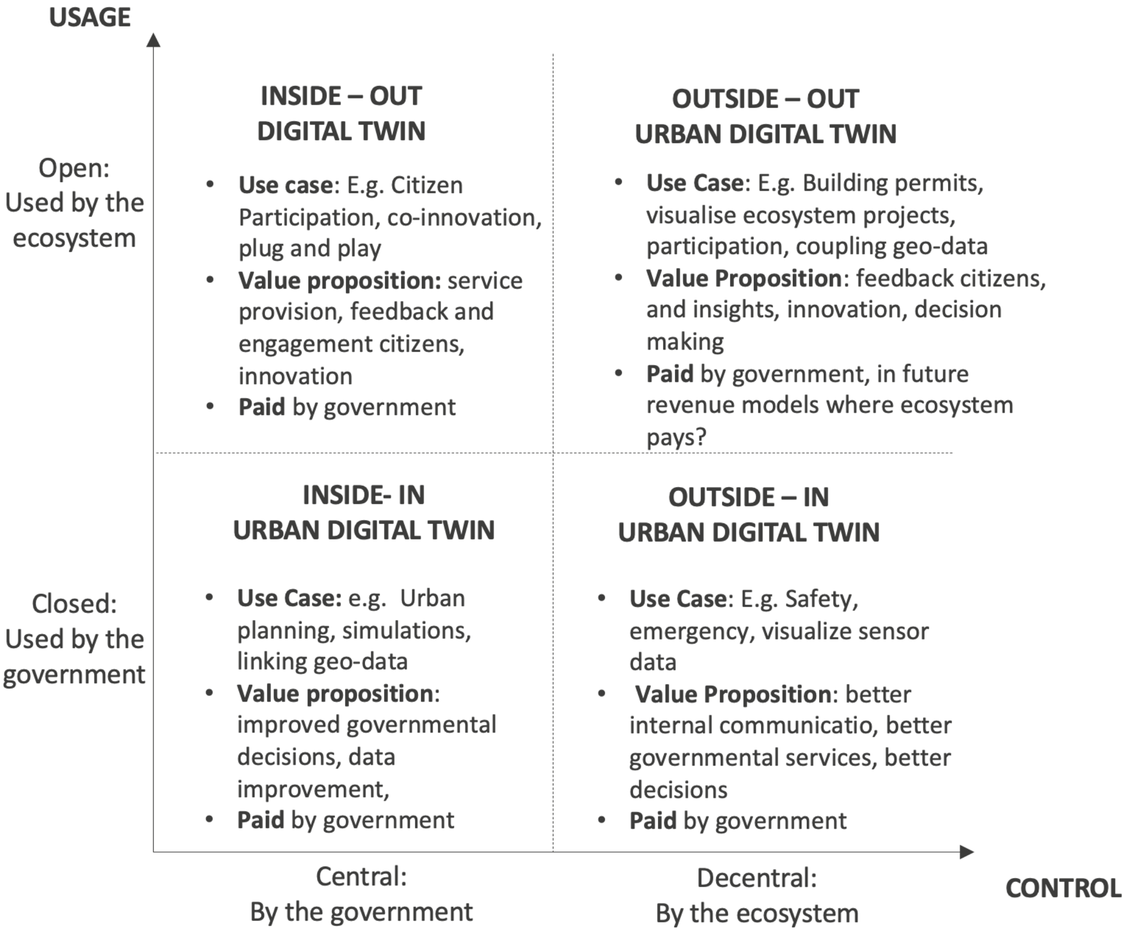
USAGE
Open: Used by the ecosystem
Closed: Used by the government
INSIDE – OUT DIGITAL TWIN
• Use case: E.g. Citizen Participation, co-innovation, plug and play
• Value proposition: service provision, feedback and engagement citizens, innovation
• Paid by government
OUTSIDE – OUT URBAN DIGITAL TWIN
• Use Case: E.g. Building permits, visualise ecosystem projects, participation, coupling geo-data
• Value Proposition: feedback citizens, and insights, innovation, decision making
• Paid by government, in future revenue models where ecosystem pays?
INSIDE- IN URBAN DIGITAL TWIN
• Use Case: e.g. Urban planning, simulations, linking geo-data
• Value proposition: improved governmental decisions, data improvement,
• Paid by government
OUTSIDE – IN URBAN DIGITAL TWIN
• Use Case: E.g. Safety, emergency, visualize sensor data
• Value Proposition: better internal communicatio, better governmental services, better decisions
• Paid by government
Central: By the government
Decentral: By the ecosystem
CONTROL

Figure 4: Use cases, value proposition and revenue models of Urban Digital Twins

This choice depends on the desired purpose, the data that needs to be included in the Digital Twin and the maturity of the Digital Twin. The four types of Digital Twins can co-exist, depending on the use case and not on the city. Thus, in one city, there can be the need for both an Inside-In Digital Twin for policy purposes and an Outside-Out Digital Twin for ecosystem purposes. The value can be created internally for the government (in the Inside-In and Outside Out Digital Twins); in these cases, the government also pays for the Digital Twin. In the Inside-Out and Outside-Out Digital Twin, the value can be internal (government) and external (for the ecosystem). In most cases, the government pays for the development and services from governmental funds. Still, alternative revenue models could be identified in the future where ecosystem actors (such as citizens, companies, and NGOs) would also pay for using the Urban Digital Twin infrastructure to access the data or the results.

For Cloud Solution Providers, the outcomes of the business model analysis show different Cloud Business Requirements which can be developed by DUET. Some of the challenges that the different Urban Digital Twins face can be related to the cloud strategy, as various components will need to be developed to facilitate the development of the Digital Twins. Therefore, the main features are the harmonisation of data, cloud-native brokers and support in moving away from vendor lock-in challenges.

The above models and approaches were first studied in DUET. The maturity model (for cities) provides a crucial guideline in determining exploitation and commercialisation strategies for cities integrating the DUET best practices on bringing multi-domain data and simulation models together. Regarding the Cloud technological solutions, it is essential to scrutinise how to open Digital Twin solutions like DUET will position themselves in the Urban Digital Twin ecosystem. The future exploitation and commercialisation strategy is not only a first test of the maturity Model. Still, it will also explore the potential for commercialising open cloud infrastructure and data spaces related to open digital twin solutions.

6. Recommendations to increase awareness

To conclude this deliverable, a list of nine main recommendations is derived based on the DUET experience with cloud computing and data spaces. The experiences translated into recommendations go beyond the Digital Twins itself. This is obvious given that a Digital Twin can be seen as an important consumer of cloud infrastructure and data spaces. The cloud and dataspaces are the main enablers to realise a Digital Twin and its needed governance. Six recommendations are related to the use of the cloud and the data spaces, while three focus on the legal and ethical aspects.

6.1 Recommendations regarding the use of the cloud and data spaces

Recommendation 1: A further shift from infrastructure-driven thinking about cloud infrastructure towards more data and information-driven approaches like data spaces is needed;
Recommendation 2: Data spaces are, from an Urban/Local Digital Twin perspective, an essential but only partial part of the solution. Next to data transfer is also a transfer of messages needed;
Recommendation 3: Data spaces can be extended with a simulation model and/or AI algorithm (sub)space to support smart data processes;
Recommendation 4: Interoperability in and between dataspaces is key to helping break down policy cycles to allow future Local Digital Twins to help realise ambitious policy goals, e.g. the Green Deal;
Recommendation 5: Learn from long-lasting open data initiatives like the P4All standardised datasets including the Open Land Use map, Open Transport Map and Smart Points of Interest dataset;
Recommendation 6: Extend initiatives like interoperable Europe and Inspire a wide range of Smart City-related datasets needed for future policy prediction, making and visualisation.

6.2 Recommendations regarding the ethical and legal aspects

Recommendation 7: Cloud services and data space stakeholders should take an active part in regulators' investigations of the industry by helping shape appropriate regulations/remedies to tackle persisting problems, such as abusive practices by leading service providers (gatekeepers) or the vendor lock-in problem.
Recommendation 8: It is evident that legal frameworks are curated for an infrastructure focused on cloud infrastructures rather than data and information focused data spaces. For Smart Cities, it will be important to define data sovereignty principles and clear ownership of data to allow for protection of personal data as discussed in Section 4, while ensuring the free flow of data, such as through data altruism. Furthermore, inspiration should be taken from other data space initiatives for future legal framework evolution.
Recommendation 9: Regulators should make full use of the available and prospective legal toolkits, including from the DMA, and future Data Act as discussed in 4.2.2., to tackle identified problems in the cloud industry. Under enforcement of this area may lead to sustaining or even deepening structural or behavioural issues in the industry.