D6.6 AVAILABILITY OF MEASURED MAINTENANCE DATA OF INFRASTRUCTURE FOR PUBLIC DOMAIN

@AshvinH2020
ASHVIN H2020 Project
www.ashvin.eu

Project Title	Assistants for Healthy, Safe, and Productive Virtual Construction Design, Operation & Maintenance using a Digital Twin
Project Acronym	ASHVIN
Grant Agreement No	958161
Instrument	Research & Innovation Action
Topic	LC-EEB-08-2020 - Digital Building Twins
Start Date of Project	1st October 2020
Duration of Project	42 Months

Name of the deliverable	Availability of measured maintenance data of infrastructure for public domain
Number of the deliverable	D6.6
Related WP number and name	Social Innovation and Standardisation
Related task number and name	T6.5 Safe, environment & productive maintenance planning
Deliverable dissemination level	PU
Deliverable due date	31-09-23
Deliverable submission date	23-11-23
Task leader/Main author	Selma Toktas (EUR), João Gonçalves (EUR), Hind Serkouh (EUR), Enrique Canadas (EUR), Jason Pridmore (EUR)
Contributing partners	
Reviewer(s)	ASI, AUS

ABSTRACT

This deliverable concerns the theoretical grounding of open data efforts for maintenance data of public infrastructure, as well reports of ASHVIN practices in this regard. It also provides a framework and recommendations that can be applied to other kinds of public data and describes the exemplary implementation of these recommendations in the ASHVIN Zadar Airport demonstration case. It provides an hierarchical structure for maintenance data availability, and a set of five ADAPT principles that guide data availability decisions and implementation.

KEYWORDS

Open data, maintenance, infrastructure, data equity

REVISIONS

Version	Submission date	Comments	Author
V0.1	18/10/2023	Internal Review Version	EUR
V0.2	17/11/2023	Internal Review	AUS
V0.3	20/11/2023	Internal Review	ASI
……			
…….			

DISCLAIMER

ACRONYMS & DEFINITIONS	
AI	Artificial Intelligence
ADAPT	Availability, Demonstration, Accountability, Proportionality, Transparency
ASI	Australian Standards Institute
AUS	AUSTRALO Marketing Lab
BIM	Building Information Modelling
CERTH	Centre for Research and Technology Hellas
DPIA	Data Protection Impact Assessments
EUR	Erasmus University Rotterdam
GDPR	The General Data Protection Regulation
GISI	GIS Integrator for Digital Twin-based Asset Management
IoT	Internet of Things
ISMS	Information Security Management Systems
ISO	International Organisation for Standardisation
KPI	Key Performance Indicator
MatchFEM	Multi-physics Model Matching Tool for Status Assessment of Bridges and Buildings
NGO	Non-governmental Organisation
PDCA	Plan-do-check-act
PIMS	Privacy Information Management Systems
RISA	Risk-based Status Assessment Tool with KPI Dashboard
Open BIM	Open Building Information Modelling
4DV-D	Construction Site Simulator for Early Design Phases with 4D Visualizer

ASHVIN PROJECT

ASHVIN aims at enabling the European construction industry to significantly improve its productivity, while reducing cost and ensuring absolutely safe work conditions, by providing a proposal for a European wide digital twin standard, an open-source digital twin platform integrating IoT and image technologies, and a set of tools and demonstrated procedures to apply the platform and the standard proven to guarantee specified productivity, cost, and safety improvements. The envisioned platform will provide a digital representation of the construction product at hand and allow to collect real-time digital data before, during, and after production of the product to continuously monitor changes in the environment and within the production process. Based on the platform, ASHVIN will develop and demonstrate applications that use the digital twin data. These applications will allow it to fully leverage the potential of the IoT based digital twin platform to reach the expected impacts (better scheduling forecast by 20%; better allocation of resources and optimisation of equipment usage; reduced number of accidents; reduction of construction projects). The ASHVIN solutions will overcome worker protection and privacy issues that come with the tracking of construction activities, provide means to fuse video data and sensor data, integrate geo-monitoring data, provide multi-physics simulation methods for digital representing the behaviour of a product (not only its shape), provide evidence based engineering methods to design for productivity and safety, provide 4D simulation and visualisation methods of construction processes, and develop a lean planning process supported by real-time data. All innovations will be demonstrated on real-world construction projects across Europe. The ASHVIN consortium combines strong R&I players from 9 EU member states with strong expertise in construction and engineering management, digital twin technology, IoT, and data security / privacy.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

1 INTRODUCTION

The Morandi bridge collapsed amidst heavy rain in Genoa, Italy in 2018. Thousands of buildings collapsed after the earthquake that took place in number of different cities in Turkey and Syria in 2023. Part of the main street of Fukuoka in Japan collapsed in 2014, with officials linking this to ongoing subway construction nearby. After each of these disasters, the same fundamental question was raised: How could these disasters have been prevented? Even if, in the case of an earthquake, the conditions may not have been foreseen or prevented, were there ways in which the effects could have been reduced and the consequences been less detrimental? One way to do this is to begin making infrastructure data more public and subject to different levels of scrutiny. The key question to be clarified here is whether open and usable data accessed by interested stakeholders, ranging from industry professionals, research scientists and local interest groups, will lead to better infrastructure accountability. To find an answer to this question, this deliverable aims to discuss whether and what maintenance infrastructure data can be made available in demo sites where ASHVIN digital twin is applied. At the same time, we need to define what risks and benefits there are for making infrastructure data public. From the ASHVIN experience, general guidelines can be derived for other maintenance data in the European context.

These suggested guidelines emerge in the context of a long-standing debate about how open public data creates economic and social benefits and value for all stakeholders, including citizens, companies, and public organisations. While in part open data allows for accountability, "the publishing of government data in a reusable format can strengthen citizen engagement" by increasing citizen trust in the government and their participation in administrative process (Huijboom & van der Broek, 2011, p.1). A primary focus in this regard has been on data about citizens, but the same rationale can be transposed to public infrastructure contexts which may serve to increase government trust and accountability. However, there is a tension with ensuring open and clear forms of data publication with issues of security. A clear sense of how open data can and should be open while ensuring that risks and vulnerabilities of public infrastructure are limited is important.

Given the proliferation of big data and advancing data analytics, especially as this extends into the construction sector, there has been an increase in investments in new technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), cloud computing and more within a data-driven business environment. As argued by Jati and others, AI is believed to produce better and more precise insights with more data and open data is seen to play a crucial role in enabling effective AI (Purnama-Jati et.al., 2022, p.875). With the advent of these technologies, making data openly available has come to the forefront as one of the main strategies in ongoing governance discussions intended to provide a more transparent and participatory service to its citizens. Those that promote open data claim it brings a better governance strategy and a more democratic approach.

At the same time, there is an increased demand by citizens for government data to be made public. This stems from policies and practices that value a more participatory management approach in which the private sector organisations, citizens, and governmental bodies partner together in decision-making processes, rather than an approach in which governmental bodies are sole decision makers. The basic principles of this form of governance focus on concepts such as openness, transparency, accountability, sensitivity to citizen demands and participation. A primary way to achieve this is to develop and implement open data-oriented strategies such as developments in open science and open data. Open science focuses on the publication of research results while open data is focused on making data publicly accessible. The intention of both practices is to increase citizen engagement in order to spread open science/citizen science to democratise knowledge production, and open public data discussions to allow for a more participatory management process. Although this focus highlights the value of openly available public data, it also “can be perceived as an attempt by governments to provide private companies with access to valuable resources, which they can use to address societal challenges while still caring for their bottom lines” (Jetzek et.al., 2019, p.706). In short, making public data available in an open form is done for a variety of reasons, ranging from increasing the efficiency of operations, providing better services, and supporting research and policy making. All these aims to realise the basic promises of a more open, participatory, and democratic governance approach.

In the context of this report, open data is conceptualised as data that is made available to enable free and reuse of generated data from public infrastructure. This data, which is made publicly available in accordance with the laws and regulations of each country, is mostly shared without being subject to restrictions such as copyright and patents. This report provides an assessment of different initiatives on open data from multiple perspectives, before discussing the extent to which it is possible more specifically to make maintenance data from sites open using the ASHVIN digital twin tool. To do this, we demonstrate the conceptualisation of open data and its effects, starting from a baseline understanding of data equity or data justice as an ethical way to promote open data practices. This then indicates ways forward in making maintenance data open in the digital transformation process of the construction industry.

The perspective taken in this text is that while open data can provide some form of information justice, it can also create more problems when it is not dealt with appropriately. Johnson notes that “open data has the quite real potential to exacerbate rather than alleviate injustices” (2014, p.263). For instance, certain established and successful companies in the maintenance field could benefit from having cost-free access to data to improve their market position even further in a way that less advantaged companies would not be able to do. Therefore, a clear and transparent approach to open data requires a focus on data equity.

In the current political context in Europe, an awareness of and desire for open data has increased predominantly at a conceptual level, but a culture of practice has not yet emerged in the same way. One of the key factors relates to concerns about commercially confidential and sensitive personal data. Many organisations are struggling to share their internal data with others due to the lack of “knowledge on how to adapt sensitive data suitable for open data publication” (Donker & Leonen, 2016, p.300). If these difficulties are addressed, fair data governance becomes a practical possibility, including for maintenance data.

Open data plays a major role in ensuring fairness and equity through transparency, equitable access, and engagement. This corresponds with the concept of data justice, a core concept of the ASHVIN approach to open data follows. A data justice approach indicates the main paths for data to be produced, collected, analysed, and used in a more fair, secure, sustainable, and autonomous framework in an increasingly datafied1 construction sector. Establishing data justice ensures that the outcomes of open maintenance infrastructure are positive and desirable, preventing negative consequences such as corporate monopolies or data misuse.

For the reasons outlines above, the concept of data justice is followed within this deliverable, focusing on how maintenance data produced at the sites using ASHVIN digital tools can be made publicly available as well as offering broader suggestions for best practices with open infrastructure data. Translating data justice into practical recommendations, the ASHVIN project offers five steps that should be taken when opening public infrastructure data. This is described as an ADAPT (Availability, Demonstration, Accountability, Proportionality, Transparency) framework. By complying with these five principles, not only the risks and concerns about opening infrastructure data can be addressed but the potential for unequal access and unfair competition among stakeholders is minimised.

1.1 SCOPE AND OBJECTIVES

This deliverable focuses on the availability of measured maintenance data of infrastructure for public domain. It describes the public use of data related to the physical behaviour of assets needed for the maintenance of public infrastructure. The report also describes the exemplary perspectives on open data policy on ASHVIN demonstration projects, particularly on the Zadar airport case. It also discusses how to persuade the public and private sectors to promote the combined use of digital twin technology and open data, and how to expand the disclosure of data obtained through digital twin technology. While searching for the ways to do this, answers are sought to these two basic questions:

1- How can public organisations or private companies overcome barriers to opening maintenance infrastructure datasets?
2- What should be the guiding principles for open data?

In addition to these key questions, the overarching question posed inherent to the concept of open data is how to balance and integrate privacy and transparency with the needs for autonomy and integrity. Undoubtedly, this situation depends on country-specific laws and regulations, as well as cultural norms (Scassa and Conroy, 2017,

[1] As discussed in detail below datafication points to converting events, activities, behaviours, and processes in real-world into digital data that can be stored, processed and analysed.

p.338), but the ASHVIN data pyramid presented in Section 2 aims to visually depict ways to address this balance.

D6.6 sets out the theoretical foundations and assumptions for openness of infrastructure data, recommends paths to openness in the context of digital twin technology, and showcases how data policy was implemented in public infrastructure maintenance demonstration sites under the ASHVIN project. By reporting on these aspects, this deliverable aims to provide the reader with:

1. A fundamental understanding of the key theoretical assumptions underlying data availability.
2. A set of concepts and guidelines on availability of public infrastructure data that can be adapted and applied to a wide range of public infrastructure maintenance projects.
3. An overview of open maintenance data policy implementation on ASHVIN.

This deliverable can benefit any reader associated with public infrastructure maintenance data, however, the theoretical concepts, best practices, and practical recommendations outlined here are most likely to directly apply to members of organisations that make decisions regarding the availability of data.

1.2 STRUCTURE

This deliverable is structured in six sections. The datafication process, which forms the basis for the disclosure of open data and especially maintenance data in the construction sector, is given in general terms in Section 2. Section 3 provides the conceptual discussion on open data is presented and provides answers to the following questions: What counts as open data? What methods and procedures does open infrastructure data generation involve? How is it conceptualised by the actors from the construction industry? What are the challenges and benefits of open data? Section 4 discusses how open and usable data potentially increases security and safety risks and how we can prevent that open data is used by actors with harmful intentions. Section 5 presents the status of data policies on maintenance demonstration sites under the ASHVIN project, providing insights into why some public owners of infrastructure data manage to open their infrastructure databases while others do not want to or struggle to do so. Finally, several recommendations are presented to make maintenance data available in the construction industry for a more transparent society and to highlight the benefits and values of making infrastructure data available while ensuring privacy and security for government agencies and companies, particularly the society.

1.3 METHODOLOGY

This deliverable is based on desk research focused on open infrastructure data, data justice and the risks and concerns about opening infrastructure. It is supplemented by nine semi-structured interviews. So open data practices for public infrastructure were assessed with three main resources: (1) an extensive literature and legal review regarding data availability and its theoretical underpinnings; (2) Four interviews with project managers and experts involved in the ASHVIN maintenance projects and five interviews with non-ASHVIN partners to see the general perspective on public infrastructure data in the construction industry; (3) formal consultations with data stewards and privacy officers to advise on the ASHVIN data policy. Since some of the interviewees wanted their names to remain confidential, these people were anonymised while quotes from the interviews are included in this deliverable.

Pseudonym	Position	Organization
Gary	Associate Professor	University
Fernando	Senior Researcher	Research Centre
Elena	Managing Director	Consulting Company
David	Civil Engineer	Construction Company
Simon	Director	Municipality
Jan	Manager Design	Construction and Development Company
Thomas	Team Lead and BIM Manager	Construction and Design Company
Matt	CFO	Construction and Development Company
John	CEO	Construction Design Company

Table 1 Overview of Interview Respondents

This deliverable also draws from and is closely related to D1.5 "Safety and Privacy for Digital Twins in the Construction Industry"[2] and D9.3 "Data Management Plan"[3]. Also, some arguments are based on the data provided in D7.1 "ASHVIN Technology Demonstration Plan."[4]

2 DATA EQUITY AND FAIRNESS AS ETHICAL DIRECTIONS FOR A DATAFIED BUSINESS ECOSYSTEM

In the last few years, discussions have focused on how digital twins, algorithms and artificial intelligence create real-time insights and positive effects such as efficiency and profitability in production processes. These discussions extend to the construction

[2] D1.5 https://zenodo.org/records/7220040
[3] D9.3 Not yet published.
[4] D7.1 https://zenodo.org/records/5542985

sector, including the maintenance stage of building life cycles. While maintenance data may seem very specific, it is still a form of datafication, related to discussions about social justice, open science, open data, and more transparent and accessible understandings of data. Datafication consists of the transformation of different events into quantifiable information to gain further insights on processes, and this includes human and social behaviour. The technologies that enable the generation, storage, and distribution of data stem from an influx of companies that enable such practices. These include media, retail and finance sector, technologies in which AI, digital twins, IoT, and machine learning have become major components of the digital transformation process. These changes have included a transition in the construction industry, with digital processes being seen to (have the potential for) positive effects on efficiency, production, and safety, among other outcomes. Although this transformation process progresses more slowly in construction than other sectors, the ASHVIN innovations show that datafication within the construction industry cannot be ignored. Therefore, in this section, after explaining what datafication means, especially in the construction industry, data justice discussions will be applied as a response to the data inequality caused by datafication. Then, the open data pyramid developed to demonstrate the data justice approach of the ASHVIN project will be presented.

2.1 Datafication and the broader impact of infrastructure data

The digital transformation in the construction industry not only provides many benefits, from increasing the efficiency of the sector, to creating safer and healthier working environments, to eliminating potential risks, it also creates an important amount of precision data. For example, it provides the means for determining the active and passive working times of employees through sensor data. Considerations about how to use different resources more efficiently according to these times and determining infrastructure safety and status with drones or other devices become possible in real-time or near real-time. One of the advantages of this data is that it also provides resources for building information modelling (BIM) systems to allow for more efficient planning of deliveries, on-site processes, and subsequent maintenance.

With the help of digital twin technology, a significant amount of data can be obtained in relation to the building life cycle. This data can show the various stages of a building project, from the moment it is designed to its handover to a maintenance organisation and potentially even to its demolition or renovation. Data generated through digital twin technology during this lifecycle can be collected and stored, managed, and reused. Making this data open can provide significant convenience and benefits to stakeholders in the subsequent maintenance processes in a sector where datafication has started to increase rapidly. With the real-time monitoring and tracking opportunities provided by these technologies, companies have increasingly turned into data-driven enterprises. Datafication process helps companies to create value for their business by understanding the requirements of industry, allowing for increased risk management, and building long term and profitable business relationships.

The exponential rise of digital data technologies has expanded datafication worldwide, along with the globalisation of data analytics; science, governance, business, and civil society are deeply involved in the ‘datafication of everything’ (Mayer-Schönberger & Cukier, 2013). Mayer-Schoenberger and Kenneth Cukier define the concept of datafication as the transformation of social actions into numerical data that enables real-time monitoring and predictive analysis (2013, p.30). According to them, datafication means much more than converting symbolic materials, or more generally analogue materials, into digital form. Following Mayer-Schoenberger and Kenneth Cukier, Mejias and Couldry (2020) define datafication as a combination of "the transformation of human life into data through quantification processes, and the generation of different kinds of value from data". From manufacturing industry to construction industry, datafication is becoming an important element to structure the workflow and improve efficiency and safety. As seen in Figure 1,[5] through creating a digital twin of the bridge for the highway network during the operation in ASHVIN demo site #7, datafication enables multiple ways to better plan interventions on bridges in highway networks in Catalonia, Spain. Maintenance data does not relate only to the infrastructure itself, but the need and type of maintenance also concerns how these infrastructures are used and the social life that surrounds them. Maintenance of a highway bridge will affect the thousands of individuals who use it daily, which in turn may be funding the maintenance through their taxes or tollgate payments.

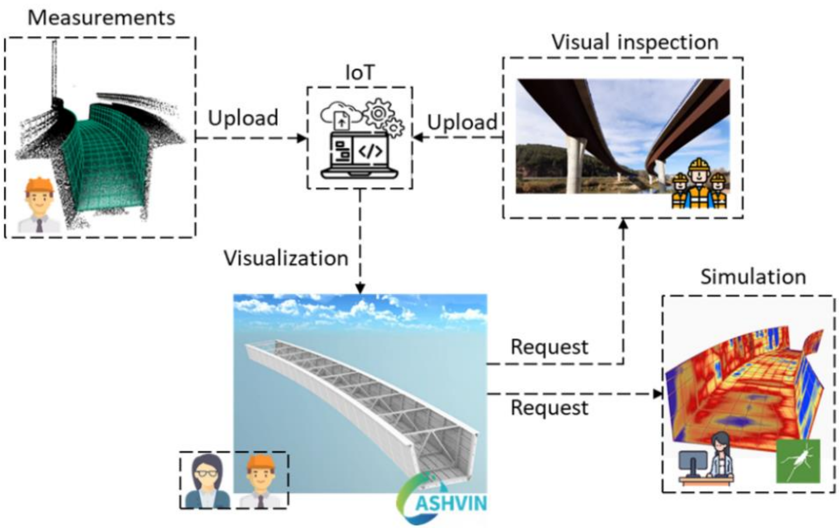

Figure 1 ASHVIN Demonstration site #7 Datafication Process

Another example of the datafication processes of ASHVIN demo sites is the maintenance of Zadar airport runways. On this demo site, ASHVIN tools support the maintenance of airport runways, especially with preventive models. A digital twin model of assets containing information about the runway structure and buried materials was created in the maintenance of the airport. This supports visual inspection and damage assessment procedures.

---

[5] For further detail, see https://www.ashvin.eu/2023/09/11/revolutionising-infrastructure-maintenance-the-ashvin-projects-digital-twin-technology/

Figure 2 Datafication of Zadar Airport Maintenance

Datafication then transforms contexts and spaces into ones that can be easily evaluated. What is unclear is whether opening these data to a public review would directly generate real value or not, in line with a broader concern about "how, or even whether, open data repositories generate any significant value" (Jetzek et al, 2019, p.702). In addition to creating new business models, datafication also transforms the relationships of public undertakings and private companies with citizens. Making data more open exposes the need for a better, more transparent management process of the public sector. In other words, considering the construction industry specifically, making infrastructure data open not only improves infrastructure planning and maintenance and supports the development of infrastructure-related services, but also enables citizens, businesses, and governments to participate and make informed decisions. Inclusive decision making strengthened by making data open emerges as an important factor to achieve more data justice environments. Open infrastructure data provides a foundation for data justice by promoting transparency, citizen participation, unbiased decision-making, and accountability in infrastructure development and management. It helps distribute infrastructure resources more equitably among all stakeholders.

2.2 Data justice as a response to datafication

As indicated above, the process of datafication involving digital twins concerns not only the actual infrastructure, but also the social, economic, and cultural contexts that surround it. Therefore, in this report, the concept of data justice is applied as the starting point that will provide an ethical and fair economic, social, and political approach to opening public infrastructure data. Thus, while this report recommends making data open for the public good and a better governance plan, it also does not ignore the importance of a fair data ecosystem and the socio-political implications of its maintenance. To contextualise the analysis on open data equity, we make use of Linet Taylor's concept of data justice.

Data justice is a solution to eliminate data inequality between different groups and stakeholders. By revealing the basic dimensions of this inequality, data asymmetries are eliminated with data governance strategies and a stronger representation feature is given to the data. This report distinguishes three groups in open data equity in a datafied business ecosystem: (1) organisations, (2) technology, (3) political and social mobilisations towards data technologies. Organisational issues would include data governance, whereas technical issues would involve 'justice-by-design', technology must be technically designed considering fair or justice guidelines. In the ASHVIN project`s approach to open data, all three groups are considered. Organisations directly involved in maintenance data management contribute to and abide by ASHVIN's data management plan (D9.3), technology development on the ASHVIN system follows both privacy-by-design principles and enhances interoperability, and its standardisation and policy efforts address the political also social dimensions of open data. Examples of how the framework for data equality is implemented on the ASHVIN maintenance demo sites are included in the next section of this deliverable.

Linet Taylor (2017) presents the concept of data justice as a framework for data governance, which can be applied to each of the three groups mentioned above. She introduces a three-pronged framework for data justice that can underpin data governance to identify ethical pathways through datafication: (in)visibility, (dis)engagement with technology; and anti-discrimination.

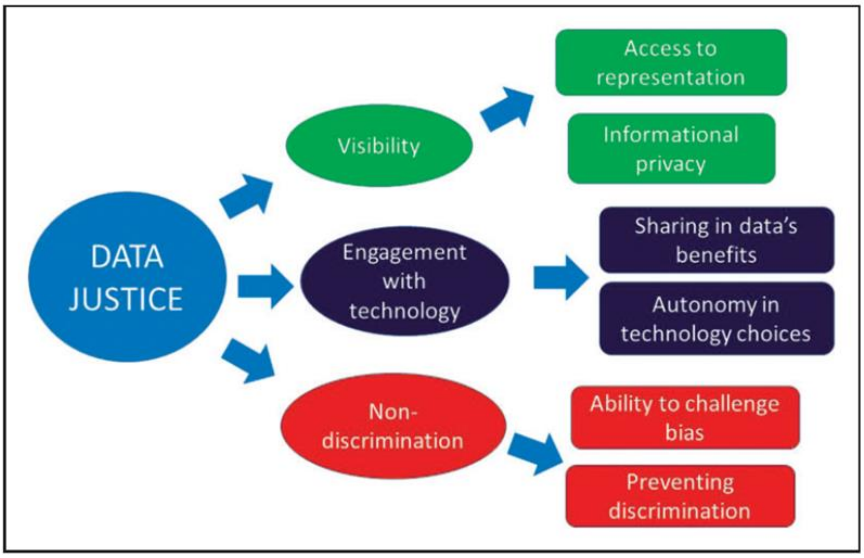

Figure 3 Three pillars of data justice (Taylor, 2017, p.9)

Figure 3 also displays how open maintenance data contributes to data justice by promoting transparency, equity, and accountability at different levels. The first pillar addresses not only visibility and representation but also privacy concerns. It covers the discussion of how much data can be treated as public domain. Making maintenance data public increases data justice by enhancing visibility. Visibility empowers especially stakeholders of maintenance sites about ongoing projects. This also brings the engagement and participation of citizens as a fundamental aspect. For instance, if available maintenance data only concerns a few neighbourhoods of a city, this can cause other areas, and its residents, to be invisible and underrepresented in decision making processed that include renovation and investment decisions.

The second pillar, engagement with technology, is an essential component of any data justice framework. This pillar underpins the power over one's visibility. If technology is not aligned with justice principles, open data practices end up benefiting a limited and privileged group of people. For instance, if open data lacks open file formats, this will introduce vendor locking and ultimately restrict the number of companies that can benefit from data. Ensuring that IoT platforms, such as the ASHVIN platform, are also open source is one way to counter these harmful possibilities.

The third pillar focuses on discrimination consisting of two dimensions: `the power to identify and challenge bias in data use, and the freedom not to be discriminated against' (Taylor, 2017, p.9). In a context in which data has exponentially increased and become a principal resource in contemporary society, data asymmetry inevitably emerges. Although Johnson (2014) thinks that open data exacerbates this asymmetry and inequality, he still presents open data as a solution to the problem of inequality by linking information justice. Following this idea, this text focuses on open data as playing a key role in minimising asymmetries, a tool that can bring public interests to the fore.

In principle, bias in infrastructure data may not be readily apparent. After all, how can sensor data be biased in the same way as financial data, for instance? In response, consider something as simple as the same type of infrastructure having different maintenance frequencies in different neighbourhoods of a city based on socioeconomic considerations. Wealthier neighbourhoods may see more frequent maintenance than less wealthy ones. This then begins to make clear how certain groups can benefit or be harmed by maintenance decisions. Opening data about this maintenance could be a first step to correct these biases.

In short, open infrastructure data makes it possible to provide more data equity to ensure fair and equitable distribution of opportunities and risks associated with data-driven applications and technologies. The ASHVIN project aims to ensure that all stakeholders, including public organisations, oversight boards, companies, citizen groups and citizens themselves have equal access and participation in the data ecosystem. The ASHVIN project seeks to minimise harm to both privacy and data security and promote ethical and inclusive data practices.

2.3 The ASHVIN open data pyramid

This section focuses on how to create a fair environment when making data openly accessible. Underlying these discussions is the assumption that not all data can or should be treated equally in relation to its commercial, social, and cultural context. For instance, making certain trade secrets public would have a disproportionately negative effect on the company that own this information. As such, differentiating between different levels and types of data is a key component of the data justice approach proposed by the ASHVIN project. To clarify these discussions, a four-level pyramid serves to visualise how to incorporate datafication and its good practices through data transparency, accessibility, and agency. In this way, the ASHVIN open data pyramid distinguishes between four types of data and the associated levels of disclosure: meta data, raw data, report data, confidential data.

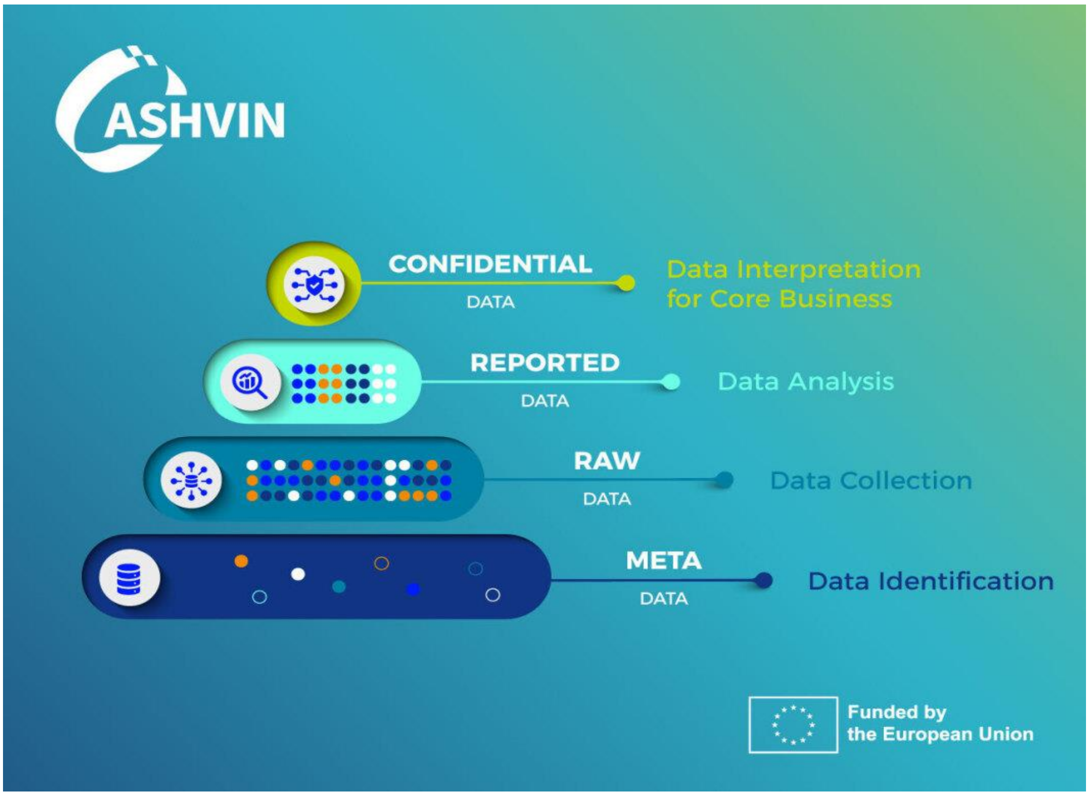

Figure 4 ASHVIN open data pyramid

Metadata corresponds to the information about a piece of data, i.e., time, location, typology, etc., of the data in question. In the scope of a maintenance project, this translates to information about when, how, and where maintenance data was collected. This baseline level is, in principle, the least concerning or controversial in terms of its public dissemination. Given the public funding of infrastructure projects, it is expected that the public should be aware of the fact that maintenance is being conducted, as well as tender procedures for public projects. In the ASHVIN project itself, this level of metadata dissemination can often be observed in articles or other communication materials about activities in the demonstration sites. This level of data does not contain the raw data, but rather informs the reader about the collection of data in the demonstration sites.

Raw data corresponds to the information collected – thus, all raw data has some meta data indexed in it unless it is explicitly removed. This can be sensor data, raw image footage, audio recordings, or whatever information that is collected directly on the infrastructure. This type of data requires further care in what can be shared, given that it might include personal data (e.g., an individual's face captured in the video footage), or other types of sensitive or confidential information (e.g., civil infrastructure tied to military infrastructure). By default, this data should be made accessible on request or, if privacy concerns are either addressed or absent in the particular data, publicly shared for reuse in an accessible data repository.

Report data is information extracted by contextualising and processing the dataset; it could be considered the first level of data interpretation. This includes for instance the outcomes and interpretation of machine learning models that are applied to the raw data, or conclusions drawn by analysts based on an interpretation of the data (e.g., that the values of tension for a bridge exceed the safety thresholds). Further caution needs to be exercised when sharing this kind of data, given that it may include proprietary analysis processes or conclusions and results may be misinterpreted. For instance, it may give away vulnerable points in key infrastructure that could be exploited by attackers. Thus, report data should be redacted and made more appropriately accessible before sharing, so that the positive outcomes of transparency are achieved while minimising risks of potential negative circumstances for the public or the report authors.

Finally, **confidential data** refers to a second level of interpretation and level of decision; it could be considered the interpretation and discussion of data to address diverse decisions and solutions. It may often contain sensitive business information or refers to processes that may be protected by intellectual property or confidentiality agreements.

In summary, we highlight there some key concerns for understanding data equity in a context of open infrastructure data:

- Discussions of open data and **datafication** invariably **involve other discussions about social justice and open science**.
- Open data carries risks and requires the minimisation of these risks by setting **thresholds of access** over the various data collected.
- Four levels and thresholds of agency and access defined by the ASHVIN project are: **meta data, raw data, report data and confidential data**.

3 OPEN (PUBLIC) DATA MANAGEMENT

In today's digital environment, the power of open data to promote transparency, collaboration and innovation increases day by day, however this power has different importance for every sector. Each industry has various strategies in its own circumstances to make its data available to the public and its stakeholders. Therefore, in this section, it will be explained what open data means for the construction industry and what the benefits and challenges of making infrastructure data open are.

3.1 IDENTIFYING OPEN DATA

Open data is one of the most discussed issues on the European agenda. In recent years, interest in open data has increased as it plays an important role in 'improving government accountability and service delivery, empowering citizens engagements and participation to make better decisions in public administration, creating economic opportunities and solving major public problems' (Verhulst et al., 2020, p.4). Janssen (2012) underlines that there are four main drivers behind governments' disclosure of their data. *First*, public data that is vital for transparency and accountability. It is important for a transparent government that citizens know what their governments are doing, but Janssen does not ignore the fact that transparency is possible not only by accessing this data, but also by being able to use and disseminate it without any restrictions. *Second* is the increasing role of participatory governance. Open data becomes important so that citizens can participate in the decision-making process. *Third*, open government data is a catalyst for innovation and economic growth. These data are important for the development of new applications and services and the creation of new business models. *Fourth* and finally, public data facilitates internal functioning within the public sector. Janssen draws attention to the fact that this will increase the efficiency of public services (2012).

One of the first examples of open data initiatives is the Open Data Institute, which was founded in 2012 by Tim Berners-Lee and Nigel Shadbolt. The purpose of the Open Data Institute is to create a secure ecosystem by including all stakeholders of the open data ecosystem in the process and ensuring their interoperability. Some current active initiatives such as The Offical Portal for European Data[6], Data Justice Lab[7], UNI Global Union[8] and Decidim[9], embrace the innovations of datafication to enhance citizen participation. Within the ASHVIN project, we transpose this idea to the domain of maintenance data. While it does not suggest that citizens will start to conduct bridge load tests themselves, they can play a role in interpreting or complementing load test data, especially if they possess the expert knowledge to do so.

---

[6] https://data.europa.eu/en
[7] https://datajusticelab.org/
[8] https://uniglobalunion.org/
[9] https://decidim.org/

The European Commission has noted that the re-use of government data is of fundamental importance to citizens, businesses and society, and governments themselves: (1) data can drive growth and therefore the creation of new jobs; (2) evidence-based policy-making and efficient administrative framework to provide a better public service. Making data available in the public sector leads to a reduction in reuse fees. The EU Commission points out that one of the best examples of this is the Danish Enterprise and Construction Authority. In this case, the reuse of open data provides a 10-fold growth in the market. (European Commission, 2011).

However, the benefits outlined above often encounter practical difficulties, starting with the definition of the concept of open data itself. The Open Knowledge Foundation (2012) defined open data as a data that `can be freely used, reused, and shared by everyone`. According to this definition, open data is built on three fundamental aspects: (1) availability and access; (2) reuse and distribution and (3) universal participation. This last point, universal participation, suggests that everyone should use the data without any restrictions (Open Knowledge Foundation, 2012, p.6). While the concept of open data directly relates to the right of availability of information, it contends with the related concept of personal data, which is by the definition tied to a right to privacy. The main difference between open data and personal data is that open data does not directly contain data about an identifiable person. The two terms are often discussed in opposition, but it is important to clarify that they are not mutually exclusive, as will be explained in section 4.

Figure 5 The concentric shell model of Backx (Donker and Leonen, 2016, p.235).

In contrast, transparency and open data are often used together or even interchangeably. The concepts of transparency, accountability, and open data became more popular when Barack Obama, on his first day in office as president of the United States in January 2009, announced that he would launch a transparency strategy that would mean an unprecedented level of openness in government. Both concepts are in line with the same basic goal: `improve participatory and governance and increase government accountability` (Cahlikova and Mabillard, 2020, p.664). Cahlikova and Mabillard present three key propositions where transparency and open data differ slightly: *First*, transparency can be seen as a prerequisite, a goal, or a potential consequence of open data. *Second*, they say that open data is a broader concept, and that transparency is an objective of open data. *Finally*, disclosure of information may bring about transparency, but it does not guarantee it (2020, p.665).

Just because data is accessible does not mean it is defined as ‘open data.’ This data must also be usable and re-shareable. In addition, it must be intelligible by both humans and machines and comply with the principle of interoperability. So, the complete openness of data relies on its economic, legal, technical, and political availability. A basic summary of all these definitions and classifications of open data is clearly seen in Backx's concentric shell model (Figure 5).

Donker and Leonen point out that this model of Backx is a guide to the steps a user should take to evaluate whether the open data is suitable for their needs. According to this model, data can be defined primarily by users, and it should be known where to get it. Then the data should be accessible to users under all circumstances. Finally, the data should be of appropriate quality for users' intended use (Donker & Leonen, 2016, p.289). According to them, open data should answer these questions to understand whether the layers meet the requirements: “Are the data identifiable and where can data be obtained? Can the user obtain the data, and under what conditions? Can the user assess the quality of the data?” (2016, p.288) In addition to these questions it is necessary to determine which data will be open to whom, a matter that was addressed by the ASHVIN open data pyramid above. The answers to all these provide a summary of the limitations, pitfalls, and promises of open data. At this point, evaluating the advantages and disadvantages of publicly available data on a sectoral basis will enable this process to be concluded more effectively.

3.2 THE PROMISES AND PITFALLS OF OPEN PUBLIC INFRASTRUCTURE DATA

One of the sectors where open data has become a pressing agenda point is the construction sector. Recording and storing data from the design process to the maintenance process and sharing it with the relevant stakeholders will positively affect the success and profitability of the sector. Of relevance is the maintenance process of a building in which there is an unnecessary loss of time and cost when no previous data is available. When this sort of structural data is made public, or at least accessible on request, risks can be better estimated, and maintenances processes can be better estimated and managed faster and cheaper.

The City of Rotterdam platform, which bears similarities to the ASHVIN digital twin platform, can be used to demonstrate the benefits of open infrastructure data. This open urban platform aims at data management, open financing and citizen trust and participation by developing a digital twin of the physical city. One of the representatives of this platform underlines the importance of making the data open in dozens of projects in the city of Rotterdam.

> For example, a resident would like to have a rubbish bin, basketball court or a tree placed somewhere. The digital reality immediately shows whether these fit. For example, because the location is an approach route for emergency services, or because there are pipes running somewhere. When this data is shared with other stakeholders, it provides us with speed and easy insight as well as transparency (Gary, Associate Professor at University).

Open data is like a double-edged sword. It carries both positive and negative features at the same time. Open data encourages potential collaboration and consultation among stakeholders, from site managers to architects, contractors, and policymakers. This is because, with making maintenance data publicly available, the project’s timeline and details are more widely discussed to ensure that the safety and security of the infrastructure is preserved. Making infrastructure data available to all stakeholders in the construction industry means that gaining access to data that was previously unavailable, enabling new approaches to maintenance.

Many of the benefits mentioned above regarding governments can also apply to the construction sector itself, from the design process to the maintenance process. Some benefits of open infrastructure data include:

- Improving transparency and accountability
- Increasing innovation
- Better decision-making
- Cost savings
- Encouraging public engagement and retaining digital trust
- Preventing monopolisation and reducing corruption

These benefits are more prominent for public institutions, but they are also valid for private companies. As Deloitte's 2012 report shows, a successful open data ecosystem has three key components: government, business, and citizens.

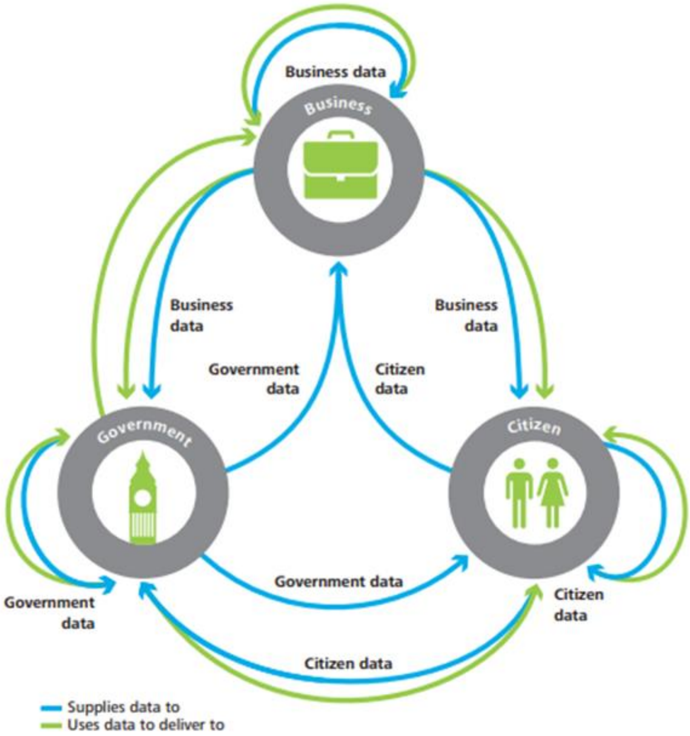

Figure 6 Open Data Ecosystem (Deloitte Analytics, 2012, p.9)

As seen in Figure 6, open data is not limited to government data, given that the private sector is one of the key stakeholders of this ecosystem. Although companies and organisations attribute their reluctance to data sharing to security and privacy concerns, they also often do not want to lose commercial competitive advantages. Verhulst and others point out that this unwillingness “may simply be perpetuating its own monopolistic ownership and, in the process, exacerbating existing inequalities” (2019, p.16). They underline that tackling these problems requires private individuals or teams in the private sector who can review and implement opportunities to unlock the public value of a company's data. According to them “data stewards need to be established and empowered across the private sector to seek new ways to create public value through cross-sector collaboration” (Verhulst et.al., 2019, p.47). This relates to the hierarchical availability of data outlined in the ASHVIN open data pyramid, but also on the ADAPT principles described in Section 6.

The discussions above show that a critical point is the good governance of open data. Although the concept of open data has been the subject of debate, most government agencies have difficulty in implementing it on the operational level. In many cases, open data governance may be well organised on strategic level, but it is not reflected in operational level. Donker and van Leonen underline that highly valued data are kept and are not made available due to `the lack knowledge on how to adapt sensitive data suitable for open data publication` (2016, p.300).

The results of a survey conducted by MIT in 2020 confirm this situation. This survey conducted with experts having knowledge of or experience of AI over 1000 executives across 11 sectors. Survey respondents perceive open data as a useful tool for generating value and opportunities. When asked what the greatest benefits of sharing data with companies in their own or adjacent industries would be, they highlight greater speed and visibility, innovative product development and efficient and new models are the major benefits of open data (see Figure 7).

Figure 7 Benefits of Open Data

These shows that private companies are aware of the benefits of sharing data with other companies in their own or adjacent industries. However, most private sector organisations are not willing to make their data public. They consider what type of data sharing and cooperation with other stakeholders is inappropriate and not legally safe to disclose their commercial secrets. The MIT survey confirms this situation, 64 percent of company executives are reluctant to adopt open data policies due to uncertainty in regulations (MIT Technology Review Insight, 2020). Willingness to share data also varies by industry, as shown by Figure 8.[10] Despite sectorial differences, on average, the majority is willing to share the data. However, notably, the construction sector in not included in this survey, illustrating its late arrival at discussions regarding open data.

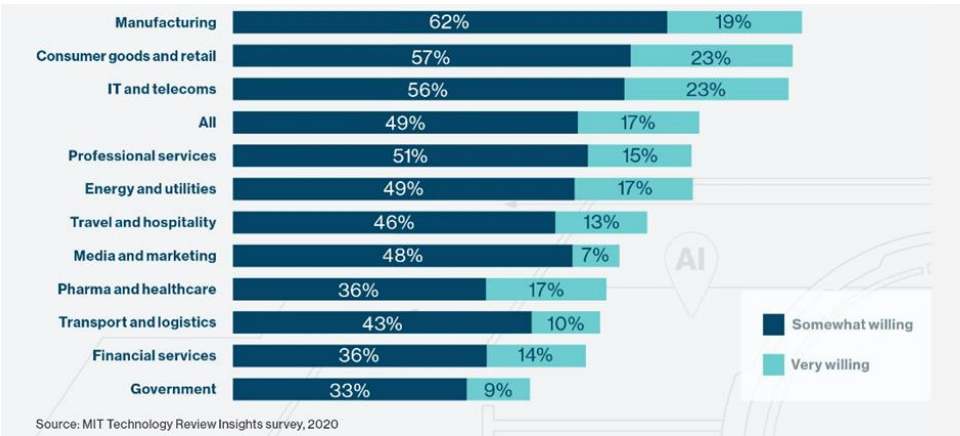

Figure 8 Distribution of sectors' willingness to share their data with other stakeholders

---

[10] See MIT Technology Review Insight, 2020.

As discussed above, willingness to share data does not always translate to open data practices. In the construction industry, the current situation regarding the sharing of data with citizens, other public institutions and private companies seems to be less open. The answer given by one of the interviewees embodies the reluctance to make infrastructure and maintenance data available:

> This has to do with what the result is of having open data by default. It's different in all bridges because if I say ok, I can upload it… It's like a second of my time, but there's no need for such a thing. What is the use of knowing these data? (Gary, Associate Professor at university).

Based on the interviews, it can be said that whether they are private or public organisations, the reluctance to make data public is dominant even some part of industry is aware of the benefits and values that data sharing. One of the non-ASHVIN interviewees pointed out:

> In fact, making BIM data available to the public can be incredibly beneficial, considering the profits and benefits to be gained. Especially in this process, it is exciting to think of the partnerships that will be created by the interaction with other companies. But the reality is different. While even municipalities are reluctant to share their projects with the public or with us, this is almost impossible for private companies (John, CEO at Construction Design Company).

A real estate trading platform, zillow.com, illustrates the economic benefits of open data and showcases how private companies generate public value through open data. This website keeps historical data such as tax, sales, rent, mortgage, geographical data of more than 110 million houses in the USA in its database. Using this database and some algorithms, the current rental or sale values of the houses are estimated (Jetzek et al., 2019, p.703). The existence of such a platform contributes to price sustainability by minimising the speculation in property value.

Similarly, if the necessary environment is provided for the design, construction, and maintenance data on the ASHVIN platform to be opened securely, the infrastructure database will be enriched over time with the characteristic features (price, required labour, risk value, materials required for maintenance, sensor metadata, and required actions) of structures such as bridges, buildings, and roads to be built. This database would make a significant contribution to new and ongoing projects' risk predictability. Thus, the costs of both newly constructed buildings and maintenance projects will be reduced and their efficiency will be increased. In addition to these benefits, open infrastructure data also enables stakeholders to work more effectively together if data overload is avoided. One of the interviewees describes this situation in this way:

> Open maintenance data would be beneficial. We step in a project, and we must take over part of the infrastructure. This was transferred from a contractor to subcontractor. We have lack of data of our infrastructure. It

> would be very helpful if we have the historical maintenance data. Takes half year or a quarter year completely check the infrastructure and find all errors deal with the clients. If we have data, we just scan it and look the deviations. We don`t know what the real maintenance interval is and are really missing real data of maintenance construction. (Jan, Manager Design at Construction and Design Company)

Some companies use the openBIM (Open Building Information) program to prevent the situation described above. This programme, an initiative of BuildingSMART, provides fast data sharing, especially in multi-stakeholder projects and using open data formats. It also provides a common language for information exchange at every stage, from design phases to implementation of construction and operational phases. This software facilitates the connection and interoperability of different stakeholders.

Based on theoretical insights and the findings from interviews, the benefits of open data can be attributed to three main actors: public institutions, business, and citizens. All these generate economic and social benefits from the open public data. Public institutions implement “good governance strategies”; citizens get transparent and better services and products; business increases its share and visibility in its own industry and general economy by achieving efficiency and sustainable value.

Sharing the data on public domain structures enables a more effective management planning, especially by making real-time data available. In the European directive on open data and the reuse of public sector information (2019), real-time data comprises “documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors”. According to this directive, open real time data is should be made by public organisations and undertakings for reuse. Considering the construction industry in particular, making the data of buildings constructed or maintained using digital twin technology of infrastructures such as bridges, roads, footbridges open in real time could be an essential resource for public sector bodies and public undertakings to create an efficient governance strategy, while providing a sustainable economic value to companies as well as providing a more secure environment to citizens.

Open data is also used for urban regeneration goals such as decarbonisation, sustainability, and value creation. Municipalities that run AI- supported digital cities projects see open public data as the main tool of the transformation process to gain the public value. In addition, open public data enables citizens to be involved in the evidence-based local government process. For example, it is very important for citizens to see where the taxes they pay are used and whether this use is used effectively. In particular, engagement of citizens in the data storage process also provides one of the requirements of democratic society. Opening this data to other companies may not only prevent the monopoly of a single company, but also saves cost and time by providing accessible information on previous works. For example, in a road maintenance carried out before, knowing where the water pipes pass through, which ones have changed, etc. not only provides benefits in terms of time and cost, but also provides an effective planning that will eliminate possible risks.

While the positive features of open data are often emphasised, its challenges and negative aspects should not be ignored. A major concern is related to privacy and security. The construction industry handles data that is as fragile and sensible as personal data; in this case, public infrastructure data can be utilised for a range of malicious purposes, from theft or breach of intellectual property or financial information to personal records. One of the ASHVIN demo site representatives pointed out that:

> Data analysis method can be shared but there are some restrictions related to Intellectual Property Rights and consortium agreements. We should ensure that any limitations or constraints regarding disclosure of sensitive data or information are considered (Fernando, Senior Researcher at Research Centre).

This response supports Donker and Leonen's debate over ownership and control of open data. As they point out, open data must have an owner who will encourage and coordinate open data activities (2016, p.296). Improper control of data causes various privacy and security risks. Therefore, determining by whom the data will be managed can also eliminate the risks in the shared data.

As illustrated by the ASHVIN open data pyramid, data to be shared needs to be well defined. Which data will be kept confidential is important both in terms of privacy and in terms of trade secrets. Sometimes shared data inherently brings with it data that must be kept confidential. This situation, which is conceptualised as the Mosaic effect, causes various risks. While the shared data does not cause any privacy problems or security gaps on its own, it creates risks and problems when combined with other information. Therefore, it is crucial to standardise which information is needed at which stage and how it should be shared.

The other negative aspect is insufficient quality of data. It is essential for the parties that the data will be shared that the data is accurate and consistent. As it is discussed in the next section, standardisation plays a vital role here. Data that is out of date, that is, poor quality or inconsistent, will undermine the credibility of open data, and can lead to undesirable outcomes. One respondent explains the importance of this situation:

> Which data you have, what you can do with this data… We have a lot of data and most of time we don`t know we have the data. Having information is not enough. It's also important to know what you have. And this data must be properly recorded and shared (Thomas, Team Lead and BIM Manager at Construction and Design Company).

Considering the construction industry, this would also have potentially harmful consequences for the maintenance or construction of structures themselves. The quote above also hints at another potential issue, data overload. If stakeholders are overwhelmed with too much data, its effective use is compromised.

In general terms, the main reason hindering open maintenance data relates to privacy and security gaps. As Johnson underlines “whether public or private, open data generally consists of a commitment to make data available publicly in non-proprietary, machine-readable formats at the lowest level of granularity possible” (2014, p.264). In relation to this goal, it is necessary and important to evaluate the current standards and regulations governing open data.

In summary, the key issues related to open public data management include:

- Data that is accessible does not necessarily mean it can be seen as 'open data'. This data must also be **usable** and **re-shareable**.
- Open data encourages **collaboration** and consultation among stakeholders, improves **transparency** and **accountability**, provides better decision making and cost savings, encourages public engagement and retaining **digital trust**, prevents monopolisation and reduces corruption.
- While it has positive features, open data can also cause the violation of some personal data and privacy. For this reason, open data must be **governed well in accordance with agreed upon standards.**

4 DATA PROTECTION, SECURITY AND STANDARDISATION OF MAINTENANCE DATA

Various data breaches and vulnerabilities in recent years have led to a greater focus on data security. These threats to personal and corporate data have increased with the spread of information technologies. The ASHVIN project embraces open data policy as one of the strategies to address these threats. Open data introduces transparency and accountability, which are fundamental components of governance that can be harnessed to address security threats. Strategies regarding data security are also guided by data privacy laws and standards.

As emphasised throughout deliverable, one of the important elements of open data governance and data protection process is data security. Data security is a set of measures and strategies that ensure the protection of data, especially in the digital environment. These measures include technological applications such as encryption, authentication systems, and network firewalls. To reduce data breaches and mitigate damages to the privacy of governments, companies, and individuals, it is necessary to implement data protection practices, especially regarding open infrastructure data.

It is difficult to build public trust and security assurance without a corporate and government commitment to security transparency. It is one of the key components of assuring security, and as the fundamental foundation of the ASHVIN open data pyramid, it aims to be a core value for public undertakings and companies in the construction industry as well.

However, security regarding data cannot be considered only in regard to protection of the data itself, but also raises issues concerning how data openness or data breaches can compromise the security and safety of the infrastructure that they refer to. For example, if any weak points of critical infrastructure are revealed by maintenance data, malicious agents may use that information to attack and compromise that very same infrastructure. This requires proportionality in data sharing, which minimises critical risks tied to data accessibility, but also strategic consideration of the goals, benefits, and harms of open data. As we saw when we approached the concept of datafication, infrastructure cannot be isolated from its social and geopolitical contexts, which should also be considered when determining benefits and risks.

As can be seen in the ASHVIN Open Data Pyramid given in Section 2, as the datafication process becomes more prominent, precautions to open the data increase and the importance of data security increases in direct proportion to this. In other words, as the layers go up in the pyramid, transparency decreases, while the density of the data that needs to be more protected increases and in connection with this, the need for greater security emerges. Transparency supports a more secure data environment, but the critical point here is the correct implementation of transparency, that is, a good balance of transparency and security must be established. Therefore, the standardisation of the policies to be applied while making the data open may inform security strategies. In practice, organisations should determine which data will be disclosed by whom and who has the right to access these data.

The General Data Protection Regulation (GDPR)[11] lays the foundations in Europe for these considerations as it requires all organisations that process personal data of individuals located in the EU to take the necessary steps to protect the confidentiality, integrity, and availability of said personal data. As will be discussed in Deliverable 6.5 “Best practices for digital twin-based privacy”, the GDPR covers quite a wide range of concerns related to personal data by anticipating how to manage not only its general use and purpose but also unauthorised access, disclosure, alteration, and destruction. In the context of maintenance data, which is available for public domain, the entities handling it ought to take the necessary steps to protect the available data in accordance with the GDPR. This may include implementing secure authentication methods, limiting access by way of contextual integrity, and encrypting the data.

The GDPR focuses on personal data and how it ought to be collected, processed, and stored, however it does not discuss in an explicit manner of open data nor its regulations. It does not legislate on how it shall be processed or stored or what should be made publicly available; rather it specifies that personal data contained in open data which will become public domain must be carefully anonymised or pseudonymised so that individuals cannot be identified. While maintenance data collected in the scope of the ASHVIN solution does not contain personal data, this concern should nonetheless be considered when opening maintenance data to the public domain. Additionally, in case organisations publish open data which contains personal data, said entities shall are expected to present a privacy notice that thoroughly explains how the data is being used and how individuals can contest, comply, or question said purposes under the GDPR.

However, the GDPR sets out exemptions that would allow organisations to publish open data that includes personal data without anonymising or pseudonymisation it. The first case considers whether publishing said data benefits the public interest; this is possible if doing so is essential to protect public safety. The other case that allows for an exemption concerns scientific research; in this case, personal data may be published if it is necessary to advance a crucial study or analyse events important to maintain public safety. Besides these exceptional circumstances, organisations ought to comply to all other GDPR requirements, including obtaining explicit consent for the use of individuals’ personal data before collecting it, prioritising privacy by design, and conduct data protection impact assessments (DPIA).

In addition to the GDPR, the International Organisation for Standardisation (ISO) provides additional standards which organisations ought to account for when considering making maintenance data publicly accessible. Specifically, ISO/IEC 27001:2013[12] and ISO/IEC 27701:2019[13] are two international standards designed to enhance existing Information Security Management Systems (ISMS) – or, in other words, sets of policies and procedures that manage and protect an organisation’s information assets – and Privacy Information Management Systems (PIMS), which are

[11] https://eur-lex.europa.eu/eli/reg/2016/679/oj
[12] https://www.iso.org/standard/27001
[13] https://www.iso.org/standard/27001

specifically designed to protect individuals' privacy. ISO/IEC 27001:2022 is a generic standard that can dictate how ISMS's are governed regardless of the organisations' size or industry; ISO/IEC 27701:2019 instead is an extension of the aforementioned standard that addresses personal data protection processes. These standards differ from the GDPR as they roll out on a plan-do-check-act (PDCA) basis, which allows for a cyclical and continuous approach when taking action to improve said systems. These regulations are also commonly recognised as best practices for information security and privacy management as they reinforce the compliance with other regulations, such as the GDPR. Additionally, the implementation and use of ISMS and PIMS as well as the PDCA cycle do reduce the risk of data breaches and other disruptions as their operational efficiency is continuously reviewed and improved. Deliverable 1.5 "Safety and privacy for digital twins in the construction industry" details the cybersecurity and sociotechnical approach from the ASHVIN solution in relation to data privacy and safety, so this aspect will not be approached in detail here.

One fundamental aspect of making maintenance data open is that this process should not be seen only as a passive process of extending access, but one that requires an active consideration of data protection and how possibilities of reuse can be maximised. The standards mentioned above, while covering aspects regarding data privacy and security, do not establish provisions to the collection and sharing of maintenance data and its specificities. This was a difficulty encountered in the scope of the ASHVIN project, which triggered a push toward the standardisation of data collection and data sharing methods, for instance, regarding drone video and image data.

While standardisation efforts in the scope of the ASHVIN project are mostly the focus of D6.2 "Recommendations and options for future standardisation for Digital Building Twins at a European scale" (to be published in March 2024), it is important to note that in standardisation workshops the re-use of data was one of the main gaps identified by participants that could be resolved through a standard. Standardisation could also enhance the potential for contributions from citizens and other societal actors that are not directly responsible for maintenance efforts. For instance, a standardised process for labelling cracks in concrete based on maintenance images could allow citizen scientists to participate in data labelling efforts, or to secure additional training data examples by taking pictures of cracks on roads or walls.

Use of non-proprietary file formats and software is also a way of enhancing reuse of maintenance data. This is particularly important in the scope of digital twin data, given that the real time nature of data collection and data sharing may not give interested parties the opportunity to convert data between formats. For instance, citizens or interested parties such as journalists, scientists, or non-governmental organisations (NGOs) cannot be expected to invest in expensive licences for data visualisation software in order to access maintenance data. As such the open-source nature of the ASHVIN platform and the dashboard are important components of this active component of data sharing.

This section has discussed the relevance of protecting and standardising data that will be made open and highlighted several key points:

- While maintenance data rarely contains **personal information**, it can provide critical information on important infrastructure. This aspect needs to be accounted for in releasing information.
- Open data policies need to be considered in the context of **data protection**, so that open data practices do not introduce vulnerabilities in systems;
- Data sharing is an active process, requiring consideration of how **access** and **reuse of data** can be maximised.

5 THE AVAILABILITY OF OPEN PUBLIC MAINTENANCE DATA IN ASHVIN DEMO SITES

As explained in detail in D7.1 "ASHVIN technology demonstration plan", the demo sites #1, 2, 3, 7 and 9 are in the maintenance phase of the construction process as part of the ASHVIN project. By using the data provided in D7.1, the table below has been generated to focus the maintenance phase related demo sites with their stakeholders, provided data and used tools.

Demo Sites	Stakeholders	Type of IoT Measurement and Sensing Devices Data	Used Tools[14]
#1 **Bridges for highspeed railways in Spain**	* **Renfe:** Public company that represents the Spanish operator of the railway networks. * **Adif:** Administrator of Railway Infrastructures, state-owned company. * **Geosica :** private company highly specialised in diverse activities.	- Deflection at midspan and displacement of supports - Inclination - Acceleration - Environmental conditions (Temperature and Humidity) - Deflections (remote sensing)	MATCHFEM RISA GISI
#2 **Building renovation in Poland**	* **Municipal Buildings and Housing Administration of Gdynia:** public body, owner of the demonstration building.	- Temperature, humidity, CO2, CO, VOC, PM 2.5 and PM 10, pressure - Wall Temperature	4DV-D

[14] For more information, please look at https://www.ashvin.eu/digital-toolkit/

	*** Occupants:** People leaving in the social flats *** Designer:** Person or the company hired by the Housing Administration to perform the design. *** FASADA:** partner of the ASHVIN consortium, responsible for contacts with the Housing administration		MATCHFEM
#3 **Airport runway in Croatia**	*** Owner - Zadar airport:** publicly owned with different percentage between the state, city and municipality * **Zadar airport Ltd**.: managing daily operations and maintenance of all airport areas * **Personnel performing the inspection and maintenance:** the ones hand on performing visual inspection, damage detection, storing data about inspection, making decisions about maintenance. * **Maintenance contractor:** the one performing actual repair and maintenance works	- Images - Environmental data (temperature, humidity, wind)	3DRI (Method) DDCV (Method) RISA GISI

#7 **Bridges in highway network in Spain**	* **The Ministry of Transport, Mobility and Urban Agenda**: The department of the Government of Spain * **BAGH Técnica:** Private company specialised in diverse activities.	- Inclination at key supports and/or bearings - Acceleration along several spans of the bridge - Environmental conditions (Temperature and Humidity) - Cameras for traffic measurement (anonymised) - Cameras for telemetry - Thermocouples	MATCHFEM RISA GISI
#9 **Sport Stadium Roof Structure**	* **City of Munich:** Owner of the stadium * **Stadtwerke München (SWM):** one of Germany's largest municipal utility and service companies. It is owned by the City of Munich. * **Olympiapark München GmbH (OMG):** wholly owned subsidiary of the City of Munich. * **Behnisch Architects:** Architects schlaich bergermann und partner/sbp	- Measurement of deflections of the cable net - Images of the roof cladding (Plexiglas)	MATCHFEM RISA GISI

Table 2 Details of Maintenance Related the ASHVIN Project Demo Sites

This table below shows the relationship between the used tools and the targeted key performance indicators (KPIs). Four demo sites out of five demo sites that are in the maintenance phase in ASHVIN use MatchFEM, RISA, GISI tools. As seen in Figure 9, these tools are used to improve the ‘reduction of cost” KPI value. So, it can be said that the maintenance data used in this project is mainly used for cost reduction.

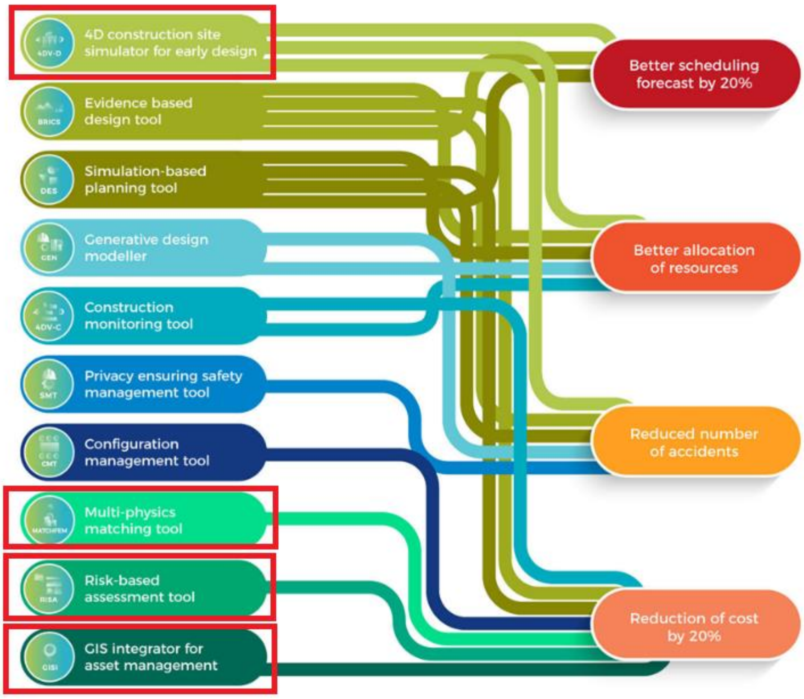

Figure 9 The Relation Between ASHVIN Tools and KPIs (Adopted from D7.1)

Therefore, as mentioned above, some demo site representatives think that sharing this data will harm commercial sensitivity:

> Making public resources available to citizens and other stakeholders can be an essential component of a strong democracy, but the owners of the sites think that the making open of commercial secrets weakens their competitive power in the market (David, Structural Analysis Engineer, Construction Company).

The table above, which details the maintenance-related demo sites in the ASHVIN project also displays that all these demo sites have both public and private stakeholders. This shows that the maintenance data is not owned by only one stakeholder which means that the decision to make the maintenance data public should be taken together with stakeholders on both public and private sector sides. Results from interviews with ASHVIN's partners also support this claim:

> Although the site is publicly owned, all decisions regarding the process are taken together with all stakeholders. Therefore, all stakeholders need to have a common opinion on such a sensitive issue regarding making data public.

> The working logic of the construction industry is based on multiple stakeholders. Even the final decision is made by the public owner, if one of the stakeholders does not want the data to be shared, it becomes much more difficult than you think (David, Structural Analysis Engineer, Construction Company)

More generally, when asked whether maintenance data could be published as open data, several interviewees gave sceptical answers. While some thought that it was not very necessary, others emphasised that it was not easy to manage this process. However, several sites do make maintenance data open.

Comparing the ASHVIN project`s demo sites against the 5-stage data journey developed by Deloitte (see figure 10), these sites can be seen as just at the beginning of the open data journey. Maintenance demo sites are mostly located in phases 1 and 2. Demands and activities in the construction industry regarding the opening of infrastructure maintenance data seem immature yet.

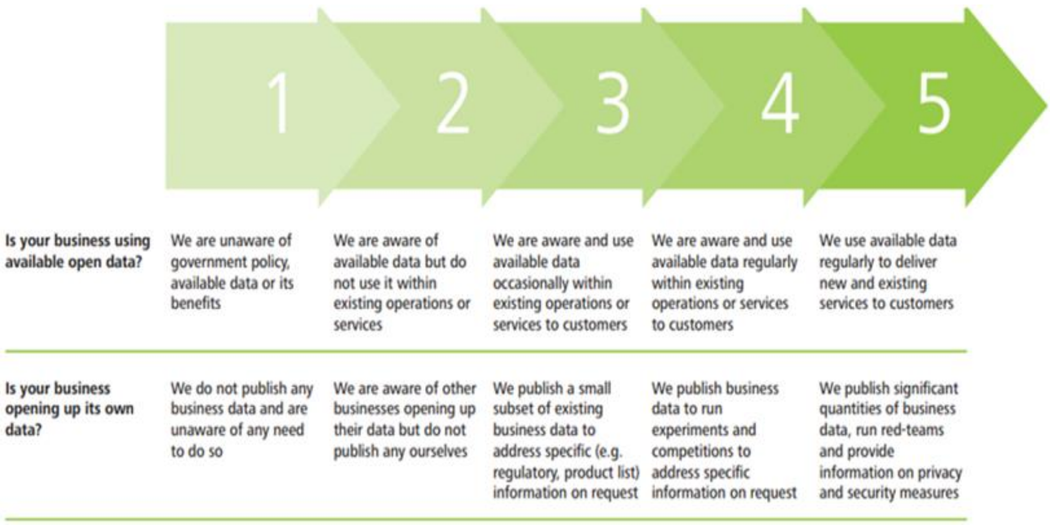

Figure 10 Five Step Open Data Journey (Deloitte Analytics, 2012, p.31).

Millions of points of data are generated from the design process to the completion process of products and services and their delivery to the end user. However, the awareness and willingness of government and companies to use this data and make what it can open, especially that which is for the public interest, has not yet fully matured. Just as in other sectors, the awareness of the public and commercial gain and value to be obtained by making data public remains low in the construction sector. In aiming to close this gap, this deliverable offers basic recommendations to open infrastructure data by prioritising privacy and security issues.

5.1 APPLYING DATA SHARING PRINCIPLES: THE CASE OF ZADAR AIRPORT

The ASHVIN Zadar Airport demonstration case concerns the maintenance of the airport runway by collecting drone footage in a mosaic approach. This drone footage has the purpose of identifying cracks in the runway surface and tire marks, thus supporting inspectors in deciding when and how to take appropriate maintenance action. Zadar airport also has the special constraint of having both a civil and a military runway, meaning that there are additional considerations in data sharing regarding classified and classified information.

The first step in opening data regarding the Zadar Airport demonstration case concerns the **meta data** which informs the public that data collection has taken place and the channels through which it took place. ASHVIN fulfilled this recommendation of the open data pyramid through a publication in its website.

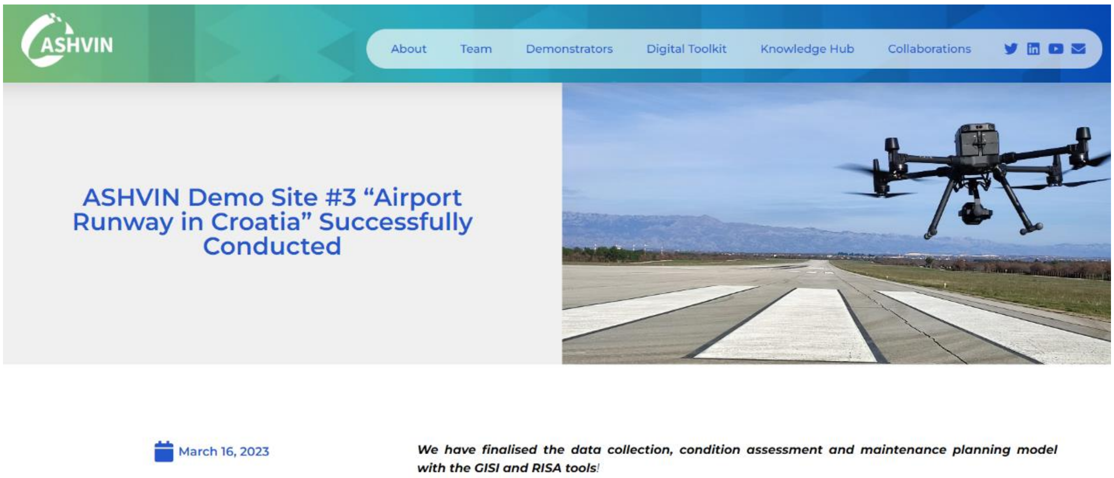

Figure 11 Meta Information About Maintenance Data Collection[15]

The second level of the pyramid, **raw data** sharing, was conducted after discussions between all stakeholders who have an interest in the data. In the case of ASHVIN, discussions involved mostly three partners of the ASHVIN project (INFRAPLAN, Erasmus University and CERTH. After discussion, it was agreed that a selection of the concrete runway images could be published for open access if the specific location information is removed. While this means that a full digital mosaic reconstruction is not viable with the available data, these raw images can nevertheless be used as a benchmark and reference to other maintenance projects or to developers who require training examples for crack and tire mark detection algorithms. The raw data was shared on Zenodo and made publicly available for reuse.

In the case of Zadar airport, the third level of data (**reports**) concerns the fully reconstructed runway and the outputs generated by the machine learning models that detected cracks and tire marks. Following the recommendations established in the ASHVIN open data pyramid, this information is not simply released to the public domain but is instead contextualised and redacted so that transparency requirements are met, but risks of misinterpretation and misuse are minimised. In the scope of ASHVIN, this is achieved through deliverable 3.1, which showcases examples and results from the crack and tire detection machine learning algorithms without disclosing the entire prediction files to the public. Information was curated in a way that still contributes to the goals of open data without compromising intellectual property or confidentiality.

---

[15] https://www.ashvin.eu/2023/03/16/ashvin-demo-site-3-airport-runway-in-croatia-successfully-conducted/).

Finally, **confidential data** concerning meetings and discussions surrounding how to use report data for decision-making within airport maintenance practices are not released in any form to the public, but interview transcripts and meeting minutes are stored securely in ASHVIN's internal work folder on Nextcloud.

The Zadar airport demo site, which makes infrastructure data publicly available, represents an important example as it ensures data justice by moving from a conceptual level to a practical level.

As seen in Zadar Airport demo site these processes and principles facilitated with opening infrastructure data by ensuring transparency and participation of other stakeholders make democratic management processes possible in the construction industry. This demo site creates a fair data environment by not neglecting to include other stakeholders in the process while making this data public. For instance, in the scope of ASHVIN, data release was considered in conjunction with data protection experts (EUR), technical experts (CERTH, INFRAPLAN), communication teams (AUS) and standardisation bodies (ASI). This contributes to impartiality while collecting data and revealing the information set to the public when necessary. It also does not ignore privacy concerns and possible negative security threats.

6 RECOMMENDATIONS

ASHVIN identifies the risks and concerns associated with opening infrastructure data and develops recommendations accordingly.

In general, there is a reluctance to expose infrastructure data for reasons such as security, commercial confidentiality, or privacy. Therefore, while deciding which data to open, some criteria should be taken into consideration. The most important of these, personal data, privacy, public and national security should be given relevance. Even if the information is anonymised while making it public, measures should be employed to prevent it from being deanonymised again when shared with other sources. Although the right to share the data is in the public domain, the security of the information containing the commercial secrets of the companies from which service is provided is also important. At this point data minimisation can be seen as an efficient strategy to balance privacy with transparency. As Scassa and Conroy point out “data minimisation can reduce the amount of personal information both by limiting collection only to that which is specifically necessary, and by limiting retention only for as long as is necessary” (Scassa & Conroy, 2017, p.350).

Improving the quality of data and determining which data will be shared with whom and how is essential in open public data strategies. Data stewards who encourage data sharing and ensure that data is managed and protected well can have an important mission in the data this data management process. Data stewards can also develop strategies in line with regulations and standardisations to reduce privacy risks and maximise security.

Regarding opening public infrastructure data specifically, the work carried out on ASHVIN informs the following five recommendations that should be adapted to each individual case, which are based on the previous sections of this deliverable. We label this the **ADAPT** (**A**vailability, **D**emonstration, **A**ccountability, **P**roportionality, **T**ransparency) framework:

I. Review the **Availability** of data in ways that does not exclude security of data.

When considering open data practices, it is important to understand that, just because data is made available and accessible, this does not mean that security concerns should be abandoned. As outlined in deliverable 1.5, access is only one dimension of cybersecurity. Adequate versioning and data protection mechanisms, such as digital object identifiers, should be used to prevent tampering with the data. Additionally, while released open data may not contain sensitive or confidential information, organisations should consider also what information can be inferred from released data (e.g. movements on the military Zadar airport runway inferred from civil runway footage).

II. Determine appropriate ways for the **Demonstration** of the benefits and harms of opening infrastructure data, including who is affected.

While opening data is generally recommended within the scope of the European Union (EU), there should be a clear purpose for opening data that is related to its potential beneficiaries. In the case of public infrastructure data, citizens, researchers and businesses can potentially benefit from access to and contributing to efforts in transparency. However, direct and indirect harms should also be considered, such as how opening maintenance data can also reveal intellectual property secrets or reinforce data imbalances by providing leading companies with additional free data.

III. Define who is **Accountable** for the benefits and the harms of opening infrastructure data.

Ultimately, the benefits of open data can only be realised if someone is available to answer questions about the data and collect the additional insights and outcomes that open data generates. Conversely, if there are negative consequences in relation to opening data, appropriate mitigation strategies and a responsible person or organisation are essential to limit the potential harms. As such, when releasing infrastructure data, it should be determined which entity or individual is accountable for that data.

IV. Data should be shared in **Proportion** to the potential risks and gains.

The principle of proportionality of data established in the EU's AI Act can also be applied to infrastructure data. While open data often leads to expected and unexpected benefits, the extent and amount of data made available should be considered in proportion to these benefits. To this effect, the ASHVIN open data pyramid may provide an adequate framework to consider these proportionality efforts.

V. Contextualise and create **Transparent** participation mechanisms in open data efforts.

One of the main advantages of open data is that organisations can benefit from a wide range of perspectives and expertise regarding the content of the data. To make it possible for these benefits to materialise, organisations should make channels available for individuals to submit analyses of the data (e.g., send academic papers that make use of the data to the accountable entity) or to contribute to the dataset (citizens submitting photos of cracks on the concrete).

7 REFERENCES

Chalikova, T., Mabillard, V. (2020). Open Data and Transparency: Opportunities and Challenges in the Swiss Context, *Public Performance & Management Review 43(3): 662-686.* https://doi.org/10.1080/15309576.2019.1657914.

Deloitte Analytics (2012), Open Data: Driving Growth, Ingenuity and Innovation, https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf

Donker, F.W. and Leonen, B. (2016). How to Assess the Success of the Open Data Eco System? International Journal of Digital Earth 10(3): 284 - 306 https://doi.org/10.1080/17538947.2016.1224938

European Commission, (2011). Decision 2011/833 - 2011/833/EU: Commission Decision of 12 December 2011 on the reuse of Commission documents, https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vivi69stinnd

Huijboom, N., & Van den Broek, T. (2011). Open data: An international comparison of strategies. European Journal of ePractice, 12(1), 1-12. https://www.researchgate.net/publication/285735704_Open_data_An_international_comparison_of_strategies

Janssen K (2012) Open Government Data and the Right to Information: Opportunities and Obstacles. The Journal of Community Informatics 8(2): DOI:10.15353/joci.v8i2.3042

Jetzek, T., Avital, M., Bjorn-Andersen, N. (2019). "The Sustainable Value of Open Government Data," *Journal of the Association for Information Systems*, 20(6), DOI: 10.17705/1jais.00549, https://aisel.aisnet.org/jais/vol20/iss6/6.

Johnson, J.A. (2014) From Open Data to Information Justice. *Ethics and Information Technology* 16(4): 263–274. https://doi.org/10.1007/s10676-014-9351-8

Mayer-Schonberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*, New York: Eamon Dolan/Houghton Mifflin Harcourt.

Mejias, U.A., Couldry, N. (2019). Datafication, *Internet Policy Review* 8(4): 1-10. DOI: 10.14763/2019.4.1428

MIT Technology Review Insight (2020). The Global AI Agenda, https://mittrinsights.s3.amazonaws.com/AIagenda2020/GlobalAIagenda.pdf

Purnama-Jati, P. H., Lin, Y., Nodehi, S., Cahyono, D. B., & van Reisen, M. (2022). FAIR versus open data: A comparison of objectives and principles. *Data Intelligence*, *4*(4), 867-881. https://doi.org/10.1162/dint_a_00176

Scassa, T., Conroy, A. (2017). The` Privacy/Transparency Balance in Open Government in *Government 3.0 – Next Generation Government Technology Infrastructure and Services Roadmaps, Enabling Technologies & Challenges*, Eds. Adegboyega Ojo and Jeremy Millard, Cham: Springer, p. 333-353.

Taylor, L. (2017). What is Data Justice? The Case For Connecting Digital Rights and Freedoms Globally, *Big Data & Society* 4(2): 1-14, https://doi.org/10.1177/2053951717736335

Verhulst, S., Young, A., Winowatan, M., Zahuranec, A.J. (2019). Leveraging Private Data For Public Good: A Descriptive Analysis and Typology of Existing Practices, https://datacollaboratives.org/static/files/existing-practices-report.pdf.

Verhulst, S.G., Young, A., Zahuranec, A.J., Aaronson, S.A., Calderon, A., Gee, M. (2020). How To Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing, Open Data Policy Lab https://opendatapolicylab.org/images/odpl/third-wave-of-opendata.pdf%20