Ref. Ares(2022)4834246 - 01/07/2022

Optimal Construction Management & Production Control

D9.8 – DMP Plan

WP9 – Dissemination, standardization, exploitation

Version 2

Issue date:	31/07/2021
Author(s):	Michele Scotton, Stefano Giulitti (UniSMART)
Editor:	Michele Scotton (UniSMART)
Lead Beneficiary:	Partner 13 – UniSMART
Dissemination level:	Public
Type	ORDP: Open Research Data Pilot

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 958398

EXPLANATIONS FOR FRONT PAGE

Author(s): Name(s) of the person(s) having generated the Foreground respectively having written the content of the report/document. In case the report is a summary of Foreground generated by other individuals, the latter have to be indicated by name and partner whose employees he/she is. List them alphabetically.

Editor: Only one. As formal editorial name only one main author as responsible quality manager in case of written reports: Name the person and the name of the partner whose employee the Editor is. For the avoidance of doubt, editing only does not qualify for generating Foreground; however, an individual may be an Author – if he has generated the Foreground – as well as an Editor – if he also edits the report on its own Foreground.

Lead Beneficiary of Deliverable: Only one. Identifies name of the partner that is responsible for the Deliverable according to the BIM2TWIN DOW. The lead beneficiary partner should be listed on the front page as Authors and Partner. If not, that would require an explanation.

Internal Reviewers: They should not belong to the authors. They should be any employees of the remaining partners of the consortium, not directly involved in that deliverable, but should be competent in reviewing the content of the deliverable. Typically, this review includes: identifying typos, identifying syntax and other grammatical errors, altering content, adding or deleting content.

BIM2TWIN Key Facts

Project title	BIM2TWIN: Optimal Construction Management & Production Control
Starting date	01/11/2020
Duration in months	42
Call (part) identifier	H2020-NMBP-ST-IND-2020-singlestage
Topic	LC-EEB-08-2020 Digital Building Twins (RIA)
Fixed EC Keywords	-
Free keywords	Digital Twin; Graph database; BIM; IA; Machine Learning; Image recognition; process optimization; safety improvement
Consortium	17 organizations

BIM2TWIN Consortium Partners

	Partner	Country
1	CSTB: CENTRE SCIENTIFIQUE ET TECHNIQUE DU BATIMENT	FR
2	TECHNION: ISRAEL INSTITUTE OF TECHNOLOGY	IL
3	UNIVERSITY OF CAMBRIDGE	UK
4	TUM: TECHNISCHE UNIVERSITAET MUENCHEN	DE
5	INRIA: INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE	FR
6	FIRA GROUP OY	FI
7	INTSITE LTD	IL
8	FUNDACION TECNALIA RESEARCH & INNOVATION	ES
9	ACCIONA CONSTRUCCTION SA	ES
10	RUHR-UNIVERSITAET BOCHUM	DE
11	SPADA CONSTRUCTION	FR
12	UNIVERSITA POLITECNICA DELLE MARCHE	IT
13	UNISMART – FONDAZIONE UNIVERSITÀ DEGLI STUDI DI PADOVA	IT
14	ORANGE SA	FR
15	SIEMENS AKTIENGESELLSCHAFT	DE
16	IDP INGENIERIA Y ARQUITECTURA IBERIA SL	ES
17	AARHUS UNIVERSITET	DK

DISCLAIMER

Copyright © 2020 by BIM2TWIN consortium

Use of any knowledge, information or data contained in this document shall be at the user's sole risk. Neither the BIM2TWIN Consortium nor any of its members, their officers, employees, or agents shall be liable or responsible, in negligence or otherwise, for any loss, damage or expense whatever sustained by any person as a result of the use, in any manner or form, of any knowledge, information or data contained in this document, or due to any inaccuracy, omission or error therein contained. If you notice information in this publication that you believe should be corrected or updated, please get in contact with the project coordinator.

The authors intended not to use any copyrighted material for the publication or, if not possible, to indicate the copyright of the respective object. The copyright for any material created by the authors is reserved. Any duplication or use of objects such as diagrams, sounds or texts in other electronic or printed publications is not permitted without the author's agreement.

EXECUTIVE SUMMARY

This document presents the Data Management Plan (DMP) of the BIM2TWIN project and, as such, it is submitted as Deliverable 9.8. The aim of this deliverable is to provide a general guideline on how to collect, structure, share and store information gathered or produced during the life cycle of the project. In addition, it includes policies and principles that must be adhered during data collection, storage and processing. Initial information about the types and nature of data and how data will be managed across the partners are identified.

The DMP document will be iteratively updated and improved in accordance with Consortium Partners' needs and with the actual outcomes of the project. At the time of writing (M9) not all the data to be managed across the entire lifespan of the project is precisely determined, and only the available information about datasets will be reported. This is because some specifications of the DBT platform are still at a conceptual stage, while some other data rely on information that will emerge in a further phase. For example, some use-case scenarios are still to be defined in detail and sensors/equipment are not installed yet so it is not possible to foresee the types of data that will be collected for the pilots in detail.

Further updates of the document will include additional information as the project progresses, covering all the datasets to be handled. As such, this release is the first version of a living document that will be properly updated, detailed and adapted alongside with the project development. Future versions are foreseen to be released as the project proceeds, aiming at integrating new datasets, updating the previously reported ones or adapting the plan according to specific data and knowledge-handling requirements. For this reason, the DMP has been structured in a flexible format, in order to accommodate these future needs.

This Data Management Plan plays an essential role in supporting and providing a fundamental regulatory framework for the implementation of dissemination and exploitation activities of project results. In fact, the DMP will facilitate the definition of a clear guideline for smooth and correct data collection, elaboration, diffusion, preservation and reuse, which are fundamental for dissemination, protection (IPR strategy) and exploitation activities. The purpose of this document is to provide a holistic analysis of the data management policy that will be adopted and followed by the Consortium in relation to the project research data, that during the development of the project will be gathered, collected and produced.

The DMP is expected to cover the complete lifecycle of the research data managed throughout the project and an agreed time span after the project termination (typically 5 years). Basically, it includes information about the types of data that will be produced or collected during the project, the adopted standards for data and metadata management, how data will be archived and preserved and the policy about data sharing. Moreover, the Data Management Plan is aligned to be compliant and consistent with the Communication and Dissemination Plan (CDP) (D9.1, D9.2 and D9.3), with the Exploitation Plan (D9.6 and D9.7) and with IPR requirements (and IPR Plan, D9.9). It aims to ensure that research data is disseminated through proper channels and modalities. Additionally, it specifies the conditions to make data available into open domains if they are publicly shareable or to be stored into restricted access repository if IPR concerns arise. In this plan and in the future versions, efforts will be focused on the best practices for exploitation of results in compliance with the Exploitation Plan and IPR protection, aiming at producing the largest possible impact on the market.

This document presents an overview and a first characterization of the datasets that are expected to be collected or generated by the different partners at this stage of the project. However, it will be enriched with new datasets and more information, describing additional practical data management procedures, as the Consortium provides more details and requirements during the project

development. Future versions of the DMP will be shared as additional updates of this deliverable at later stages of the project.

TABLE OF CONTENTS

EXPLANATIONS FOR FRONT PAGE **2**

BIM2TWIN KEY FACTS **3**

BIM2TWIN CONSORTIUM PARTNERS **3**

EXECUTIVE SUMMARY **4**

TABLE OF CONTENTS **6**

LIST OF FIGURES **7**

LIST OF TABLES **7**

ABBREVIATIONS **8**

1 INTRODUCTION **9**

2 PRINCIPLES AND REGULATIONS **11**

2.1 FAIR DATA PRINCIPLES 11

2.2 GENERAL DATA PROTECTION REGULATORY (GDPR) 11

3 DATA MANAGEMENT PLAN CONTENT AND PURPOSE **13**

4 TYPE OF DATA OVERVIEW **16**

5 BIM2TWIN PROJECT DATASETS **18**

5.1 SECTION 1: DATASET AND DATA COLLECTION 18

5.2 SECTION 2: STANDARDS AND METADATA 18

5.3 SECTION 3: DATA SHARING 19

5.4 SECTION 4: ARCHIVING AND PRESERVATION 20

6 DATASETS RELATED TO BIM2TWIN PARTNERS **22**

6.1 UNIVERSITY OF CAMBRIDGE 22

6.2 TECHNISCHE UNIVERSITAET MUENCHEN 24

6.2.1 DTP Data Model 24

6.2.2 BIM Integration Tool 27

6.3 TECNALIA 29

6.4 RUHR-UNIVERSITAET BOCHUM 32

6.4.1 Real World Image Dataset 32

6.4.2 Synthetic Image Dataset 34

6.5 UNIVERSITÀ POLITECNICA DELLE MARCHE 37

6.6 UNISMART – FONDAZIONE UNIVERSITÀ DI PADOVA 40

6.7 ORANGE SA 44

***7* CONCLUSION** **48**

APPENDIX A: LITERATURE **49**

LIST OF FIGURES

Figure 1 - Data Management flow chart 14

LIST OF TABLES

Table 1 - Dataset information collection 14
Table 2 - Table 2 - Initial overview of project expected impact and innovation introduced by BIM2TWIN (as reported in GA) 17
Table 3 - University of Cambridge's Dataset 24
Table 4 – TUM's Dataset 1/2 27
Table 5 – TUM's Dataset 2/2 29
Table 6 – Tecnalia's Dataset 31
Table 7 - RUB's Dataset 1/2 34
Table 8 - RUB's Dataset 2/2 37
Table 9 - UNIVPM's Dataset 40
Table 10 - UniSMART's Dataset 44
Table 11 - Orange's Dataset 47

ABBREVIATIONS

AI	Artificial Intelligence
BIM	Building Information Modeling
CDP	Communication and Dissemination Plan
D	Deliverable
DBT	Digital Building Twin
DEC	Dissemination-Exploitation-Communication
DMP	Data Management Plan
DoW	Description of Work
EEA	European Economic Area
EU	European Union
GA	Grant Agreement
GDPR	General Data Protection Regulation
H2020	Horizon 2020 Programme
IP	Intellectual Property
IPR	Intellectual Property Right
KPI	Key performance indicator
M	Month
NDA	Non-Disclosure Agreement
PDCA	Plan-Do-Check-Act
WP	Work Package

1 INTRODUCTION

In the BIM2TWIN project, all the 17 Consortium partners will have to deal with data – both collected and created– which requires a clear Data Management Plan. Each of the partners has specific expertise that enables implementing a high-quality collaborative project incorporating research and innovation data. The adoption and use of a Data Management Plan are of course, recommended, but overall necessary to regulate data collection, production, use, flow, access and preservation.

The Consortium partners of the BIM2TWIN project are multifaceted and heterogeneous, which affects the area of competence, but also which data will be handled and how they will be managed. The roles of partners also change according to the different work packages and tasks they are involved in. For example, there are data providers, data collectors, data integrators and data "re-users". As a result, the DMP needs to pursue a holistic approach and cover all the aspects that could be related to the management of data by the BIM2TWIN project's players. The intention is to provide an instrument able to maximize the protection and exploitation of the project results and to guarantee a smooth data flow and management.

Data management and sharing could raise issues related to IPR protection and sensitive information. Looking at the specific case of BIM2TWIN, the project aims to build a Digital Building Twin (DBT) platform for construction management that implements lean principles to reduce operational waste of all kinds, shortening schedules, reducing costs, enhancing quality andsafety and reducing carbon footprint. In relation to the implementation of the project, several data (including sensitive information) are going to be created, collected, shared, stored – in one word: managed. It must be clear that a potential disclosure of information could represent possible damage to protection, commercialization and industrialization of results. In this scenario, specific restrictions could be adopted as a legitimate reason to not compromise the protection of project results. Nevertheless, specific strategies will be adopted to limit such restrictions, in agreement with the project partners and the project coordinator, to guarantee data accessibility as recommended for EU-funded research projects.

The Data Management Plan, by regulating the use (thus the re-use) of data, has as a primary objective the broad and fast dissemination of project results for the benefit of research institutes, industry, citizens, and other stakeholders. Technology discovery processes could be accelerated and more quickly reach the market as well as easily available research data could increment visibility of researchers, increasing the chance for new potential collaboration and citations. In parallel to project results publication and dissemination, all the data needed to implement, validate, and exploit the results must be archived and preserved.

All the datasets that will be handled in the BIM2TWIN project will be included in the Data Management Plan. As previously stated, at the time of writing of the first DMP document (M9) only clearly available and identified datasets will be included in the report. Research data refers to information, both quantitative and qualitative, facts and numbers, collected or generated to be examined and set the basis for implementing the project. In the project context, examples of data include literature review, statistics, results of simulations and experiments, measurements, digital models, system architectures, processes, methodologies, business models, observation from fieldwork, video, and images. The purpose of this document is to provide a set of shared methodologies and policies about data usage and management, which shall ensure coherence across the partners. As the BIM2TWIN project is intended to be developed across 42 months, it is easily predictable that potentially new concerns about data management will arise during the project lifespan. Keeping this into account, the DMP is thought to be an always-reliable instrument, repeatedly updated, able to be a holistic point of reference for

data management. All the Consortium partners will be compliant with the DMP procedures and principles and contribute to its detailed development.

The core of this deliverable is composed of five sections:

- Main DMP principles that will be adhered in BIM2TWIN, namely FAIR principles and the GDPR regulations;
- Data Management Plan content and purpose;
- Overview of the data types considered in the project
- Structure of the different datasets to be managed;
- Detailed description of the datasets to be managed by the Consortium members;

2 PRINCIPLES AND REGULATIONS

Two key principles will be followed for the implementation of the BIM2TWIN project and thus will be strictly adhered in all the steps related to data handling and management. The first is the "FAIR Data Principles", which are a set of guiding principles in order to make data Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016). The second one is the GDPR regulation which regards about the data protection and privacy for all individuals within the European Union (EU) and the European Economic Area (EEA).

2.1 FAIR DATA PRINCIPLES

FAIR Guiding Principles for scientific data management and stewardship were published in 2016. They define principles for data, metadata, vocabularies, protocols in the following directions in order to increase data reusability:

- Findable - Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for the automatic discovery of datasets and services, so this is an essential component of the FAIRification process;
- Accessible - Once the user finds the required data, she/he needs to know how they can be accessed, possibly including authentication and authorization;
- Interoperable - The data usually need to be integrated with other data. In addition, the data need to inter-operate with applications or workflows for analysis, storage, and processing;
- Reusable - The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

FAIR sharing community developed a special curated portal (https://www.FAIRsharing.org) for datasets, which fulfill FAIR principles. These principles shall be adhered as much as possible in all the BIM2TWIN data management.

2.2 GENERAL DATA PROTECTION REGULATORY (GDPR)

The protection of personal data is of vital importance in the BIM2TWIN project. Datasets should be managed according to a system of proportionate governance. This means that personal data are only shared if necessary for research with the potential for high public value. BIM2TWIN project involves carrying out data collection in different cases, scenarios, and contexts (e.g., demo cases), so the Consortium will have to comply with all European and national legislation and directives relevant to the country where the data collections are taking place. The collection, processing and transmission of personal data will be analysed under principles of:

a) The recently published General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679);
b) The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data; and
c) the national laws applying its provisions.

Any additional regulations at the national level that do not fall under the GDPR and apply to data protection or any other sensitive information will also be considered. Data managed during the project will be processed only under the following preconditions which need to be met: (i) When the data subject has given her/his consent; (ii) When the processing is necessary for the project implementation; (iii) When processing is necessary for compliance with a legal obligation; (iv) When processing is necessary in order to protect the vital interests of the data subject. To this end, personal data managed within BIM2TWIN will be anonymized and stored in a form that does not permit users' identification. Moreover, data processing will be done regarding the purposes for which the data were collected or for which they are further processed while ensuring appropriate protection for personal data stored for longer periods for historical, statistical or scientific use. BIM2TWIN will establish a data management framework that guarantees the security of collected personal data from potential abuse, theft, or loss.

3 DATA MANAGEMENT PLAN CONTENT AND PURPOSE

The Data Management Plan described in this document will define how to select, structure, store, and share the information gathered or produced during the project's life cycle. Each type of data will make up a dataset. Considering these datasets, this document will contain the following information for each one of them:

- Dataset reference and name;
- Dataset description;
- File format;
- Standards and metadata;
- Data sharing;
- Archiving and preservation.

A "Question-and-Answer" approach will be followed across the four different sections, to tackle and give answer to the following questions:

1. Dataset and data collection: "What data will be collected/generated during the project and who is responsible for these data?"
2. Standards and metadata: "What standards/methodologies will be adopted for data and metadata?"
3. Data sharing: "What/How data will be shared?"
4. Archiving and preservation: "How data will be preserved/archived?"

This set of information has been collected by WP9 leader UniSMART from project partners through a specific template in the form of a table, as showed in Table 1:

SECTION		GUIDANCE	STATEMENT
1 - Dataset and data collection	Dataset reference and name	Name, publisher, maintainer, responsible for quality/data security/sharing …	…
	Dataset description	Description, provenience, type of data, usefulness, similar data, re-use and integration	…
	File format	Nature and format (text, pictures, video; pdf, doc, ppt, avi, …)	…
2 - Standards and metadata		Metadata description, standards adopted, quality assurance standards	…
3 - Data sharing		What data is suitable for sharing and why, license, openness/access policy, restrictions, confidentiality, sharing authorization, software/hardware requirements, repository, quoting policy, potential users, …	…
4 - Archiving and preservation		Data management, back up, archive, size; short/long term preservation policies, preservation costs (if applicable), data security, sensitive data protection, security standards, risks to data security (in terms of confidentiality and security: level and likelihood, impact,	…

	monitoring, mitigation), ethics and legal compliance …	

Table 1 - Dataset information collection

A qualitative description of the adopted methodology to manage data is represented by the flow chart in Figure 1.

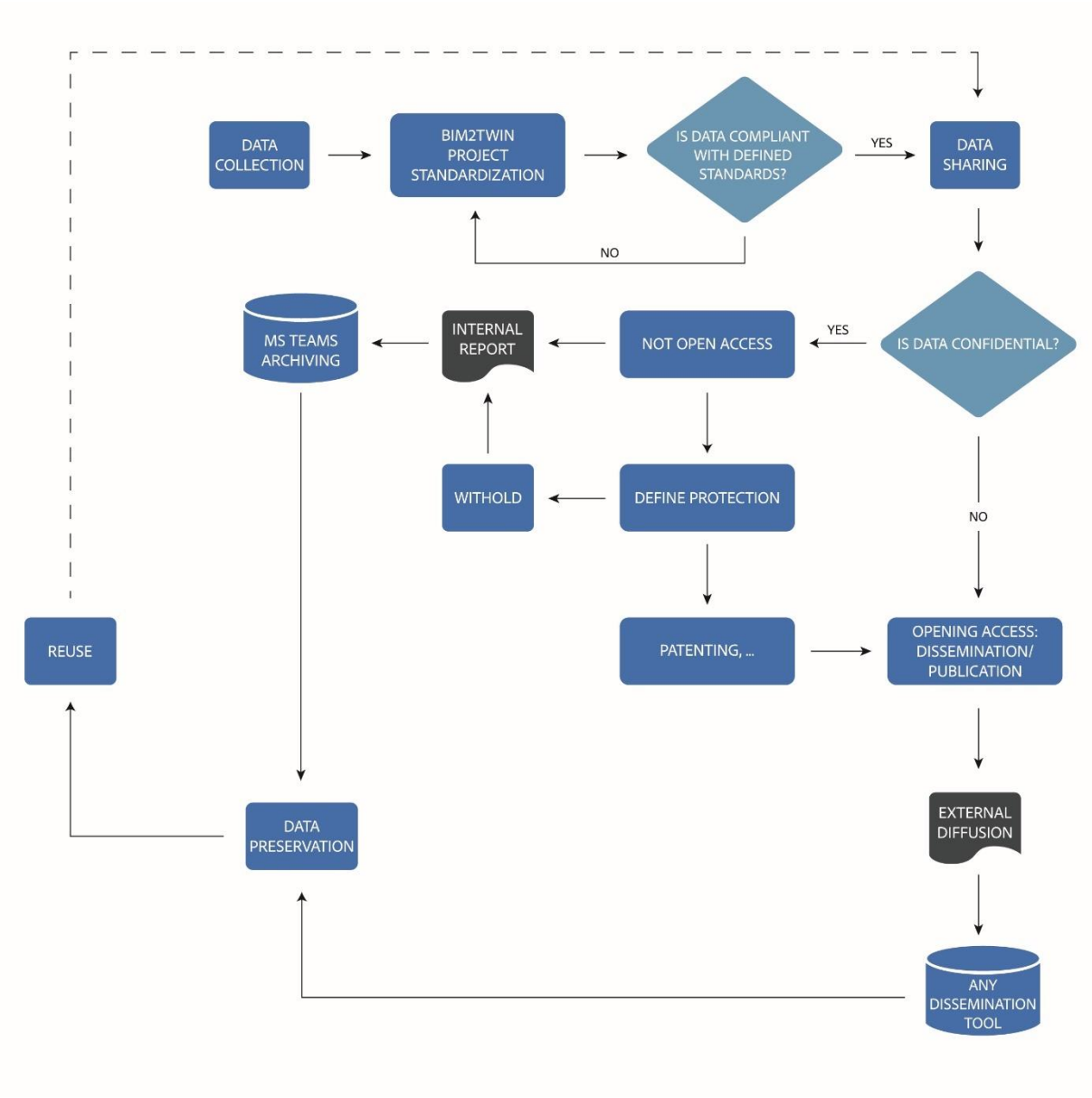

Figure 1 - Data Management flow chart

As already pointed out, the DMP has the purpose of outlining the key points of the data management policy and methodologies that will be adopted and followed by the project partners during the project life cycle – and beyond, if the case, regarding all the gathered and produced data. As long as the project proceeds, the DMP is expected to move from an initial overview of the datasets produced during the project towards a more detailed description of the practical procedures related to data management. It is important to specify that all the data collected during the project development must be well structured and preserved according to agreed methodologies: this is for the best use, re-use, and integration of data, even if some of those data could seem secondary in the very moment of its collection. These cited methodologies rely on standards, quality level, and sharing policies that have been agreed upon and fixed.

4 TYPE OF DATA OVERVIEW

This first version of the DMP will analyse the data types to be managed by consortium partners, with the information available at the moment of writing. Combining this with the information included in the BIM2TWIN proposal, an overview of the indicative types of data and the associated partner is shown in the following paragraphs. The initial overview of the expected impacts from the project implementation is reported below, in order to provide a first qualitative idea of the areas of competences developed in BIM2TWIN, as indicated in the Grant Agreement. This list of impacts provides a preliminary framework where the project know how will be implemented during the project development. In the following section of this document, each dataset will be described more in detail.

EXPECTED IMPACT	BIM2TWIN'S INNOVATION
Impact on project schedule and cost	BIM2TWIN will provide a DBT system design and prototype that can achieve this target by transforming production management from reactive to proactive (AI software, situational awareness, production control)
Impact on Resource Allocation and Operation	The BIM2TWIN platform will provide accurate, reliable information about available equipment including monitored records of its performance and capacity over time and under different conditions
Impact on Construction Safety	BIM2TWIN will enhance construction safety thanks to its applications (prediction of risky situation, prevention of hazardous environments, real-time risk monitoring and detection, improving workers' awareness, …)
Standardisation for Digital Building Twins at a European Level	BIM2TWIN will contribute to the future standardisation for Digital Building Twins (set of information models, support semantic interoperability, develop a standardisation plan)
Impact on project portfolio management	BIM2TWIN platform will enable project portfolio management activities so that so that similar work/phases do not overlap, resulting in accurate scheduling and predictability of production
Standardization of the work and workplace	Adoption of PDCA approach (Plan-Do-Check-Act), resulting in higher quality with less skilled resources. Implementation of Lean process. Certifications of new buildings may rely on the digital footprint stored in the platform

Impact on innovative solutions	Innovative approaches for automated data collection fostering innovation across the entire construction value chain
Environmental impacts	Reduced waste of materials, reduced wasteful operation of equipment, increased recycling due to material tracking, better inventory of carbon footprint through data collection

Table 2 - Table 2 - Initial overview of project expected impact and innovation introduced by BIM2TWIN (as reported in GA)

5 BIM2TWIN PROJECT DATASETS

This section is intended to provide a deeper specification of the different datasets to be produced and managed in BIM2TWIN, in relation to what has been identified at this stage of the project. At the time of writing the first version of this deliverable, Consortium partners provided their inputs about datasets characterization and management through a specific template in a table format, as explained above. A guideline to better understand the content of this table template is reported below, where four sections will describe in detail which information is expected to be collected. As already outlined, the DMP is a living document that will evolve during the project's development, where more and more detailed information will be provided. As long as project results and data will be produced and, necessarily, managed, new and more detailed dataset descriptions will be produced.

5.1 SECTION 1: DATASET AND DATA COLLECTION

This first section of the Data Management schema can be considered as the unique identifier for a specific dataset to be produced. It will provide preliminary and descriptive information, in particular dataset reference and name, dataset description, and structural way of collecting data (namely file format). This section is intended to clearly provide a first explanation of the different types of data that will be produced and managed by each partner in BIM2TWIN. Given the fact that the nature and characteristics of DMP's datasets can progress throughout the project, more detailed descriptions could be provided through the different versions of the document. As an example of the expected content of this section, dataset description could include details about data generation/collection (source, origin, provenience, etc.), nature and file format, the usefulness of data (why and to whom), re-use and integration. In few words, this section will act as an "identity card" of the managed dataset, with specific concern on the qualitative description. A set of questions to be answered through this section could be the following one:

- What data will you collect or create?
    - Type, format and volume (keep into account implications in terms of storage, backup, access and interoperability)
    - Indicate any existing reusable data
    - Indicate data source if the case
- How will data be collected or created?
    - Specify any standard or specific methodology for collecting, creating and processing data
    - Data organization: naming conventions, folder structure, file versioning…
    - Data collection quality assurance processes
- How is data planned to be protected?
- How is data planned to be disseminated?
- Who will be responsible for data collection and management?
    - Consider data capture, metadata production, data quality assurance, etc.

5.2 SECTION 2: STANDARDS AND METADATA

Metadata, literally “beyond data”, stands for “data about data”. They represent accompanying information that helps other users interpret and comprehend the published/shared/disseminated data. Metadata are also fundamental for data organization, search and access since they play a crucial role in data identification and localization. In general, this section will host a clear description of adopted metadata and any accompanying documentation: as an example, this set of data documentation could provide information about the creator, time reference, origin, access conditions, terms of use, or even technical aspects, such as units of measurements or software used to collect data. Metadata could also refer to specific and recognized standards that are maintained by standard organizations or refer to chosen guidelines or methodologies in relation to their purpose, type or domain: metadata standards, indeed, refer to specific disciplines and areas of relevance. These standards or guiding methodologies can also be referred to quality assurance issues, such as standards or processes for calibration, validation, or review. It is not compulsory to refer to a specific metadata standard. However, it must be outlined what metadata will be created and how. A set of questions to be answered by this section could be:

- What documentation and metadata will accompany the data?
    - What information is needed for the data to be read and interpreted when queried (creator, contributors, date of creation, the definition of variables, units of measurement, software used to collect data, etc.)?
    - Description of data in terms of content and quality.
    - How will support documentation and metadata be created and where will it be recorded?
    - Indicate any metadata standards that will be used.

5.3 SECTION 3: DATA SHARING

Considering the fact that BIM2TWIN is a collaborative project based on the Horizon 2020 framework and developed by a Consortium of partners playing specific and significant roles, data sharing, exchange and dissemination are topics of crucial importance. For this reason, this third section must face in a comprehensive manner all those aspects related to sharing and dissemination of information: what data will be shared, how data will be shared, access policies, restrictions, technical needs for sharing, openness, etc. The first main distinction can be considered the data accessibility: publicly available or for internal use of the Consortium. In this latter case, a specific tool or repository is typically used, and this is also the case of the BIM2TWIN project, where the MS Teams portal plays this role. For any other shareable data that do not require to be undisclosed, specific ways, procedures and requirements for dissemination must be outlined and respected for the best impact of the project on society and to safeguard partners’ interests. It must be pointed out that partners wishing to publish or disseminate any content or material should notify this intention to the Consortium in order to enable the consortium board to assess whether the requested action is suitable in terms of IPR protection. Accessibility and shareability issues will be specifically faced through the CDP document and the development of this DMP. In any case, since the BIM2TWIN project aims at proposing improvement models for buildings through the implementation of new technical and economical solutions, the achievements are required to be communicated publicly and the results to be distributed openly. However, some specific data, such as quality and performance data of the Manufacturing enterprises, will be considered private, therefore accessible after granting permission. In parallel, specific datasets collected or generated during the BIM2TWIN project, such as research data about modelling procedures, KPI validation, event modelling, inspection, real-time quality control and system optimization, will be distributed freely. The Data Management Plan will be developed with a specific

focus on the different datasets managed by the consortium partners of the BIM2TWIN project, where specific tables will indicate which, whether and how data will be shared and/or made open, managed, curated, preserved.

Section three is intended to examine the pointed-out topics by the following questions:

- How will access and security be managed?
    - Are there any possible risks to data security?
    - Is there access control to keep data secure?
    - How will secure access be guaranteed to partners?
    - Indicate any relevant standard followed
- How will safe data transfer be guaranteed?
- How will data be shared?
    - Use of repository, secure data service, direct data handling, etc.
    - When will data be available for sharing? Will any period of embargo be defined?
    - Describe technical processes for dissemination (any software or tool needed, etc.)
    - How will potential users find out about shared data?
    - Who is the addressee of data sharing?
    - Will there be any access restriction to shared data and why?
    - Indicate any action to minimize restrictions (NDA, etc.)

5.4 SECTION 4: ARCHIVING AND PRESERVATION

The project consortium will share project results, both internally and externally, through specific repositories, without compromising the protection and exploitation strategy. For what concerns the data to be shared within the Consortium, the MS Teams platform will be the primary instrument to operate this function. On the other hand, shareable and open data should be made publicly available and stored on the BIM2TWIN website, as well as on Open Access repositories. Of course, the project website shall be the reference instrument for archiving and storing public data. However, several co-existing options could be considered to this aim. There are specific institutional repositories, or subject-based repositories that could host scientific publications, as well as general research data repositories or centralized repositories that could host the generated research data and make them available to third parties. As an example, the academic institutions taking part in BIM2TWIN could have access to specific repositories, as well as project partners' websites could act as a redirection of specific content.

Data files and associated metadata will be backed up and replicated in multiple copies concerning the specific associated repository. Currently, the project consortium uses the MS Teams platform primarily to exchange and store project-related files. If there is a need to use other repositories which are in line with the general principles of the DMP, they can be used and will be indicated in the next version of the DMP.

An example of possible questions to be answered through section four is as follow:

- How will the data be stored and backed up?
    - Partner's internal storage / Common Data Environment / Cloud, etc.
    - How and where will data be backed up? How often? How many copies?
    - How will data be potentially recovered?
    - Who is responsible for data backup and recovery?
    - Will specific existing procedures be followed?

- Which data will be stored for a long-term period?
    - How will long-term data be stored?
    - How long will the data be retained and preserved?
    - How will data be reused (validation, new studies, teaching, etc.)?
    - Is there any preservation plan for the dataset?
- How will ethical issues be managed?
    - Consider consent aspects of data preservation and sharing.
    - Consider privacy issues of involved people and sensitive data handling (anonymization, formal agreements, etc.).
- How will copyright and IPR issues be managed?
    - Who is the data owner?
    - How will the data be licensed for reuse (patenting, etc.)?
    - Who will be able to reuse the data?
    - Are there any restrictions on the reuse of data by third parties?
    - Will data sharing be postponed? (e.g., for publishing or patenting issues)

The four sections outlined above should be preparatory for a better comprehension and interpretation of how datasets providers are managing or foresee to manage information across data generation and collection. The following paragraphs will report partners' inputs available at this stage of the project, reflecting the table format presented in **Erreur ! Source du renvoi introuvable.**.

6 DATASETS RELATED TO BIM2TWIN PARTNERS

6.1 UNIVERSITY OF CAMBRIDGE

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Cambridge on-site spatial and visual dataset
	Publisher	UCAM
	Maintainer	UCAM
	Responsible for data quality	UCAM
	Responsible for data security	UCAM
	Responsible for data sharing	UCAM
Data set description	Description	Indoor and outdoor as-built low-level geometry and visual data (points and pixels).
	Provenience	Experimental data
	Type of data	Mainly quantitative
	Usefulness	Provide 3D and 2D data from the various data acquisition technologies, in preparation for the Interpret and Merge functions.
	Similar data	No
	Re-use and integration	Data could be published as open-source dataset and used for a wider diffusion of project results or features, towards different markets, to reach a wider audience and to set the basis for further development.
	Data organization	TBD (may be MS Teams)
File format	Nature	Images and point clouds
	Format	.jpg, .avi, .las
	Metadata description	Capture spatial, visual and thermal data of construction site at different phases

Standards and metadata	Standards adopted	D10.1 "Technical quality assurance and risk management plan" will be kept as a guideline for data collection processes. Not specific metadata standards will be adopted.
	Quality assurance standards	D10.1 "Technical quality assurance and risk management plan" will be followed to ensure proper quality to managed data.
Data sharing	What data is suitable for sharing and why	The entire dataset will be published as open source dataset
	License	Owned by UCAM
	Openness/Access policy	Data is confidential before publication. Confidential data emerging from project development will be available among partners.
	Restrictions	Confidential before publication
	Data availability	Data will be shared whenever consented by UCAM
	Sharing authorization	The dataset is required to be cited when used
	Software/Hardware requirements	High-resolution camera, infrared camera and LiDAR
	Repository	Local storage
	Quoting policy	The dataset is required to be cited when used
	Potential users	The potential users will be different stakeholders in the AEC industry, such as building owners, managers, utility- and other building/construction-related service providers, end users, investors, research centres, and any other possible user interested in implementing BIM2TWIN solutions.
Archiving and preservation	Data management	Data will be mainly stored in local server
	Back up	Data will be backed up in local server
	Archive	Data will be archived in local server
	Size	TBC (based on the areas being scanned)
	Short/Long term preservation policy	Disseminated data will remain available for the whole duration of the project

	Preservation costs	Cost of hard disks with large storage space
	Confidential and sensitive data	Confidential before publication
	Security Standards	Data security will be ensured by service providers (in the case of social media and Cloud services). Moreover, data will be transferred through secure and encrypted channels.
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	No risks are identified
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	UCAM will verify IPR of data related to dissemination material.

Table 3 - University of Cambridge's Dataset

6.2 TECHNISCHE UNIVERSITAET MUENCHEN

6.2.1 DTP Data Model

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Digital twin data model
	Publisher	TUM
	Maintainer	TUM
	Responsible for data quality	Partners providing partial data models (Core data model, progress evaluation, quality evaluation, safety control, equipment optimization and operation and maintenance)
	Responsible for data security	All partners following data security policy
	Responsible for data sharing	Owner partners and TUM

Data set description	Description	Digital twin data model that describes object- and process-related entities together with their interconnection as a basis for the central graph database
	Provenience	Partially derived from the WP-specific BPMNs
	Type of data	Qualitative
	Usefulness	Useful for all project partners since it visualizes the connections between different types of data, clarifies which information can be retrieved from the DTP and how information that is send to the DTP needs to look like.
	Similar data	There are existing data models that are partially similar to the DTP data model that is to be developed (e.g. the DICO data model).
	Re-use and integration	Can be used in all WPs to define the data exchange between the developed software tool and the central platform
	Data organization	For MS Teams platform storage, commonly agreed standards will be followed. For local storage, specific company's policy will be followed.
File format	Nature	Diagrams in UML format (or a similar format)
	Format	.pdf, .png, .jpg
Standards and metadata	Metadata description	The data will be accompanied by specific information in order to specify the data creator, date of creation, data owner and data objective.
	Standards adopted	No specific metadata standards will be adopted.
	Quality assurance standards	Primarily UML modeling standards will be followed to ensure quality.
Data sharing	What data is suitable for sharing and why	Data will be shared with the BIM2TWIN project. (Sharing for research purposes?)
	License	To be verified with specific data providers (partners).
	Openness/Access policy	Data emerging from the project development will be available among the partners.
	Restrictions	Data owner will specify any data restrictions or necessity for sharing.

	Data availability	Digital building twin data model will be shared whenever consented by the data owners.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	Only basic applications for viewing PDFs and graphical content.
	Repository	MS Teams will be used as a repository for project related data.
	Quoting policy	Will be specified by data owners in case it is needed.
	Potential users	Mainly partners from WP2 and WP3-7 who will develop tools that will interact with the central platform.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	Data will be stored and backed up in MS Teams.
	Archive	Data will be archieved in MS Teams and in private servers.
	Size	Approximately 10GB
	Short/Long term preservation policy	All the information will remain available for the whole duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined
	Confidential and sensitive data	Data owners will label confidential and sensitive data accordingly.
	Security Standards	Data security will be ensured by service providers (MS Teams).
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	Low risk to data security. The partners' risks are to be specified by the data owner.
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.

	Copyright and IPR issues	TUM will verify IPR of data related to the data model.

Table 4 – TUM's Dataset 1/2

6.2.2 BIM Integration Tool

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	BIM integration tool
	Publisher	TUM
	Maintainer	TUM
	Responsible for data quality	Mainly TUM
	Responsible for data security	All partners following data security policy
	Responsible for data sharing	TUM
Data set description	Description	BIM integration tool that is necessary to map input data, e.g., as-designed and as-planned models to the graph database of the central digital twin platform
	Provenience	Several sources that will influence the development of the BIM integration tool: IFC documentation (and other documenations), xBIM, Orange Thing'in Injectors, …
	Type of data	Qualitative
	Usefulness	Useful for all partners that provide input data for the central graph database since the tool provides a mapping that converts, e.g. .ifc files and .mpp schedule files to nodes and relationships in the property graph
	Similar data	Similar to Orange Thing'in Injectors
	Re-use and integration	TBD
	Data organization	Organized in a Visual Studio project with all the necessary files
File format	Nature	Visual Studio project written in C#
	Format	.cs and others

Standards and metadata	Metadata description	The data will be accompanied by specific information in order to specify the data creator, date of creation, data owner and data objective (git versioning).
	Standards adopted	No specific metadata standards will be adopted. (git?)
	Quality assurance standards	Common practices in coding will be followed to ensure the quality of the developed tool.
Data sharing	What data is suitable for sharing and why	Source code will only be shared within TUM
	License	To be verified with specific data providers (partners).
	Openness/Access policy	BIM integration tool will be integrated into the central digital building twin platform and can be used by any project partner.
	Restrictions	Data owner will specify any data restrictions or necessity for sharing.
	Data availability	BIM integration tool will be shared whenever consented by the data owners.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	Visual Studio or a similar IDE.
	Repository	Gitlab.lrz will be used as a repository for BIM Integration related source code.
	Quoting policy	Will be specified by data owners in case it is needed.
	Potential users	Mainly partners from WP2 and WP3-7 who will develop tools that will interact with the central platform.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	Data will be stored and backed up in gitlab.lrz.
	Archive	Data will be archived in gitlab.lrz and in private servers.
	Size	Approximately 20GB

	Short/Long term preservation policy	The BIM integration tool will remain accessible for the whole duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined
	Confidential and sensitive data	Data owners will label confidential and sensitive data accordingly.
	Security Standards	Data security will be ensured by service providers (gitlab.lrz).
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	Low risk to data security. The partners' risks are to be specified by the data owner.
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	TUM will verify IPR of data related to the data model.

Table 5 – TUM's Dataset 2/2

6.3 TECNALIA

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Images for defect detection
	Publisher	TECNALIA
	Maintainer	TECNALIA
	Responsible for data quality	Partners developing core data.
	Responsible for data security	All partners following data security policy.
	Responsible for data sharing	Owner partners and Tecnalia
Data set description	Description	Dataset to create deep learning-based models for defect detection
	Provenience	---

	Type of data	Images
	Usefulness	The dataset will be used in the development of the vision module for defect detection and characterization of surfaces in WP4.
	Similar data	---
	Re-use and integration	The requirement for data re-use is the respect of the IP.
	Data organization	To be defined according to the use case and availability of data
File format	Nature	Usable data for development of the vision module
	Format	.jpg, .png
Standards and metadata	Metadata description	The data will be accompanied by defect type or category
	Standards adopted	There are no standards identified
	Quality assurance standards	There are no standards identified
Data sharing	What data is suitable for sharing and why	DEC data will be shared and spread through different channels in compliance with IPR.
	License	To be verified with specific data providers (partners).
	Openness/Access policy	Confidential data emerging from project development will be available among partners. Public data will be disseminated, when relevant, and agreed with partners.
	Restrictions	Data owners will specify any data restrictions or necessity for sharing.
	Data availability	DEC data will be shared whenever consented by data owners.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	Not specific software or hardware necessities

	Repository	MS Teams will be used as repository for project related data. An open access repository will be adopted to store publicly disclosable project results.
	Quoting policy	Data owners will specify any quoting policy, if considered.
	Potential users	The potential users will be different stakeholders in the AEC industry, such as building owners, managers, utility- and other building/construction-related service providers, end users, investors, research centres, and any other possible user interested in implementing BIM2TWIN solutions.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	Data will be stored and backed up in MS Teams.
	Archive	Data will be archived in MS Teams and in private servers, when relevant, and in the project website.
	Size	Estimated approx. 50GB.
	Short/Long term preservation policy	Dataset will remain available for the whole duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined.
	Confidential and sensitive data	Data owner will specify confidentiality of data and the presence of sensitive data.
	Security Standards	Data security will be ensured by service providers (in the case of social media and Cloud services). Moreover, data will be transferred through secure and encrypted channels.
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	No risks are identified
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	TECNALIA will verify IPR of image dataset

Table 6 – Tecnalia's Dataset

6.4 RUHR-UNIVERSITAET BOCHUM

6.4.1 Real World Image Dataset

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Real world image dataset (Equipment)
	Publisher	RUB
	Maintainer	RUB
	Responsible for data quality	RUB + Partners
	Responsible for data security	RUB + Partners
	Responsible for data sharing	RUB
Data set description	Description	Images from real construction sites. From different perspectives (Bird`s, Perspective of a surveillance camera).
	Provenience	From camera captured images/videos
	Type of data	Mainly qualitative.
	Usefulness	Captured images are needed to generate and evaluate different machine learning models.
	Similar data	For the creation of the machine learning models, data is used that was created during the project. This means it is available as long as the project is running. Data from other work or from other project participants may also be used.
	Re-use and integration	The requirement for data re-use is the respect of the IP. Data can be re-used for a wider diffusion of project results or features, towards different markets, to reach a wider audience and to set the basis for further development.
	Data organization	With prior consultation with the project partners, generally agreed standards will be followed for storage on the MS Teams platform. For local storage, specific company policies are followed.
File format	Nature	images, video (from real construction sites)
	Format	jpeg, png, mp4 (h.264)

Standards and metadata	Metadata description	The data will be accompanied by specific information in order to specify data creator, date of creation, data owner and data objective.
	Standards adopted	No specific metadata standards are adopted.
	Quality assurance standards	Methods from other published works are followed to obtain a consistent data set.
Data sharing	What data is suitable for sharing and why	The data is mainly shared in publication. The images may also appear in press reports related to BIM2TWIN.
	License	To be verified with specific data providers (partners).
	Openness/Access policy	Confidential data resulting from project development will be available among partners. Public data will be disseminated when relevant and agreed with the partners.
	Restrictions	Data owners will specify any data restrictions or necessity for sharing.
	Data availability	Data will be shared whenever consented by data owners.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	Not specific software or hardware necessities
	Repository	MS Teams will be used as repository for project related data. An open access repository will be adopted to store publicly disclosable project results (for example kaggle).
	Quoting policy	Data owners will specify any quoting policy, if considered.
	Potential users	The potential users are: Research centres and all other potential users interested in implementing BIM2TWIN solutions.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	The data is stored and backed up in MS Teams and on private servers.

	Archive	Data will be archived in MS Teams and in private servers. Public data will be stored on publicly accessible servers.
	Size	Estimated approx. 10 - 50 GB
	Short/Long term preservation policy	Disseminated data will remain available for the whole duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined.
	Confidential and sensitive data	Data owner will specify confidentiality of data and the presence of sensitive data.
	Security Standards	Data security will be ensured by service providers (in the case of social media and Cloud services). Moreover, data will be transferred through secure and encrypted channels.
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	No risks are identified
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	RUB will verify IPR of data related to dissemination material.

Table 7 - RUB's Dataset 1/2

6.4.2 Synthetic Image Dataset

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Synthetic image dataset (Equipment)
	Publisher	RUB
	Maintainer	RUB
	Responsible for data quality	RUB + Partners
	Responsible for data security	RUB + Partners

	Responsible for data sharing	RUB
Data set description	Description	Images created from 3D renderings. Images will represent different perspectives (bird's eye view, surveillance camera perspective).
	Provenience	The synthetic images are generated by means of 3D models. For this purpose, a renderer is used that generates photorealistic images.
	Type of data	Mainly qualitative.
	Usefulness	Synthetic data is used to enrich the data set for training.
	Similar data	For the creation of the machine learning models, data is used that was created during the project. This means it is available as long as the project is running.
	Re-use and integration	The requirement for data re-use is the respect of the IP. Data can be re-used for a wider diffusion of project results or features, towards different markets, to reach a wider audience and to set the basis for further development.
	Data organization	With prior consultation with the project partners, generally agreed standards will be followed for storage on the MS Teams platform. For local storage, specific company policies are followed.
File format	Nature	images, video (From a 3D environment)
	Format	jpeg, png, mp4 (h.264)
Standards and metadata	Metadata description	The data will be accompanied by specific information in order to specify data creator, date of creation, data owner and data objective.
	Standards adopted	No specific metadata standards are adopted.
	Quality assurance standards	Methods from other published works are followed to obtain a consistent data set.
Data sharing	What data is suitable for sharing and why	The data is mainly shared in publication. The images may also appear in press reports related to BIM2TWIN.
	License	To be verified with specific data providers (partners).

	Openness/Access policy	Confidential data resulting from project development will be available among partners. Public data will be disseminated when relevant and agreed with the partners.
	Restrictions	Data owners will specify any data restrictions or necessity for sharing.
	Data availability	Data will be shared whenever consented by data owners.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	Not specific software or hardware necessities
	Repository	MS Teams will be used as repository for project related data. An open access repository will be adopted to store publicly disclosable project results (for example kaggle).
	Quoting policy	Data owners will specify any quoting policy, if considered.
	Potential users	The potential users are: Research centres and all other potential users interested in implementing BIM2TWIN solutions.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	The data is stored and backed up in MS Teams and on private servers.
	Archive	Data will be archived in MS Teams and in private servers. Public data will be stored on publicly accessible servers.
	Size	Estimated approx. 10 - 50 GB
	Short/Long term preservation policy	Disseminated data will remain available for the whole duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined.
	Confidential and sensitive data	Data owner will specify confidentiality of data and the presence of sensitive data.

	Security Standards	Data security will be ensured by service providers (in the case of social media and Cloud services). Moreover, data will be transferred through secure and encrypted channels.
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	No risks are identified
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	RUB will verify IPR of data related to dissemination material.

Table 8 - RUB's Dataset 2/2

6.5 UNIVERSITÀ POLITECNICA DELLE MARCHE

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	UNIVPM on-site collection data
	Publisher	UNIVPM
	Maintainer	UNIVPM
	Responsible for data quality	UNIVPM and partners cooperating with UNIVPM.
	Responsible for data security	UNIVPM and all partners following data security policy.
	Responsible for data sharing	UNIVPM and partners cooperating with UNIVPM.
Data set description	Description	Indoor and outdoor as-built data collected in T2.3: 2D images at different wave length, 3D point cloud from LIDAR and data from positioning system. Some data are collected in laboratory conditions.
	Provenience	Experimental data.
	Type of data	Qualitative and quantitative

	Usefulness	The produced dataset will be useful to integrate the datasets present in the literature and share knowledge with BIT2TWIN partners. Provide data for the Interpret and merge function and the AI platform developed in WP4. The collected data will be useful also for training a neural model able to classify and perform semantic segmentation of concrete pathologies.
	Similar data	For dissemination activities will be almost exclusively used data created during the project, so available as long as project proceeds.
	Re-use and integration	The requirement for data re-use is the respect of the IP. Data can be re-used for a wider diffusion of research results and to set the basis for further development.
	Data organization	MS Teams platform storage, where commonly agreed standards will be followed and local storage.
File format	Nature	dataset of images, hypercubes, 3D point clouds and other experimental data.
	Format	.docx, .jpg, .png, .xlsx, .pdf, .npy, .DAT
Standards and metadata	Metadata description	Capture spatial, visual and hyperspectral data of construction site at different phases
	Standards adopted	D10.1 “Technical quality assurance and risk management plan” will be kept as a guideline for data collection processes. Not specific metadata standards will be adopted.
	Quality assurance standards	D10.1 “Technical quality assurance and risk management plan” will be followed to ensure proper quality to managed data.
Data sharing	What data is suitable for sharing and why	The dataset produced will be available on for consortium members.
	License	owned by UNIVPM
	Openness/Access policy	Data is confidential before publication. Confidential data emerging from project

		development will be available among partners.
	Restrictions	Confidential before publication
	Data availability	Data will be shared whenever consented by UNIVPM
	Sharing authorization	The dataset is required to be cited when used
	Software/Hardware requirements	Monochromatic cameras, hyperspectral cameras and LIDAR
	Repository	Local storage
	Quoting policy	The dataset is required to be cited when used
	Potential users	The potential users will be different stakeholders in the AEC industry, such as building owners, managers, utility- and other building/construction-related service providers, end users, investors, research centres, and any other possible user interested in implementing BIM2TWIN solutions.
Archiving and preservation	Data management	Data will be stored in local server
	Back up	All the data will be backed up in UNIVPM private backup server.
	Archive	All the data will be archived in local UNIVPM private backup server.
	Size	To be defined based on the areas being scanned
	Short/Long term preservation policy	Disseminated data, deliverables and collected data will remain available for the whole duration of the project and also for a period afterwards to be defined. Scientific publications will be public also after the end of the project.
	Preservation costs	To be defined.
	Confidential and sensitive data	Dataset of images and collected experimental data are confidential. Scientific publications will be public and the

		deliverables will have the dissemination level agreed by the consortium.
	Security Standards	*Data security will be ensured by service providers (in the case of social media and Cloud services). Moreover, data will be transferred through secure and encrypted channels.*
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	Risks connected research and dissemination activities could be the dissemination of confidential data outside the consortium.
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	UNIVPM will verify IPR of data related to dissemination material.

Table 9 - UNIVPM's Dataset

6.6 UniSMART – Fondazione Università di Padova

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Dissemination and Exploitation Plan
	Publisher	UniSMART
	Maintainer	UniSMART
	Responsible for data quality	Partners developing core data.
	Responsible for data security	All partners following data security policy.
	Responsible for data sharing	Owner partners and UniSMART.
Data set description	Description	Dissemination, Exploitation and communication data necessary to implement stakeholders' outreach and generate exploitation opportunities for project partners.

	Provenience	Previous WPs, DEC material.
	Type of data	Mainly qualitative (quantitative when required for quantifying specific project results).
	Usefulness	Data used for Dissemination-Exploitation-Communication activities will be useful to promote BIM2TWIN project and its results, with the objective of reaching relevant stakeholders, potential customers, penetrate the market, protect IPR knowledge, maximize the project impact.
	Similar data	For DEC activities, data created during the project will be almost exclusively used, therefore available as long as project proceeds.
	Re-use and integration	The requirement for data re-use is the respect of the IP. Data can be re-used for a wider diffusion of project results or features, towards different markets, to reach a wider audience and to set the basis for further development.
	Data organization	To be defined with data owners when the case, in compliance with D10.1 “Technical quality assurance and risk management plan”. For MS Teams platform storage, commonly agreed standards will be followed. For local storage, specific company's policy will be followed.
File format	Nature	Usable data for DEC activities: text (publications, social media posts), graphic materials (dissemination/communication events, project features, demo sites photographs, leaflet, poster, rollup), video (interviews, presentations, promotional campaigns).
	Format	.doc, .pdf, .jpg, .png, .mp4, .avi, .ai
Standards and metadata	Metadata description	The data will be accompanied by specific information in order to specify data creator, date of creation, data owner and data objective.

	Standards adopted	D10.1 “Technical quality assurance and risk management plan” will be kept as a guideline for data collection processes. Not specific metadata standards will be adopted.
	Quality assurance standards	D10.1 “Technical quality assurance and risk management plan” will be followed to ensure proper quality to managed data.
Data sharing	What data is suitable for sharing and why	DEC data will be shared and spread through different channels in compliance with IPR. They will be shared through dissemination activities, such as social media, publications, exhibitions, promotion campaigns to reach the widest audience.
	License	To be verified with specific data providers (partners).
	Openness/Access policy	Confidential data emerging from project development will be available among partners. Public data will be disseminated, when relevant, and agreed with partners.
	Restrictions	Data owners will specify any data restrictions or necessity for sharing.
	Data availability	DEC data will be shared whenever consented by data owners – in compliance with openness requirements as required by H2020 framework.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	Not specific software or hardware necessities, beside basic applications for text, graphics, video and browser.
	Repository	MS Teams will be used as repository for project related data. An open access repository will be adopted to store publicly disclosable project results.
	Quoting policy	Data owners will specify any quoting policy, if considered.

	Potential users	The potential users will be different stakeholders in the AEC industry, such as building owners, managers, utility- and other building/construction-related service providers, end users, investors, research centres, and any other possible user interested in implementing BIM2TWIN solutions.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	Data will be stored and backed up in MS Teams.
	Archive	Data will be archived in MS Teams and in private servers, when relevant, and in the project website.
	Size	Estimated approx. 50GB.
	Short/Long term preservation policy	Disseminated data will remain available for the whole duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined.
	Confidential and sensitive data	Data owner will specify confidentiality of data and the presence of sensitive data.
	Security Standards	Data security will be ensured by service providers (in the case of social media and Cloud services). Moreover, data will be transferred through secure and encrypted channels.
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	Risks connected to dissemination activities could be the not-effective diffusion of BIM2TWIN project, penetration of the market, project awareness. Dissemination plan will include monitoring and mitigation measures. Partners' risks are to be specified by data owners.
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.

	Copyright and IPR issues	UniSMART will verify IPR of data related to dissemination material.

Table 10 - UniSMART's Dataset

6.7 ORANGE SA

SECTION	GUIDANCE	STATEMENT
Data set reference and name	Name of data set	Digital Building Twin
	Publisher	Orange
	Maintainer	Orange
	Responsible for data quality	Partners pushing data in the Digital Twin
	Responsible for data security	Partners pushing data in the Digital Twin
	Responsible for data sharing	Owner partners and Orange
Data set description	Description	Data shaping the Building Digital Twin for "as planned" and "as performed" views: building objects and relationships, plannings, etc.
	Provenience	Existing data like BIM files or repositories data Collected data like IoT sensors data
	Type of data	Qualitative
	Usefulness	Digital Building Twin is the core feature of BIM2TWIN project.
	Similar data	Digital Building Twin data is built on the basis of existing source data but brings a higher level of abstraction.
	Re-use and integration	The requirement for data reuse is the IP and data proprietary consent.

	Data organization	Data is organized and stored in a graph database.
File format	Nature	Graph database
	Format	Proprietary but the metamodel is based on the NGSI-LD standard, data are modeled with semantic datamodels (ontologies) which are documented and data can be exported in different standard or open formats: RDF, JSON-LD.
Standards and metadata	Metadata description	Metadata associated with the data are data owner, date of creation and date of modification.
	Standards adopted	The metamodel of the Digital Twin is based on NGSI-LD standard, it is free to use any standard semantic datamodel (ontology) to model and describe the Digital Twin.
	Quality assurance standards	To be defined
Data sharing	What data is suitable for sharing and why	The benefit of a Digital Twin is to share information between building stakeholders to improve building construction processes.
	License	To be defined with partners.
	Openness/Access policy	To be defined with partners.
	Restrictions	To be specified by partners owning data.
	Data availability	Digital Twin data aims at being available during project duration, an extension afterwards of Digital Twin data availability

		may be decided by the BIM2TWIN consortium.
	Sharing authorization	Data owners will specify sharing restrictions or specific authorization policy, if needed in compliance with Consortium Agreement.
	Software/Hardware requirements	No specific hardware nor software.
	Repository	The repository for Digital Twin data is the Thing'in platform provided by Orange.
	Quoting policy	Data owners will specify any quoting policy, if considered.
	Potential users	The potential users of shared data are the partners and end users of companies involved in the construction projects of BIM2TWIN pilot sites.
Archiving and preservation	Data management	Data owners will specify the responsible for data preservation and will specify any preservation policy for data management, documentation, curation and preservation.
	Back up	Digital Twin data will be stored and backed up athrough the Thing'in platform.
	Archive	Digital Twin data will be stored and archived in the Thing'in research platform.
	Size	To be estimated.
	Short/Long term preservation policy	Digital Twin data will remain available for the whole

		duration of the project and also for a period afterwards to be defined.
	Preservation costs	To be defined.
	Confidential and sensitive data	Data owner will specify confidentiality of data and the presence of sensitive data.
	Security Standards	The Thing'in platform provides flexible security features able to manage the security policies BIM2TWIN consortium would like to apply.
	Risks to data security (level and likelihood, impact, monitoring, mitigation)	Partner's risks have to be specified by data owners. Thing'in platform provides advanced and flexible security features to mitigate risks to data security.
	Ethics and legal issues	The Grant Agreement and Consortium Agreement are the reference for ethics and legal issues.
	Copyright and IPR issues	Each data owner will verify IPR or copyright of data related to Digital Twin data.

Table 11 - Orange's Dataset

7 CONCLUSION

This deliverable, D9.8, contains the first version of the BIM2TWIN's Data Management Plan. The Data Management Plan outlines essential procedures that must be followed to ensure proper data management needs of the project, such as handling research data, dissemination, protection and exploitation activities. The plan will serve as a clear guideline for smooth and correct data collection, elaboration, diffusion, preservation and reuse, which is fundamental for project implementation and for dissemination, protection and exploitation activities. The Data Management Plan will play a crucial role in handling valuable data and knowledge within BIM2TWIN. It also ensures that different interests of Consortium members and other stakeholders are considered during the various stages of data collection, processing, transmission and storage.

The document also provides significantly important principles and regulations in data management that must be strictly adhered in the project. The first is the FAIR Data Principles, a set of guiding principles to make data Findable, Accessible, Interoperable and Reusable. In terms of privacy, the GDPR regarding data protection and privacy for all individuals within the European Union (EU) and the European Economic Area (EEA) will be strictly adhered. However, other rules can be added to future versions if they are deemed important and compliant with these two principles and regulations.

In order to describe and properly identify the different data to be managed by BIM2TWIN's Consortium partners, the following attributes have been analysed for each dataset:

1. Dataset reference, name and description;
2. Standards and metadata;
3. Data sharing;
4. Archiving and preservation.

In relation to data handling within the project's context, a procedural methodology is prepared to govern the required control and verification in the different steps of data management. The instruction of this methodology is described in detail in Deliverable D10.1 "Technical quality assurance and risk management plan", which will be followed by all the consortium members throughout the project, unless amendments are made to it and agreed by the consortium.

A first collection of the initial data types and the related attributes of data that must be handled within the BIM2TWIN project have been provided by 5 out of 17 partners and reported in dedicated tables. If new data types arise during the project implementation or eventual changes occur on the attributes of the existing data, they will be added subsequently in the Data Management Plan.

Finally, this deliverable will be continuously updated according to the new data and information generated by the project. The dissemination leader will periodically request new updates that imply changes in the Data Management Plan. The Consortium members can also forward the request for changes by communicating to the project coordinator and the dissemination leaders, as soon as any variations occur.

APPENDIX A: LITERATURE

Online Sources

- FAIRsharing community. [Online] Available at www.FAIRsharing.org [Accessed 28/05/2021]

Bibliography

- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., “The FAIR Guiding Principles for scientific data management and stewardship” Sci Data 3, 160018 (2016).