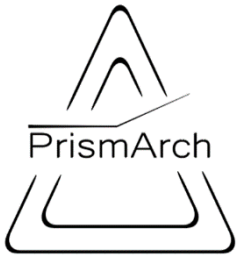

PrismArch

Deliverable No D8.2

Data Management Plan

Project Title:	PrismArch - Virtual reality aided design blending cross-disciplinary aspects of architecture in a multi-simulation environment
Contract No:	952002 - PrismArch
Instrument:	Innovation Action
Thematic Priority:	H2020 ICT-55-2020
Start of project:	1 November 2020
Due date of deliverable:	30 April 2021
Actual submission date:	29 April 2021
Version:	0.6
Main Authors:	Evangelia Vangeli Margariti (CERTH)

Project funded by the European Community under the H2020 Programme for Research and Innovation.

Deliverable title	Data Management Plan
Deliverable number	D8.2
Deliverable version	Final
Contractual date of delivery	30 April 2021
Actual date of delivery	29 April 2021
Deliverable filename	PrismArch_D8.2_v0.6.docx
Type of deliverable	Report
Dissemination level	Public
Number of pages	77
Work package	WP8
Task(s)	T8.2
Partner responsible	CERTH
Author(s)	Evangelia Vangeli Margariti (CERTH), Dimitrios Ververidis (CERTH), Spiros Nikolopoulos (CERTH)
Editor	Evangelia Vangeli Margariti (CERTH)
Reviewer(s)	Antonios Liapis (UOM), Vittoria Bava (Mindesk)

Abstract	This deliverable presents the guidelines for the data collection, storage and ownership for all PrismArch activities. Additionally, it defines and ensures that all collected data are intended to be processed, in a manner relevant and limited solely to the purposes of PrismArch. Finally, it foresees the essential security measures that will be implemented to prevent unauthorized access to personal data or the equipment used for their processing.
Keywords	Data management, GDPR, Data minimisation, FAIR

Deliverable history

Version	Date	Reason	Revised by
v0.1	02/04/2021	Alpha version incorporating the table of contents of the deliverable	Evangelia Vangeli Margariti (CERTH)
v0.2	09/04/2021	Pre-beta version incorporating the data management methodology and requesting for inputs	Evangelia Vangeli Margariti, Dimitrios Ververidis (CERTH)
v0.3	13/04/2021	Version to solicit the inputs from all partners in terms of their datasets	All partners
v0.4	21/04/2021	Beta version circulated for internal review	Evangelia Vangeli Margariti (CERTH)
V0.5	28/04/2021	Final version handed over to coordinator for proof-reading	Evangelia Vangeli Margariti , Dimitrios Ververidis (CERTH)
V0.6	29/04/2021	Final version submitted after proof-reading	Spiros Nikolopoulos (CERTH)

List of abbreviations and Acronyms

Abbreviation	Meaning
AR	Augmented Reality
DMP	Data Management Plan
DoA	Description of Action
EC	European Commission
FAIR	Findable, Accessible, Interoperable, Re-usable
GDPR	General Data Protection Regulation
H2020	Horizon 2020
MR	Mixed Reality
N/A	Not Applicable
VR	Virtual Reality
WP	Work Package
SMEs	Small and medium-sized enterprises
TBD	To be determined

Table of Contents

1. Executive summary

The PrismArch Data Management Plan (DMP) will guide the PrismArch partners in the process of recording and managing the project-related data assets and activities, while identifying and addressing potential emerging security, resource and ethical issues.

A project which aims to combine AEC's industry various functional tools into a common platform that will facilitate decision making and ultimately reshape the design process, presents several data management challenges besides the collection of data. Therefore, the data management lifecycle must be described precisely. The DMP is based on the H2020 Online Manual for *Data Management Plan*[1], which describes the data management life cycle for the data to be collected, processed and/or generated by the project. The methodology proposed by the European Commission Guidelines in line with the EU General Data Protection Regulation (GDPR) has been adopted for the compilation of this deliverable. FAIR (Findability, Accessibility, Interoperability and Reusability) Data Management is highly promoted by the European Commission and relevant attention will be given to these principles for the needs of the project.

Following these guidelines for making research data findable, accessible, interoperable and re-usable (FAIR), this deliverable includes information on:

- The handling of research data (during and after the end of the project)
- The types and formats of data collected (processed and/or generated)
- Methodologies and standards applied
- Data accessibility and restrictions
- How data will be curated and preserved (during and after the end of the project)

Moreover, this deliverable presents the guidelines for data collection, storage and ownership for all PrismArch activities. Additionally, D8.2 will define and ensure that all of the collected data are intended to be processed, in a manner relevant and limited solely to the purposes of PrismArch. Finally, it will foresee the essential security measures that will be implemented to prevent unauthorized access to personal data or the equipment used for their processing.

The main objective for this deliverable is to set the framework and an initial listing of data management activities for the project, presenting the guidelines for data collection, storage and ownership for all project activities. In this context, it will determine the strategy by which the research data generated by the project will be made open, wherever and if possible, to maximize their re-use. Finally, a data categorisation will be made with respect to (i) their confidentiality, ethical treatment

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

and possibility to be shared in the open data initiative, and (ii) their format and standardization.

2. Introduction

AEC industry doesn't possess a significant enhancement of high technology advancements (such as VR, AR, MR) in their current business operations. There may be a vast number of CAD/BIM/CAE Simulation software available, not providing though the essential compatibility that allows for cross-disciplinary collaborative work with natural and intuitive interfaces in a multi-simulation environment.

With PrismArch, the use of interactive technologies will become mainstream in the AEC industry, fulfilling the necessity for an interdisciplinary tool capable to address the unique requirements of architects and engineers from each construction level, enabling them to work on the same architectural project and perceive it in their own different way that best suits their needs. PrismArch aims to create a VR-aided design environment that will be able to host both architects and engineers towards a common goal. By supporting the major disciplines that are typically engaged in an architectural project - namely architects, structural and MEP engineers - PrismArch will enhance the overall decision making process through an action and reaction paradigm. The dynamic collaboration that PrismArch aims to offer, will allow them to iteratively co-decide, preview and evaluate the result of their decisions towards a joint optimal solution. Through advanced simulations embedded within the collaborative VR-aided design environment, superimposed with physical and functional characteristics, the designers will be able to experience in-real time not only how their decisions affect their own discipline but also the other disciplines and consequently the overall architectural project. The introduced AI-assisted design capabilities of PrismArch aim to take designers even closer to their common goal by suggesting scientifically sound design options with respect to quantitative evaluation criteria. Finally, the photorealistic representations of PrismArch will allow the designers to gain insights and obtain a visceral feeling of their creation way long before their actual construction through intuitive interfaces tailored to their individual needs and expertise.

An essential part of the project depends on different types of data (questionnaires and information sheets, data for developing and testing algorithms etc.), which makes the Data Management Plan (DMP) more than important. The DPM aims at defining the management strategy of all data within the PrismArch framework, and describes all activities and procedures to ensure that all data comply with the FAIR principles, following the template and approach recommended by the EC. While we are in favour of making key data assets produced within the project openly available and accessible, data sharing may also be restricted in several cases, taking into account

"*the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions"*[2] .

[2] Guidelines on FAIR Data Management in Horizon 2020

In order to gather all the related information for each of the datasets used (or will be used) during the lifetime of the project, a template (see Annex in section 7) has been sent to all the partners to solicit their inputs.

This document consists of two main parts. In Section 3, we present the general data management methodology of PrismArch-according to H2020 Guidelines and FAIR data- along with the FAIR data financing plan, data security measures and ethical aspects. In Section 4, we present and explain the PrismArch dataset template, and then, using this template, we document the PrismArch datasets – up to the composition date of this deliverable. Finally, section 5 concludes the deliverable. Annex includes the template form shared to all partners in order to fill in the information related to the datasets aligned with the "Template for Horizon 2020 Data Management Plan (DMP)".

3. Data Management Methodology

The methodological approach that has been used to compile this deliverable follows the "Template for Horizon 2020 Data Management Plan (DMP)", version 1.0, released on 13.10.2016 by the European Commission. The PrismArch DMP presented in this deliverable addresses the following aspects of PrismArch data:

- Data summary
- FAIR data
    - Making data findable, including provisions for metadata
    - Making data openly accessible
    - Making data interoperable
    - Increase data re-use
- Allocation of resources
- Data security
- Ethical aspects
- Other issues

In the following subsections, we briefly present the type of questions associated with each of these aspects. For each question we also provide a summary of the general strategy adopted by the project consortium for handling different dataset categories. Detailed answers for each dataset are provided in Section 4.

Updating Methodology and future versions of the PrismArch DMP

Generally, no modifications are expected on the DMP methodology during the project's lifetime, no future updates will be noted. Documents will be stored in the file repository (Google Drive) of the project and will follow the Data Security measures described in section 3.7**.**

3.1 Data Summary

The Data Summary addresses the following issues:

- Outline the purpose of the collected/generated data and its relation to the objectives of the PrismArch project;
- Outline the types and formats of data already collected/generated and/or foreseen for collection/generation at this stage of the project;
- Outline the reusability of existing data;
- Outline the origin of the data;
- Outline the expected size of the data;
- Outline the data utility.

This field describes the data that will be generated or collected, including references to their origin (in cases where data is collected), nature, scale, to whom it could be useful, and whether it underpins a scientific publication. With regard to the individual

questions, our generic DMP approach is summarized below (detailed answers for each dataset are given in Section 4).

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The main goal of PrismArch is to design a new interactive platform that will be tailored to the professional needs (i.e. architects and engineers). In this direction, our plan is to gather continuous feedback from the professional base, starting from the beginning of the project (through collecting the user requirements) to its end (through iterative testing as well as through the envisaged pilots demonstrating the developed technology and the system update based on the users' evaluation). For this vision to unfold, the following **types of dataset** are expected to be used, collected or generated:

- **Requirements analysis data** (in the form of questionnaires, interviews, focus groups, etc.) will be collected, to identify user needs, use case scenarios and desired software functionalities. The objective of collecting such data is to guide the design and development of the PrismArch platform, tools and assets towards the needs of actual users, and support the PrismArch use cases altogether. For instance, such data has already been used in WP6 for T6.1 "Define the architectural projects and usage scenarios for demonstration and evaluation".
- **Evaluation data**, such as user activity and survey data, will be collected from the end users with the aim to assess the impact and effectiveness of the proposed set of apps/tools.
- **Technical data** (existing or generated) will be collected by technical partners in order to develop and test the PrismArch applications and tools. A variety of data will be necessary, including CAD files, images, text, 3D objects etc.
- **Data related to PrismArch dissemination and communication activities**, to allow better organization of events and offer better services to attendees. Video content and photos from participants will also be used for creating dissemination content.
- **Contact data of PrismArch consortium members** (e.g. name, email, organization, etc.) used for project management activities. Selected video-conference calls may be recorded, so there is also audio-visual content of the partners involved in this category
- **Data for system evaluation** will also be collected in the context of WP6, including user questionnaires and automatically generated platform use analytics. Details on these datasets can be found in section 4.

What types and formats of data will the project generate/collect?

The project will use different types of data (CAD files, AEC-related metadata, video, images, text, communication data, system log data, etc.), both personal and non-personal, from a variety of sources (web, partners, testers, etc.) and will probably generate datasets in the process of creating the PrismArch tools and assets.

Will you re-use any existing data and how?

There will be reuse of existing datasets, for instance technical files (e.g., CAD files, knowledge management related documents, etc) will be collected by partners, contributing to the development and benchmarking of algorithms, and further develop and test the PrismArch applications and tools. Other datasets will be reused via machine learning models trained on them, e.g. for AI-assisted content creation and design suggestions a number of open-source datasets can be used for training, indicatively LIFULL HOME'S dataset[3] and RPlan dataset[4]. Note that machine-learned models on existing datasets will not distribute the datasets themselves and the resulting output of models trained on these datasets is expected to be significantly different from the source data.

What is the origin of the data?

The data comes from various origins, including, but not limited to, the following:

- Individual researchers that openly share their data in open repositories such as GitHub, Zenodo and ORE or via their webpages;
- Research and academic organizations that openly share data in open or institutional repositories;
- Open data repositories in order to provide property snapshot data, including image data from floor plans and indoor photographs for each property.
- Use case partners that share data with the technical partners of the consortium to help them define the requirements, setup the demonstrators, or even train and test their algorithms and software;
- Web pages;
- Participants, end users or project partners, after filling out a consent form;
- Questionnaires and surveys filled in by end users (evaluation questionnaires);
- Interviews and focus groups conducted with end users;
- Audio-visual content recorded for the project needs;
- Use of PrismArch software tools by the users (automatically collected data analytics);
- Dataset generated in order to build data driven software components as

[3] https://www.nii.ac.jp/dsc/idr/lifull/

[4] http://staff.ustc.edu.cn/~fuxm/projects/DeepLayout/index.html

part of the project development.

What is the expected size of the data?
Dataset sizes are discussed in section 4.

To whom might it be useful ('data utility')?
The datasets listed in this DMP are necessary to project partners for identifying user and technical requirements for the use cases, designing, developing and testing the PrismArch methodologies, algorithms and tools, and assessing the effectiveness of these tools in real-life trials involving end users. It is also crucial for increasing the project outreach and achieving high dissemination impact. Technical and evaluation data may also be useful to researchers with a focus on the development of similar architect and engineering tools.

The following subsections (3.2-3.5) about making data FAIR refer to the datasets that are (or will be) produced by the project and not those that already exist and are being used by the project.

3.2 Making data findable

This point addresses the following issues:

- Are the data produced in the project discoverable and identifiable?
- What naming conventions are followed?
- Will search keywords be provided that optimize possibilities for re-use?
- Are clear version numbers provided?
- What metadata will be created?

In general, the data collected and generated by the project will be identifiable and discoverable. With regard to the individual questions, our DMP approach is summarized below (again detailed answers for each dataset are given in section 4).

Are the data produced and/or used in the project discoverable and identifiable?
Datasets that will be made publicly available will be uploaded to open repositories like Zenodo etc., thus making it both easily discoverable and identifiable externally. With regard to datasets that will only be used internally in the project, either because of confidentiality reasons, IPR constraints, importance to commercial exploitation or due to limited value to external parties, they will only be discoverable and identifiable by consortium partners or selected institutional users involved in the processing of this data. Consequently, these datasets are not subject to the FAIR data principles.

What naming conventions are followed?

A specific naming convention is suggested to identify the various PrismArch datasets:

PrismArch_<WPno>_<serial number of dataset>_<data type>_<dataset title>

- The <WPno> reveals the WP in the context of which this data is collected or generated and processed
- The <serial number of dataset> is assigned manually in the order of presentation in this deliverable
- The <data type> field is determined according to the categorization presented below:

Acronym	Description
RTD	Supporting research and technical development
PILOT	Resulting from pilot activities

- Finally, the <dataset title> is a descriptive dataset title

Will search keywords be provided that optimize possibilities for re-use?

Keywords will be provided in the cases where this is applicable.

Are clear version numbers provided?

For datasets that will be made publicly available in open repositories, versioning will be supported by appropriate naming conventions.

What metadata will be created?

For datasets that will be shared via open repositories, the metadata standards used by these repositories will be used.

Metadata for data uploaded at the project Google Drive repository is also supported.

In general, Data discoverability will be further enhanced by associating search keywords along with the data, as well as promoting the datasets through the project's communication activities (e.g. blog posts, tweets, etc.). As a part of metadata provision, keywording must comply with the following principles:

- Who, what, when, where and why: these questions must be covered.
- Consistency among the different keyword tags needs to be ensured.
- Keywording must be relevant, understandable and clear.

3.3 Making data openly accessible

This point addresses the following issues:

- Which data produced and/or used in the project will be made openly available as the default?
- How will the data be made accessible (e.g. by deposition in a repository)?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Where will the data and associated metadata, documentation and code be deposited?
- If there are restrictions on use, how will access be provided?
- Is there a need for a data access committee?
- Are there well-described conditions for access (i.e. a machine-readable license)?
- How will the identity of the person accessing the data be ascertained?

With regard to the individual questions about data accessibility, our generic DMP approach is summarized below (again detailed answers for each dataset are given in section 4):

Which data produced and/or used in the project will be made openly available as the default?

Some of the datasets to be used in this project (as described in Section 3.1) is open data already, made openly available by third parties (for example Github). Since this data is already open, as a general policy, PrismArch will not re-share it. Sharing such data will only be pursued in cases where the data license allows it and when PrismArch researchers estimate that re-sharing of the data (in some new form) provides additional benefit to third parties.

In addition to open data, there are also privately owned datasets. These are owned by the organizations involved in PrismArch, as well as some technical and academic partners, and have been collected and created over a period of years or in the context of other projects or internal processes, independently from PrismArch. Such data may be provided to the project for research purposes, but will not be shared openly. However, effort will be made to make this (or part of this) data openly available in cooperation with the data owners, wherever this is possible.

Data that will be collected by the project in the form of questionnaires or forms addressed to end users (for user requirements analysis and/or for evaluation) will not be made openly accessible, since they may contain personal or confidential information. Wherever possible and in case there is added value from their sharing, such data will be anonymized before being shared (mainly with regard to the evaluation data).

The aforementioned data (whether public, private, or personal) will be used exclusively for achieving the project objectives. Where appropriate, the analysis

results will be made open as part of public project deliverables and publications available in open repositories.

Open software tools and publicly available datasets will be shared on GitHub[5], Zenodo[6] as well as on the EU's open access publishing platform – Open Research Europe (ORE)[7] that will allow other researchers to easily crawl and collect data and software from all the open data sources.

How will the data be made accessible (e.g. by deposition in a repository)?

Open data will be deposited in open repositories like Zenodo, GitHub and ORE. The datasets will also be shared through the PrismArch website.

Datasets destined to be used internally by project partners will be stored either on the project's file repository on Google Drive and/or in the servers of project partners.

What methods or software tools are needed to access the data?

Different methods and software tools will be required to access the data depending on the dataset. More details are provided in Section 4 (e.g. web-browser, API).

Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Where this is applicable, the relevant software and its documentation will be included.

Where will the data and associated metadata, documentation and code be deposited? Have you explored appropriate arrangements with the identified repository?

Open data will be deposited in open repositories, like Zenodo, GitHub and ORE. These adopt standard and simple procedures to allow data sharing by researchers. No need for appropriate arrangements is foreseen.

If there are restrictions on use, how will access be provided?

If such cases are identified, access could be provided either through use of consent and anonymization, or by regulating and restricting access to specific users.

[5] https://github.com/
[6] https://zenodo.org/
[7] https://open-research-europe.ec.europa.eu/

Is there a need for a data access committee?
No such need has emerged yet.

Are there well-described conditions for access (i.e. a machine-readable license)?
Such licenses will be used for the data we plan to make openly available.

How will the identity of the person accessing the data be ascertained?
This will be dealt with on a case-by-case basis. For the open datasets, no identification of the person accessing the data will take place. For the data that will be used only internally by project partners (which is stored on the project file repository or partners' servers), access control procedures are in place that define access rights and provide secure access with username/password credentials.

3.4 Making data interoperable

This point specifies what data and metadata vocabularies, standards or methodologies are followed in order to facilitate interoperability. It also addresses whether a standard vocabulary is used for all data types within the dataset, in order to allow interoperability. The specific issues covered are the following:

- Are the data produced in the project interoperable?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Are the data produced in the project interoperable?
Effort will be made to achieve interoperability on most of the data produced in PrismArch.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
In order to ensure interoperability and maximum re-use of PrismArch data, project

partners will try to collect existing and new data in standardized formats, following well-known data representation models and metadata vocabularies (i.e. IFC, RDF/OWL, XLS,OBJ, FBX, Speckle Objects).

Standard and simple data vocabularies will be adopted for different types of datasets (image data, text data, user analytics, etc.). Additionally, we will consult the OpenAIRE Guidelines for Data Archives[8].

Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?
Whenever possible, standardized vocabularies will be used to encourage the wide exchange of information and sharing of data. More specifically, for **architectural data, mechanical and engineering data:** The architectural vocabularies and graphic standards used for project development are based on specified professional standards (e.g. CAD & BIM standards[9]).

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
This will be examined on a case by case basis.

Further actions on making data interoperable will be outlined in subsequent versions of the DMP, as the project progresses. These actions refer to revisiting data and metadata vocabularies, imposing additional standards or methodologies and optimizing interoperability overall.

3.5 Increase data re-use (through clarifying licenses)

- How will the data be licensed to permit the widest re-use possible?
- When will the data be made available for reuse? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- How long is it intended that the data remains re-usable?
- Are data quality assurance processes described?

[8] OpenAIRE Guidelines for Data Archives: https://guidelines.openaire.eu/en/latest/
[9] https://popcdn.azureedge.net/pdfs/POP_CAD_BMI_Stndrds.pdf

With regard to the individual questions about increasing data re-use, our generic DMP approach is summarized below (again detailed answers for each dataset are given in section 4):

How will the data be licensed to permit the widest re-use possible?
This will be examined on a case by case basis depending on the dataset. Our general approach can be summarized as follows: ● In case of data coming from external open sources or in cases where the data comes with a license on its own, the data will be shared under the same license. ● For others, a CC-BY 4.0 (Creative Commons Attribution 4.0 International License) license will be selected, which allows open sharing but also allows keeping some control over the data (e.g. requires attribution). Most common CC –BY 4.0 licensing types are the following: ▪ **Creative commons Attribution-Share Alike 4.0** (CC BY-SA 4.0): any third party can freely copy, distribute, display and modify the datasets for any purpose. Remix, transform, or built upon data, must be distributed under the same license as the original. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. ▪ **Creative Commons Attribution 4.0 International** (CC BY 4.0): any third party can freely copy, distribute, display and modify the datasets for any purpose. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. ▪ **Creative Commons Attribution-NoDerivatives 4.0 International** (CC BY-ND 4.0): any third party can freely copy, distribute, display and modify the datasets for any purpose. Remix, transform, or built upon data, however, must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. ▪ **Creative Commons Attribution-NonCommercial 4.0 International** (CC BY-NC 4.0): third parties can copy, distribute, display and modify the datasets for any purpose other than commercial unless they get a permission by project partners first. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. ▪ **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International** (CC BY-NC-ND 4.0): third parties can copy, distribute, display and modify the datasets for any purpose other than commercial unless they get a permission by project partners first. Remix, transform, or built upon data, however, must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. Licensing will be discussed in later stages of the project with all involved parties.

Alternative license schemes may also be adopted at the discretion of the dataset owner.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
This will be examined on a per case basis. In general, effort will be made for the data to be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
This will be examined on a case-by-case basis (see section 4). The openly shared datasets will be reusable after the end of the project on Zenodo and any additional available platforms/outlets (GitHub, Zenodo, ORE, PrismArch website, partners' websites).

How long is it intended that the data remains re-usable?
The openly shared datasets will in general be perpetually reusable.

Ara data quality assurance processes described?
Automatic *data cleaning* techniques will be employed to improve data quality. Data cleaning consists of identifying incomplete, incorrect, inaccurate, or inconsistent parts of the data and then replacing, modifying or deleting such data. This is necessary for improving data quality and producing a clean, uniform, and consistent dataset for integration; the quality of the data reflects directly upon the quality and accuracy of the analysis results. For datasets including questionnaire data, a manual quality control will be performed by partners to ensure data quality.

3.6 Allocation of resources

This point addresses the following issues:

● **Estimate the costs for making the data FAIR and describe the method of covering these costs;**

- **Identify responsibilities for data management in the project;**
- **Describe costs and potential value of long term preservation.**

Estimate the costs for making the data FAIR and describe the method of covering these costs.

Since the beginning of the design of the project, data management was taken into consideration and every partner has been allocated effort for this purpose. This is embedded into the tasks dealing with data management activities, either collecting, processing, or creating datasets. Hence, all related costs for data management are already covered by the project and no additional resources will be needed. Costs for publications are covered by the project budget. Other costs for making the data FAIR will be covered by the individual partners that will share the data. Zenodo, which is free of charge, will be used to make available papers and datasets (Green Open Access model) under the PrismArch community.

Identify responsibilities for data management in the project.

Regarding the Data management role, although not specifically mentioned in the DoA, a data manager role has been established in the project to ensure that data processing actions within PrismArch are in line with the law. CERTH has been appointed as the beneficiary responsible for data management and has cooperated with technical and pilot partners to draft a detailed data management plan that clearly identifies how each dataset used or created by the project will be handled.

CERTH will be responsible for closely monitoring the execution of the data management plan and ensuring that project partners handle project datasets appropriately.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

Such a discussion is postponed until the full details of platform architecture becomes stable. During the project's lifetime, CERTH's cloud infrastructure will be used to host the intermediate versions of PrismArch's platform and facilitate the necessary developments.

3.7 Data security

This addresses secure storage and transfer of sensitive data as well as data recovery, including the following questions:

- Is the data safely stored in certified repositories for long-term preservation and curation?
- What provisions are in place for data security?

All software tools and data storage mechanisms used within PrismArch are designed to safeguard collected data against unauthorized use and to comply with all national and EU regulations. Engineering best practices and state-of-the-art data security measures along with GDPR legislation will all be incorporated following their respective guidelines and principles.

As explained, PrismArch datasets will either be openly shared (by uploading them in open repositories) or shared internally among specific partners (stored in the project file repository or partners' servers). Below, we examine the data security strategy for these options.

Open repositories

Datasets to be openly shared will be deposited in repositories such as Zenodo or Github that have in place strong mechanisms and protocols for data recovery and long-term data preservation.

PrismArch file repository

In order to be able to share files within the consortium, a Google Drive repository has been set up. Inside the repository, a Work Packages folder with one subfolder for each WP have been created so as to share internally deliverables and relevant documents for each WP. In addition, a data repository folder has been set up to share data such as architecture and engineering layouts and designs etc. Furthermore, folders for sharing data with respect to the project meetings, participation in events, communication kit, literature, templates and meetings have been created. Finally, in the Google Drive folder there is information about project administrative issues such as contract documents, contact details, the mailing list etc.

Google started in the cloud and runs on the cloud, so it's no surprise that they fully understand the security implications of powering the business in the cloud. Because Google and its enterprise services run on the same infrastructure, any organization will benefit from the protections they have built and use every day. Their robust global infrastructure, along with dedicated security professionals and our drive to innovate, enables Google to stay ahead of the curve and offer a highly secure, reliable, and compliant environment.

In order for Google Drive to be protected from unauthorized access, alteration, disclosure, or destruction of information, Google privacy policy[10] includes, but not limited to, the following:

- Encryption to keep data private while in transit
- A range of security features, like Safe Browsing, Security Checkup, and 2 Step Verification to help you protect your account
- Reviewing their information collection, storage, and processing practices, including physical security measures, to prevent unauthorized access to our systems
- Restricting access to personal information to Google employees, contractors, and agents who need that information in order to process it. Anyone with this

[10] https://policies.google.com/privacy?hl=en-US

access is subject to strict contractual confidentiality obligations and may be disciplined or terminated if they fail to meet these obligations

All partners have access to this repository and store content that needs to be shared among the consortium and it is expected that this infrastructure will be used for most project datasets.

Google Drive is in full compliance with GDPR in order to strengthen personal data protection in Europe. In order for that to be achieved, Google supports the GDPR compliance[11] efforts by:

- Committing in their contracts to comply with the GDPR in relation to their processing of customer personal data in all Google Cloud Platform and Google Workspace services
- Offering additional security features that may help the customers to better protect the personal data that is most sensitive
- Giving the documentation and resources to assist any customer in their privacy assessment of google services
- Continuing to evolve google capabilities as the regulatory landscape changes

Partners' servers

PrismArch partners have significant experience in data handling and protection both in the context of their institutional operation as well as in the context of their participation in other H2020 projects. As a result, the beneficiaries already have in place operational policies regarding potential ethics issues as well as privacy and security guidelines for data protection, adhering to national and EU regulations. Ultimately, each partner is responsible for the data protection and security mechanisms in their own servers.

3.8 Ethical aspects

This section covers any ethical or legal issues that can have an impact on data sharing, including references to ethics deliverables and ethics chapter in the DoA. Specifically, it addresses the following issues:

- Are there any ethical or legal issues that can have an impact on data sharing?
- Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

When a dataset cannot be shared, the reasons for this will be outlined (e.g. ethical restrictions, rules governing privacy and personal data protection, intellectual property, and commercial sensitivity).

With regard to the individual questions, our generic DMP approach is summarized below (again detailed answers for each dataset are given in section 4):

Are there any ethical or legal issues that can have an impact on data sharing?

[11] https://cloud.google.com/security/gdpr

Addressing legal and ethics challenges is an important part of the PrismArch work plan. As already indicated in section 5 "*Ethics and Security*" of the DoA, special attention has been paid to these issues since the very beginning of the project.

PrismArch will pay particular attention to any ethical issues that will arise and will address them in a professional way following established EU regulations and corresponding national laws about user privacy, confidentiality and consent. On that direction we have foreseen the PrismArch's Ethics Board in the organizational structure (see deliverable D8.1 [1]), which will be the responsible committee to depict and face all the rising issues that refer to ethics. In detail, the adopted ethical practices are described below.

Should personal data be used as part of the DMP, they will be anonymised, or if that is not possible, they will be pseudonymised according to the current state of the art, and the additional information necessary for re-identifying the individual will be kept separately (according to Art. 4(5) of the GDPR). Pseudonymisation is a permissible measure for data protection in research, according to Art. 89; nevertheless, if possible, further identification should not be possible. Any processing of personal data in PrismArch is covered by the appropriate legal ground. Moreover, all the practices that PrismArch will follow in order to assure the data privacy, protection and confidentiality of the participants and their data generated throughout the project implementation and the usage scenarios can be found in the deliverables D9.1 H - Requirement No. 1 [2] and D9.2 POPD – Requirement No. 2 [3]. The objective of these deliverables was to review all the related legislative framework and present guidelines on how to manage all the project generated data assuring that all ethical, privacy and legal directives are followed by the project.

All personal data that will become available during the project will be kept secure and unreachable by unauthorized entities. The data will be handled with appropriate confidentiality and technical security, as required by law in the individual countries and EU laws and recommendations.

A general policy on ethical conduct will be adopted by the PrismArch Consortium. Prior to the start of relevant PrismArch activities approval form responsible ethics committees will be requested in line with current regulations and guidelines and will explicitly address specifics related to the conduct of analysing personal data, including procedures of (electronic or written) informed consent, remote data collection, user's feedback and privacy and confidentiality and cybersecurity in the data chain of the data and possible data sharing. The PrismArch activities will comply with all applicable national, EU and international legislation, regulations and conventions around the Research on Humans (such as GDPR, Data Protection Directive 95/46/EC of European Parliament and of the Council, OHCHR, etc.).

Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?
As stated in D9.1 H - Requirement No. 1 [2] and D9.2 POPD – Requirement No. 2 [3], in order to ensure safety, the available data will be automatically anonymized or at

least become pseudonymous and will not be transmitted to third parties. Therefore, at least in this phase of the project, there is no need for informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data. In case that, during the lifetime of the project, storing and sharing personal data is needed, the participants will be fully informed through the informed consents.

The consortium guarantees that all personal data collected during the project will be kept secure and unreachable by unauthorized persons. The data will be handled with appropriate confidentiality and technical security, as required by law in the individual countries and EU laws and recommendations, mainly the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/67[12] of the EU.

Before obtaining written consent, information concerning the data processing operations will be handed to pilot participants. The specific information requirements are laid down in Art. 13 and 14 GDPR. Accordingly, in order to provide information to the data subjects in a clear manner and to give the individual participants a genuine choice with regard to the envisaged data processing, the information sheets give research participants information about, inter alia:

- Purposes of data collection, data processing and data analysis;
- Types of personal data processed;
- Transfer of their personal data between their employer/LPA and the relevant technical partner(s), involved in the trials;
- The rights they have as data subjects, and information on how to exercise them;
- The period for which the data will be stored;
- The participation at the research is entirely voluntary and the participants have the right to withdraw from the research at any time without any adverse consequences.

3.9 Other issues

Other issues refer to other national/ funder/ sectoral/ departmental procedures for data management that are used.

In general, all the research organizations and SMEs that participate in the project have in place their own data privacy and security policies, which are compliant with EU regulations and especially the GDPR.

[12] https://www.eugdpr.org/eugdpr.org.html

4. Data management plan for PrismArch datasets

This section includes information on all datasets that can be foreseen as necessary, at the time of writing this deliverable. Every table provides information on the dataset, along with explanations on whether and how this dataset will be FAIR and secure, as far as the datasets produced by the project are concerned. In the first table, we present the structure of the table along with explanations for every field contained. This table template also includes the partner responsible for data collecting and maintaining the dataset, along with an indication on whether these data will be based on existing datasets.

The datasets are categorized under two sections depending on their purpose, including a) supporting research and technical development, and b) resulting from pilot activities. According to this classification we also codify their naming under RTD and PILOT.

A template (see Annex) was shared to all partners in order to fill in the information related to the datasets aligned with the "Template for Horizon 2020 Data Management Plan (DMP)".

Below, we provide a reference Table that briefly summarizes the 17 datasets presented in this section and offers a glance at the structure of the section and its subsections.

Table 4.1: List of datasets addressed in PrismArch

DMP component	WP	Short Summary	Relevant Sub-section
Data collected for supporting research and technical development			**4.1**
RTD_INITIAL-DATASET	WP1	Structured data from small groups of selected experts, for the initial mapping of workflow	4.1.1
RTD_USE-CASE-FILES	WP1	CAD files that have been provided by our end-users to support the use cases	4.1.2
RTD_AUXILIARY-DOC	WP1	Auxiliary documents (e.g., docs, excel files, images) that have been provided by the end-users to convey the knowledge management aspect of their collaborative work	4.1.3
RTD_AI-TRAINING	WP2	Human-AI interaction data used to train designer models for personalized generation.	4.1.4

RTD_BEHAVIOUR-MEASUREMENT-DATA	WP3	Behaviour Measurement Data collected from small groups of selected experts during their interaction with digital interfaces	4.1.5
RTD_INTERVIEW-RECORDINGS	WP3	Data collected from interviews from small groups of selected experts that will help define UX and usability guidelines in VR-aided design environments	4.1.6
RTD_INTERVIEW-TRANSCRIPT-EPISODES	WP3	Dataset derived from Interviews on 4.1.6 that will be used to define UX and usability guidelines in VR-aided design environments	4.1.7
RTD_VR-INTERFACES	WP4	3D-objects (geometries, materials and textures) used to support the interfaces developed within the VR environment	4.1.8
RTD_CAE-SIMULATIONS	WP4	Data files (e.g. BIM files) used to support CAE-based simulations	4.1.9
RTD_KNOWLEDGE-BASE	WP4	Data related to architectural design and 3D models coming from architectural design tools and VR environments	4.1.10
RTD_SPECKLE-DATABASE	WP5	Data related to the architectural design and integration protocol	4.1.11
RTD_ DEVELOPMENT-DISSEMINATION-FILES	WP7	Data related to PrismArch platform development that can be used for creating dissemination content	4.1.12
Data collected resulting from pilot activities			**4.2**
PILOT_USER-REQUIREMENTS-QUESTIONNAIRES	WP6	Structured questionnaires that apply to each design, for the collection of user requirements during the pilot activities	4.2.1
PILOT_EVALUATION-DATA	WP6	Structured questionnaires, for the evaluation of the	4.2.2

		developed tools	
PILOT_EVALUATION-LOG-FILES	WP6	Log files that will be automatically created during pilot evaluation sessions	4.2.3
PILOT_DATA-SOURCES-QUESTIONNAIRE	WP6	Small groups of selected experts will be engaged in experimental settings for evaluating the user experience and spatial cognition aspects while using the PrismArch platform	4.2.4
PILOT_EVALUATION-DISSEMINATION-FILES	WP7	Data related to PrismArch dissemination and communication activities, to allow better organization of events and offer better services to attendees.	4.2.5

4.1 Datasets for the development of technologies

4.1.1 Initial Dataset

NAME	**PrismArch_WP1_001_ RTD_INITIAL-DATASET**
Data summary	**Responsible partner: ZHVR, AKT, SWECO** **Purpose:** Small groups of selected experts will participate in small-scale user studies (e.g. focus groups, interviews, questionnaires) to help identify the limitations of existing workflows in architecture and provide guidance to researchers and developers with respect to principles, rules, restrictions and interconnections within and across design disciplines as well as with respect to cross-disciplinary perspectives and multi-simulations. It will help us define a list of requirements, including both general requirements concerning the PrismArch technical functionalities and features but also requirements focusing on the specific use cases. The findings of the inquiries will be summarized in deliverable D1.3 [5]. **Type/format**: Excel documents containing questions and user responses, as well as versions of the Incident Sheets (generated in WP1) that are marked up by the interviewers and focus group of participants. **Re-use of existing data:** The data is original

	Data origin: Questionnaires filled by ZHVR, AKT and SWECO employees (shared through professional network platforms) **Expected size:** A few MB in total **Data utility:** This data will be used to report the cross-discipline principles rules, constraints, and interfaces definition for cross-disciplinary and multi-simulation perspectives in VR in WP1. Their insights will be shared to technical partners for developing PrismArch's platform. Therefore, these files address all technical partners.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The files are stored on a project's google drive folder, in which only the partners have access. So the raw data is not discoverable for third parties outside the project. The aggregated knowledge from this data will be made accessible to the public through D1.3 [5]. **Search keywords:** N/A **Versioning:** Project's google drive folder supports versioning **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, analysis results based on this data was shared with the consortium (only) and the findings of the questionnaires will be part of the D1.3, but no sensitive information was gathered since the interviews / questionnaires were anonymous and only demographic non-identifying questions were included (age and gender). **How it will be accessible:** Stored in google drive and it is **only** internally accessible by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Web-browser (only by project partners) **Repository:** The raw questionnaires are stored in Google Drive. D1.3 [5] will be available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A

Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally. **Availability for re-use:** This data is not expected to be re-used. It will be only used from partners to specify PrismArch's user requirements. The aggregated knowledge will be available on D1.3 [5]. Third parties may use it after giving appropriate credit. **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5. Moreover, given the small sample of the questionnaire a manual control was performed by ZHVR, AKT and SWECO to ensure data quality.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data will be made publicly available through D1.3 [16]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires were informed on it by adding the data usage information at the start of the questionnaires)
Other issues	N/A

4.1.2 Use Case Files

NAME	**PrismArch_WP1_002_ RTD_USE-CASE-FILES**
Data summary	**Responsible partner: ZHVR, AKT and SWECO** **Purpose:** CAD files that have been provided by our end-users to support the use cases. It will help us define a list of requirements, including both general requirements concerning the PrismArch technical functionalities and features but also requirements focusing on the specific use cases. **Type/format:** CAD files, excel and word documents. Vector-graphics representation floorplans that will be retrieved from open repositories will include the geometric and semantic information and are stored as pixel image. **Re-use of existing data:** The data is original **Data origin**: Files provided by ZHVR, SWECO and AKT employees. A number of open-source datasets are considered in order to provide property snapshot data, including image data from floor plans and indoor photographs for each property. **Expected size:** 5-10 GB **Data utility:** Data with 3D & 2D designs as well as documents with information regarding materials, processes, tolerances, and other useful information regarding all the necessary processes from design to production will be collected. This data will be used to report the cross-discipline principles rules, constraints, and interfaces definition for cross-disciplinary and multi-simulation perspectives in VR in WP1.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable:** The files are stored on a project's google drive folder, in which only the partners have access. So the raw data is not discoverable for third parties outside the project. **Search keywords:** N/A **Versioning:** Project's google drive folder supports versioning **Metadata creation:** N/A

Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, analysis results based on this data was shared with the consortium (only) and the findings will be part of the D1.3 [5]. **How it will be accessible:** Stored in google drive and it is only internally accessible by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data:** Web-browser (only by project partners) **Repository:** The raw files are stored in Google Drive. D1.3 [5] will be available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies:** The architectural vocabularies and graphic standards used for project development are based on specified professional standards (Chappell D. & Dunn M., (2016) [6], The American Institute of Architects, (2016) [7], Collective Work, (2013) [8], CAD & BIM standards[13]). **Use of standard vocabularies:** Specified professional standards used in CAE industry **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be only used from partners to specify PrismArch's user requirements. **Usable by third parties after end of project:** N/A **Re-use timeframe:** N/A

[13] https://popcdn.azureedge.net/pdfs/POP_CAD_BMI_Stndrds.pdf

	Data quality assurance process: Raw data is cleaned and pre-processed as described in section 3.5.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information. Only aggregated forms of the data will be made publicly available through D1.3. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared and the providers of data were informed about the data usage.
Other issues	N/A

4.1.3 Auxiliary documents

NAME	**PrismArch_WP1_003_ RTD_AUXILIARY-DOC**

Data summary	**Responsible partner: ZHVR, AKT, SWECO** **Purpose:** Auxiliary documents will be provided by the end-users to convey the knowledge management aspect of their collaborative work. They will contribute to the definition of list of technical and use case requirements concerning the PrismArch functionalities**.** **Type/format:** Excel and word documents, images, example code. Vector-graphics representation floorplans retrieved from open repositories will include the geometric and semantic information and are stored as pixel image. **Re-use of existing data:** The data is original **Data origin:** Files provided by ZHVR, AKT and SWECO employees. A number of open-source datasets are considered in order to provide property snapshot data, including image data from floor plans and indoor photographs for each property. **Expected size:** A few MB in total **Data utility:** This data will be used to report the cognitive issues and usability guidelines that will contribute to the definition of technical and use case requirements concerning the PrismArch functionalities.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable:** The files are stored on a project's google drive folder, in which only the partners have access. Raw data is not discoverable for third parties outside the project. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A

Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, analysis results based on this data was shared with the consortium (only) and the findings will be part of the D1.3 [5]. **How it will be accessible:** Stored in google drive and it is only internally accessible by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data:** Web-browser (only by project partners) **Repository:** Data files are stored in Google Drive. D1.3 [5] will be available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies:** N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies:** N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be only used by partners to define PrismArch's technical features **Usable by third parties after end of project:** N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5.
Allocation of	**Costs for making the data FAIR:** N/A

resources	**Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data will be made publicly available through D1.3 [5]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires were informed on it by adding the data usage information at the start of the questionnaires).
Other issues	N/A

4.1.4 AI Training and Designer Modelling

NAME	**PrismArch_WP2_004_ RTD_AI-TRAINING_AND_DESIGNER_MODELING**

Data summary	**Responsible partner:** **UOM** **Purpose:** Human-AI interaction data used to train designer models of preference and style, which can guide the generation of personalized suggestions. The data will contribute to the definition of requirements for personalized AI suggestions which are a research output of PrismArch and a part of the PrismArch functionalities**.** **Type/format:** Human-AI interaction data will be logged during dedicated user testing sessions during the demonstrators (see WP6). Data logged specifically regarding the human-AI interaction includes number of times a suggestion was preferred or interacted with, number of times a user requested for a new suggestion, the actual files of suggestions selected and suggestions non-selected, the user's own discipline or non-identifying demographics (e.g. gender and age), the context of other activities they were performing before and after their interaction with AI, etc. Data retrieved from open repositories such as vector-graphics representation floorplans will include the geometric and semantic information and are stored as pixel image. **Re-use of existing data:** The data will be collected as part of the PrismArch project, and no existing data will be reused for this task. **Data origin**: External, large-scale datasets of floorplans will be used for the purposes of machine learning, to drive AI-assisted content creation and suggestions as part of WP2. A number of open-source datasets are considered for training, indicatively LIFULL HOME'S dataset[14] and RPlan dataset[15]. LIFULL HOME'S dataset is provided to researchers by the National Institute of Informatics[16] from LIFULL Co., Ltd.[17] (formerly NEXT Co., Ltd.) and includes rental property snapshot data, including metadata such as location, age, or rent as well as image data from floor plans and indoor photographs for each property. The RPlan dataset, initially introduced in the publication by Wu et al. [17], is a large-scale dataset of floor plans from residential buildings with semantic annotations at the pixel level. All floor plans in the

[14] https://www.nii.ac.jp/dsc/idr/lifull/
[15] http://staff.ustc.edu.cn/~fuxm/projects/DeepLayout/index.html
[16] https://www.nii.ac.jp/
[17] https://translate.google.com/website?sl=ja&tl=en&ajax=1&u=https://LIFULL.com/

	RPlan dataset have no copyright issue. **Expected size:** A few MB in total **Data utility:** Human-AI interaction data will be used to train user models specific to a designer or broader user models based on the specialization/discipline of the users. The designer models will influence the priorities of the AI generators and will produce personalized suggestions.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable:** The dataset is stored on a project's google drive folder, in which only the partners have access. Raw data is not discoverable for third parties outside the project. The aggregated knowledge from this data was made accessible to the public through D2.2 [9] and D2.3 [10]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, analysis results based on this data will be shared with the consortium (only). The aggregated knowledge from this data was made accessible to the public through D2.2 [9] and D2.3 [10]. Data retrieved from open data repositories have already been used for machine learning purposes outside of PrismArch. Machine-learned models on these datasets will produce novel output that is significantly different from the source data, and special care will be given that the generative models do not recreate any of the floorplans from the training datasets. **How it will be accessible:** Stored in google drive and it is only internally accessible by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data:** Web-browser (only by project partners). **Repository:** Data files are stored in Google Drive. D2.2 [9] and D2.3 [10] will be available on the project website.

	Restrictions on access: Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies:** N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies:** N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be only used by partners to define PrismArch's technical features **Usable by third parties after end of project:** N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).

Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data are made publicly available through D2.2 [9] and D2.3 [10]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires and logged interactions were informed on it by adding the data usage information at the start of the questionnaires).
Other issues	N/A

4.1.5 Behaviour Measurement Data

NAME	**PrismArch_WP3_005_RTD_BEHAVIOUR-MEASUREMENT-DATA**
Data summary	**Responsible partner:** ETH **Purpose:** Participants (e.g. small groups of selected experts) will be engaged in psychometric tests and in experimental settings collecting behavioural data including but not limited to, reaction times and physiological signals during user-interaction with digital interfaces, with the intention to understand user-interaction with, and assess the cognitive load stemming from the use of PrismArch platform. We foresee such measures to provide an empirical basis for evaluating the requirements, including both general requirements concerning the PrismArch technical functionalities and features but also requirements focusing on the specific use cases. The findings of the questionnaires will be summarized in D3.2 [12] and D3.3 [13]. **Type/format**: Excel and word documents, Compressed Video, Audio streams **Re-use of existing data:** The data is original **Data origin:** Behavioural tracking of user interactions with the immersive environments. **Expected size:** A few MB in total

	Data utility: This data will be used to report the cognitive issues and UX and usability guidelines in VR-aided design environments in WP3. Their insights will then be shared to technical partners for developing PrismArch's platform.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: Data is stored on the project's google drive folder, in which only the partners have access. Raw data will be anonymised during data collection, and will not be discoverable for third parties outside the project. The aggregated knowledge from this data will be made accessible to the public through D3.2 [12] and D3.3 [13]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. No sensitive information was gathered since the questionnaires were anonymous and only demographic non-identifying questions were included (age and gender). After aggregation and processing, the knowledge from this data will be made accessible to the public through D3.2 [12] and D3.3 [13]. **How it will be accessible:** Stored in google drive and it is **only internally accessible** by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Web-browser (only by project partners) **Repository:** The raw questionnaires are stored in Google Drive. D3.2 [12] and D3.3 [13] will be available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be only used by partners to define PrismArch's technical

	features **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5. Moreover, given the small sample of the questionnaire a manual control was performed by ETH to ensure data quality.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data will be made publicly available through D3.1 [11]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires were informed on it by adding the data usage information at the start of the questionnaires)
Other issues	N/A

4.1.6 Interview Recordings

NAME	**PrismArch_WP3_006_RTD_INTERVIEW-RECORDINGS**
Data summary	**Responsible partner:** ETH **Purpose:** Data collected through interviews with selected experts, that will be used to define UX and usability guidelines in VR-aided design environments in WP3. **Type/format**: Compressed Video, Audio streams (various container formats) **Re-use of existing data:** The data is original

	Data origin: Interview-type conversations, recorded using a video-conferencing platform (mostly zoom). **Expected size:** 10 to 50 GB **Data utility:** This data will be used to define UX and usability guidelines in VR-aided design environments in WP3. The recordings will be transcribed and annotated by the researchers (cf. 4.1.7 Interview Transcript Episodes).
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The dataset is stored locally on external storage media and on servers that are located on ETH premises. This type of raw data is considered sensitive information. It is, hence, not discoverable for third parties outside the project. The aggregated knowledge from this data will be made accessible to the public through D3.1 [11]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, the knowledge from this data will be made accessible to the public through D3.1 [11]. **How it will be accessible:** The dataset is stored locally on external storage media and on servers that are located on ETH premises. This type of raw data is considered sensitive information. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Video player (e.g. VLC), Video cutting applications (e.g. Adobe Premiere. VLC, Adobe and Adobe Premiere). **Repository:** The dataset is stored locally and on servers that are located on ETH premises. D3.1 [11] will be available on the project website **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally

	Availability for re-use: This data is not expected to be re-used. It will be only used by partners to define PrismArch's technical features **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures** The dataset is stored locally on external storage media and on servers that are located on ETH premises. This type of raw data is considered sensitive information. Access is restricted to ETH team members immediately involved in the data analysis activities. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are not anonymised as faces and voices of researchers and interview partners can be identified from the video/audio data. However, only aggregated or anonymised forms of the data are made publicly available through D3.1 [11]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Data providers were informed regarding the data's usage)
Other issues	N/A

4.1.7 Interview Transcript Episodes

NAME	**PrismArch_WP3_007_RTD_INTERVIEW-TRANSCRIPT-EPISODES**
Data summary	**Responsible partner:** ETH **Purpose:** The transcripts are derived based on dataset Interview - Recordings (cf. 4.1.6). This data will be used to define UX and usability guidelines in VR-aided design environments in WP3. Their insights will then be shared to technical partners for developing PrismArch's platform. The findings will be summarized in D3.1 [8]. **Type/format**: Text/Markdown, Text/RTF, Text/Word Processing Document **Re-use of existing data:** The data is based on dataset Interview -

	Recordings (cf. 4.1.6). **Data origin:** Manual transcription of selected episodes of the recorded conversations in Interview - Recordings (cf. 4.1.6). **Expected size:** A few MB in total **Data utility:** The transcripted episodes will inform the empirical work in WP3. Insights will then be shared to technical partners for developing PrismArch's platform. Furthermore, a subset of the collected episodes informs the usage scenario development (WP6).
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: The dataset is stored locally and on servers that are located on ETH premises. Raw data is not discoverable for third parties outside the project. The aggregated knowledge from this data will be made accessible to the public through D3.1 [11]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. No sensitive information will be included; during transcription the researchers will apply codes so as to anonymise names, company names, personal information and other data that could reveal the identity of the participants or their organisations. After aggregation and processing, the knowledge from this data will be made accessible to the public through D3.1 [8]. **How it will be accessible:** The dataset is stored locally and on servers that are located on ETH premises and it is **only internally accessible** by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Web-browser, word processing applications, text editors (only by project partners) **Repository:** The raw questionnaires are stored in Google Drive. D3.1 [11] will be available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
FAIR Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A

FAIR Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally. **Availability for re-use:** This data is not expected to be re-used. It will be only used by partners to define PrismArch's usability and UX features. **Usable by third parties after end of project**: No such use is foreseen at the time of writing of this report. The dataset is re-usable as an empirical basis for investigating typical work practices in contemporary architectural and engineering design practices. **Re-use timeframe:** cf. above. **Data quality assurance process:** Researchers at ETH manually ensure the use of consistent formatting conventions for direct citations, annotations by the researchers, references to original sources such as interview sessions as anonymised codes, video and audio time-stamps. These formatting conventions are available together with the transcript data as it is shared internally with the project partners.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The dataset is stored locally and on servers that are located on ETH premises. Raw data is not discoverable for third parties outside the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying information (such as years of professional experience in a certain field). Only aggregated forms of the data will be made publicly available through D3.1 [11]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the interviews were informed on it by adding the data usage information at the start of the interview)
Other issues	N/A

4.1.8 VR - Interfaces

NAME	**PrismArch_WP4_008_RTD_VR_Interfaces**

Data summary	**Responsible partner: CERTH** **Purpose:** The purpose is to create a customizable virtual environment for the PrismArch project space. Data include conceptual drawings and sketches, as well as CAD and vector designed virtual interfaces exports (e.g. inventory and tools' widgets for VR immersed design manipulation). Furthermore, data include some user experience guidelines for architectural VR tools design, based on existing and original research by the partners. **Type/format**: Image and diagrams vectors, 3D CAD models, rendered images, wireframes,3D objects library. **Re-use of existing data:** Data is a result of inspiration from ready-made vectors and new concept creation in alignment with the use case partners ZHA, AKT, SWECO and ETH. **Data origin:** Conceptual drawings and diagrams from use case partners, existing literature review and original data creation for the purpose of PrismArch. **Expected size:** The data might vary from 5 MBs to 1 TB. **Data utility:** Internally, the data will be useful to CERTH, Mindesk and UoM for VR environment and tools developing and for the disciplines that collaborate inside VR space for the design manipulation. Possibility exists that data will also be experientially useful for third parties reviewing the PrismArch tool.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: Visuals for the development of VR tools and environment are created for the sole purpose of the PrismArch project and are customized as such. Visuals are extracted from software such as Adobe Illustrator, Photoshop, Sketch, Rhino, Cinema 4D, Unreal Engine, Figma etc. Any variant user experience research and interfaces visual results that will be produced out of the original models will be investigated if it is possible to be published in relevant research papers and VR or UX design related Journals and forums. **Search keywords:** VR Interfaces, VE Interfaces, UI, UX **Versioning:** Consecutive variations will be applied according to amendments **Metadata creation:** Metadata standards will be in conformed with basic publishing visuals standards. Examples of those standards are .ai from Adobe Illustrator and exporting formats such as .jpg,.tiff,.png,.eps,.pdf. Moreover, based on the serialization service of Unreal Engine, the VR interfaces will be saved as .uasset, the standard Epic Games' file format.

FAIR Data: **Accessibility**	**Data openly accessible:** Certain visuals of the VR space that prove the concept of the project will be openly accessible through the dissemination platforms and actions. However, several of the architectural projects that are used as a baseline for the project and are showcased in the virtual environment are held in a non-disclosure agreement between the use case partners and their clients. In this context, these data cannot be exposed, e.g. the interior design of well-known buildings that already exists, due to security reasons. If any variants stem out of these buildings, they can be disclosed as public data. **How it will be accessible:** The data will be available through architectural research repositories such as RIBA, top architectural curators such as ArchDaily[18], UX digital publishers such as Medium[19] or as supplementary material in journals such as Architectural Design Wiley[20] and Architects' Journal[21]. **Methods/software tools to access data**: The access to the data can be achieved with widely used commercial software such as Adobe Illustrator or Rhino, Cinema 4D but also from open source software such as Adobe Reader, EPS Viewer, OpenOffice Draw, LibreOffice Draw, GIMP, Vectr, Blender, Unreal Engine. **Repository:** VR projects are complex structures that use a mix of binary assets and source code files. VR projects with their respecting interfaces and source code will be saved in public repositories (e.g Gitlab), that pose no restrictions in terms of the supported data formats. **Restrictions on access:** The Unreal Engine projects and VR interfaces will be publicly available for use through the MIT license. However, already existing proprietary software and source code plugins from consortium partners will remain private due to protection rights.
FAIR Data: **Interoperability**	**Interoperability:** The development tools used for the VR interfaces follow a cross-functional and reusable pattern plan, which enables produced assets to be shared and utilized among different projects and authors. Therefore, data exchange can be considered as a feasible action concerning the developed VR interfaces and their respecting data (visual and code). **Data and metadata vocabularies**: .fbx, .eps, .ai, .uasset

18 https://www.archdaily.com/
19 https://medium.com/topic/design
20 https://onlinelibrary.wiley.com
21 https://www.architectsjournal.co.uk/

	Use of standard vocabularies: The aforementioned development tools use a standard vocabulary and naming conventions, allowing developers and content creators to communicate through a shared structured language. The same vocabulary will be employed for the development of the VR interfaces, making cross-disciplinary communication easier. **Mappings to commonly used vocabularies**: Not necessary
FAIR Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally. Only aggregated knowledge will be available on open source repositories (e.g. Gitlab), where user's credential verification fully complies with the European and international framework and the GDPR (see section 3.7). **Availability for re-use:** After the end of the project, data will be available for re-use. **Usable by third parties after end of project**: Not restricted. **Re-use timeframe:** Permanent **Data quality assurance process:** The development tools that will be used, hold a wide variety of quality assurance tools that automatically sanitize and ensure the desired quality of the produced data. Moreover, manual testing processes will be followed during the pilot versions of the produced software that will exploit possible problems and inaccuracies.
Allocation of resources	**Costs for making the data FAIR:** Costs of submitting to a journal covered by first author organization will be covered by the individual partners that will share the data. **Costs for long-term preservation:** TBD
Security	**Security measures:** The datasets will be stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long term preservation given:** The professional network providing the data will be fully informed regarding the data usage and share.
Other issues	N/A

4.1.9 CAE Simulations

NAME	PrismArch_WP4_009_RTD_CAE_Simulations
Data summary	**Responsible partner:** AKTII **Purpose:** Provide building simulation results such as structural analysis simulations (Finite Elements Analysis); Heating simulations; or Annual Daylight per year simulations, in order to be visualized within VR environment. This kind of data has no significant meaning to be exposed as free data as they are mainly simulation results for certain elements of buildings. We describe though the conditions of providing this data as open source for any possible reuse. **Type/format**: Arithmetic results, coloured meshes or voxels. **Re-use of existing data:** AKTII and SWECO have already extracted simulation results for projects "One Park Drive Tower" and "One Thousand Museum". **Data origin:** Re-produced variants of previous projects such as "Private Residential Villa", "One Park Drive Tower", "One Thousand Museum", and "Bankside Yards West" by AKTII, ZH, and SWECO use case partners. **Expected size:** The data might vary from 10 MBs to 1 TB. **Data utility:** The data will be useful for all AEC disciplines that collaborate inside VR space but most essentially useful to Structural and MEP Engineers.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: Simulation data related to the aforementioned residential and commercial projects contain sensitive information that can not be disclosed. Simulation results are extracted from software such as SAP2000, Sofistik, Rhino-Grasshopper plugins (HoneyBee and LadyBug). Any variant simulation results that will be produced out of the original models will be investigated if it is possible to publish them in RIBA (Royal Institute of British Architects - RIBA[22]) in "Resources" page, or as supplementary material in Journals papers. **Search keywords:** "Simulation results, Architecture, Structural Analysis, Efficient Energy Consumption" **Versioning:** Consecutive variations will be applied according to amendments **Metadata creation:**: Standards will conform main simulation software

[22] https://www.architecture.com

	such as .gh for HoneyBee -LadyBug and .sdb for SAP2000.
FAIR Data: **Accessibility**	**Data openly accessible:** Several of the architectural projects that are used as a baseline for the project are held in a non-disclosure agreement between the use case partners and their clients. In this context, these data cannot be exposed, e.g. the interior design of well-known buildings that already exists, due to security reasons. If any variants stem out of these buildings, they can be disclosed as public data. **How it will be accessible:** The data will be available through architectural research repositories such as RIBA or as supplementary material in journals such as Architectural Design Wiley[23] . **Methods/software tools to access data**: The access to the data can be achieved with widely used commercial software such as SAP2000, SOFISTIK, Rhino-Grasshoper or open source software such as FreeCAD, OpenCascade, and OpenStudio (see further analysis on D4.1 [21]. **Repository:** Simulation data are already meta-data extracted from original 3D models. **Restrictions on access:** N/A
FAIR Data: **Interoperability**	**Interoperability:** Simulation data have already existing well known formats such as .sdb (SAP2000) or STEP, IGES (Open source format). **Data and metadata vocabularies**: Simulation file formats are per se vocabularies and typologies of metadata. **Use of standard vocabularies:** .SDB, .STEP, .IGES **Mappings to commonly used vocabularies**: Not necessary
FAIR Data: **Reusability**	**License:** Attribution 4.0 International (CC BY 4.0) **Availability for re-use:** After the end of the project, data will be available for re-use **Usable by third parties after end of project**: Not restricted **Re-use timeframe:** Permanent **Data quality assurance process:** Data will be produced by senior qualified engineers of AKT and SWECO
Allocation of resources	**Costs for making the data FAIR:** Costs of submitting to a journal covered by first author organization **Costs for long-term preservation:** TBD
Security	**Security measures:** Data will regard imaginary architectural projects so as to avoid any security issues

[23] https://onlinelibrary.wiley.com

Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long term preservation given:** The professional network providing the data will be fully informed regarding the data usage and share
Other issues	N/A

4.1.10 Knowledge Base

NAME	**PrismArch_WP4_010_ RTD_KNOWLEDGE-BASE**
Data summary	**Responsible partner: CERTH** **Purpose:** Providing an effective semantic representation and storage of metadata coming from multimodal VR environments and from architectural design data repositories. The vocabulary will follow the ontological structures which will be defined in T4.1: Interconnect VR-aided design environment with BIM and CAE Simulation environments. **Type/format**: Data will be stored in RDF format. **Re-use of existing data:** Not intended for the time being. **Data origin:** Data will be provided by all partners after their involvement with several VR environments, as well as from architects who operate with architectural design tools (Rhino, Grasshoper, Revit, spreadsheets, databases). **Expected size:** The size will depend solely on the number of models that will be stored in the Knowledge Base. **Data utility:** Data will be used to support queries execution on top of the triplestore, according to the defined scenarios.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A

Fair Data: **Accessibility**	**Data openly accessible:** Data will be accessible for internal purposes in terms of PrismArch project. Raw data is not discoverable for third parties outside the project. **How it will be accessible:** Data will be accessible for internal purposes in terms of PrismArch project. **Methods/software tools to access data**: Project partners will be able to access the data using the Knowledge Base URL, while storage and information retrieval will be available using the services that will be developed. **Repository:** The dataset will be stored in GraphDB triplestore, which will be installed on CERTH's server. **Restrictions on access:** Project partners will be able to access the stored information using the developed services that retrieve semantic information and have access to GraphDB workbench mode (if needed).
Fair Data: **Interoperability**	**Interoperability:** Existing ontologies will be reused to promote interoperability. **Data and metadata vocabularies**: Data will be stored in the Knowledge Base in terms of RDF triples, following the Resource Description Framework (RDF) format. **Use of standard vocabularies:** Existing ontologies and standards will be used to formulate the PrismArch ontology. Web Ontology Language (OWL) will be used to represent the developed ontology. **Mappings to commonly used vocabularies**: The generated mappings will be based on existing ontologies and vocabularies.
Fair Data: **Reusability**	**License:** Since data are intended for internal use only, they will not be licensed. **Availability for re-use:** Not intended for the time being. **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Several evaluation methods could be provided in terms of ontologies and semantic data

	management.
Allocation of resources	**Costs for making the data FAIR**: N/A **Costs for long-term preservation**: N/A
Security	**Security measures:** GraphDB will be hosted in CERTH's server. A basic user authentication (including a username and password) will be required in order to access the repository to ensure that the triplestore is secure.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long term preservation given:** N/A
Other issues	N/A

4.1.11 Speckle Database

NAME	**PrismArch_WP5_011_ RTD_VR-SIM-INTERCONNECTION**
Data summary	**Responsible partner: CERTH** **Purpose:** Data related to the architectural design and integration protocol for VR-aided design environment and its interconnection with BIM and CAE Simulation. This data is actually the 3D models of imaginary residential and commercial buildings (the 2 use cases) that Architects and Engineers of PrismArch will design. **Type/format**: Speckle JSON format **Re-use of existing data:** The data is original **Data origin:** Data provided by use case partners during designing a residential and a commercial building **Expected size:** 300 MBs to 1TB per each of the 2 use cases. **Data utility:** This data will be used as a benchmark data for developing and testing the final prototype of the VR-aided design platform.

Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The data regards imaginary residential and commercial buildings. They will be designed by Architects and Engineers of PrismArch, having their names pseudonymised, while only members of the project will have access to these files. The data will be stored in a CERTH server or in a server in UK, as all use case partners are in UK, in order to speed up downloading/uploading speeds through internet. Towards using a server inside the project, a CERTH server with a Speckle server software instance will be used. Raw data is not discoverable for third parties outside the project. In case a server in UK will be used, we will use Speckle.xyz UK server[24] (of Speckle Systems Company[25]) that provides its services under rules that comply with the European Law regarding privacy[26] issues and terms of use[27]. The case of using the official Speckle.xyz server is preferred because data can be accessed also from wide audiences such as Architects and Engineers for free. Also, this solution reduces the costs of maintenance of the server from CERTH. Each partner of PrismArch will be responsible of the content it uploads to this server. **Search keywords:** Speckle.xyz, PrismArch Stream **Versioning:** Consecutive variations will be applied according to commit number **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible after the end of the project. **How it will be accessible:** It will be accessible in Speckle.xyz server as it will be made publicly open without the ability to append or delete. **Methods/software tools to access data**: Through Speckle interconnection software (Web) and plugins for Rhino, Revit, and Grasshoper, the data will be wide available to Architects and Engineers. **Repository:** The data will be preferable stored in Speckly.xyz server as it allows to be visible to wide audiences. The case of storing the

[24] https://speckle.xyz/
[25] https://speckle.systems/
[26] https://speckle.systems/privacy/
[27] https://speckle.systems/terms/

	data in CERTH's instance of Speckle server is weak as CERTH has limited resource for maintaining this server and it also has limited visibility. **Restrictions on access:** Shared among project partners with access to project's Google Drive and Speckle's database. Speckle has its own authentication mechanism protected with personal passwords. PrismArch members will be Collaborators inside Speckle that can be assigned by PrismArch manager (terminology: PrismArch stream administrator).
Fair Data: **Interoperability**	**Interoperability:** Through Speckle software the data can be used from all the connectors supported by Speckle, which are currently Rhino, Revit, Grasshopper and Unreal. **Data and metadata vocabularies**: JSON Speckle format **Use of standard vocabularies:** JSON Speckle format **Mappings to commonly used vocabularies**: JSON Speckle format
Fair Data: **Reusability**	**License:** Data under Attribution CC 4.0. **Availability for re-use:** This data can be re-used. **Usable by third parties after end of project**: Yes **Re-use timeframe:** Permanent **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is preferable stored on Speckle.xyz server. Speckle's database is restricted only to registered users. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7). PrismArch data will be accessible as public after the end of the project. Speckle database provides its services under rules that comply with the European Law regarding privacy issues and terms of use.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** No **Is informed consent for data sharing and long-term preservation given:** Yes
Other issues	N/A

4.1.12 Development Dissemination Files

NAME	**PrismArch_WP7_012_RTD_DEVELOPMENT-DISSEMINATION-FILES**
Data summary	**Responsible partner:** CERTH **Purpose:** This dataset will be related to PrismArch dissemination and communication activities. Screenshots and other demo files during PrismArch's platform development will be shared to professional networks in order to increase further awareness, promoting dissemination activities. **Type/format:** Photo, video and audio files **Re-use of existing data:** No **Data origin:** Files created by the development of PrismArch from partners **Expected size:** A few MB in total **Data utility:** This data will be used in the context of WP7 for the dissemination and communication activities. No sensitive information will be included.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable:** The data will not be openly accessible. After aggregation and processing, they will be included in the project's dissemination activities, with any sensible information being removed. The aggregated knowledge from this data will be made accessible to the public through D7.4 [14] and D7.5 [12]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A

Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. Part of them will only be used in the project's dissemination activities, with any sensible information being removed. The aggregated knowledge from this data will be made accessible to the public through D7.4 [14] and D7.5 [15]. **How it will be accessible:** Data will be included in all dissemination activities. **Methods/software tools to access data:** N/A **Repository:** The raw files are stored in Google Drive. D7.4 [14] and D7.5 [15] will be available on the project website. **Restrictions on access:** N/A
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies:** N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies:** N/A
Fair Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project:** N/A **Re-use timeframe:** This data is not expected to be re-used. It will be only used for the needs of the dissemination activities. **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A

Security	**Security measures:** The datasets will be stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A The aggregated knowledge from this data will be made accessible to the public through D7.4 [14] and D7.5 [15]. Third parties may use it after giving appropriate credit. **Is informed consent for data sharing and long term preservation given:** N/A
Other issues	N/A

4.2 Datasets for piloting activities

4.2.1 User Requirements' Questionnaires

NAME	**PrismArch_WP6_001_PILOT_UserRequirmentsQuestionnaries**
Data summary	**Responsible partner:** ZH, AKT and SWECO **Purpose:** Three structured questionnaires that apply to each design (Architectural, Structural, and MEP) have been developed by ZH for the collection of user requirements as a part of the deliverable D6.1 [16]. It will help us define a list of requirements, including both general requirements concerning the PrismArch technical functionalities and features but also requirements focusing on the specific use cases (e.g. VR experiences, AR experiences). The findings of the questionnaires will be summarized in D6.1 [16]. **Type/format**: Excel documents containing questions and user responses **Re-use of existing data:** The data is original **Data origin:** Questionnaires filled by ZH employees and partners' employees (shared through professional network platforms)

	Expected size: A few MB in total **Data utility:** This data will be used to structure the user requirements in WP6. Their insights will be then shared to technical partners for developing PrismArch's platform. Therefore, these questionnaires address all technical partners.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The questionnaires are stored on a project's google drive folder, in which only the partners have access. Raw data is not discoverable for third parties outside the project. The aggregated knowledge from this data will be made accessible to the public through D6.1 [16]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, analysis results based on this data was shared with the consortium (only) and the findings of the questionnaires will be of the D6.1, but no sensitive information was gathered since the questionnaires were anonymous and only demographic non-identifying questions were included (age and gender). **How it will be accessible:** Stored in google drive and it is **only** internally accessible by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Web-browser (only by project partners) **Repository:** The raw questionnaires are stored in Google Drive. D6.1 [16] is available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be only used from partners to specify the user requirements

	for PrismArch's platform. **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5. Moreover, given the small sample of the questionnaire a manual control was performed by ZH to ensure data quality.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data are made publicly available through D6.1 [16]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires were informed on it by adding the data usage information at the start of the questionnaires)
Other issues	N/A

4.2.2 Evaluation Data

NAME	**PrismArch_WP6_002_PILOT_EVALUATION-DATA**
Data summary	**Responsible partner:** SWECO, AKT and ZH **Purpose:** Structured questionnaires will be developed by WP6 partners for the evaluation of the developed tools in the context of the various use cases during the pilot trials. The questionnaires will include questions that cover issues such as usefulness, usability, visualisation and interaction, learnability, encountered problems and future expectations, etc. as well as user demographics. This dataset includes questionnaires filled by the end users to assess

	the tools developed. Data collected through the questionnaires is used exclusively for analysis and statistical purposes. **Type/format**: Word/Excel documents containing questions and user responses **Re-use of existing data:** No **Data origin:** Questionnaires filled by end-users in the context of pilot evaluation **Expected size:** A few KBs per questionnaire. A few MB in total **Data utility:** This data will be used in the context of WP6 to evaluate PrismArch technologies. The evaluation results of the three pilot cases will be used by the technical partners to improve PrismArch's tools as part of further development and commercial exploitation activities. The data will be also used in D6.3 "Report on testing and evaluating the PrismArch platform" [19] in M16, as well as their updates in D6.4 [17] and D6.5 [18].
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The questionnaires and the answers will be stored on a project's google drive folder, in which only the partners have access. Access to questionnaires will be given to participants through Google Forms. The answer data is not discoverable for third parties outside the project. Access to the statistical analysis of the results will be given to the public through D6.3 [19], which will be a public deliverable. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** TBD
Fair Data: **Accessibility**	**Data openly accessible:** Raw data is considered internal working material. Hence, interviews, questionnaires, assessments, etc. will be considered confidential and will only be accessible by the consortium. After aggregation and processing, analysis results based on this data will be shared with the consortium and included in relevant deliverables. In case of a report or paper submitted for publication, all research findings will be integrated into the report or paper. Datasets will not be added to the publication. **How it will be accessible:** N/A **Methods/software tools to access data**: N/A **Repository:** N/A. **Restrictions on access:** N/A

Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The datasets will be stored in the premises of the respective pilot leaders and (if needed) on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** These datasets may contain personal information of end-users, but they will not be shared. **Is informed consent for data sharing and long term preservation given:** Informed Consent Forms (electronic or written) will be shared to all participants (architects and engineers) for the participation in the pilots. These forms will describe all the rights of the participants and will make clear that the treatment of the data is confidential, complies with GDPR and is carried out exclusively for analysis and statistical purposes. If there is a need for data sharing and long term preservation, the participants will be informed accordingly in the Informed Consent Forms.
Other issues	N/A

4.2.3 Evaluation Log Files

NAME	**PrismArch_WP6_003_PILOT_EVALUATION-LOG-FILES**
Data summary	**Responsible partner:** CERTH, SWECO, ZHVR **Purpose:** This dataset will include log files that will be

	automatically created during pilot evaluation sessions. These files will include, but not limited to, context information, timestamps and general event information along with associated analytics which will be created during the test session by users of PrismArch technologies. This content will be used for evaluation purposes and demonstration of the results of the project. **Type/format**: Log files regarding usage of PrismArch applications **Re-use of existing data:** No **Data origin:** Log files created by the usage of PrismArch from end-users in the context of pilot evaluation **Expected size:** Depends on the number of participants **Data utility:** This data will be used in the context of WP6 to evaluate PrismArch technologies, understand how useful the tools can be to the users and demonstrate the capabilities of the project technologies
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The data will not be discoverable for third parties outside the project. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** TBD
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. If there is a need for publishing some of the analytics (for instance as a part of a deliverable) all the sensible information will either be removed or pseudoanonymized. **How it will be accessible:** N/A **Methods/software tools to access data**: N/A **Repository:** N/A. **Restrictions on access:** N/A
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project**: N/A

	Re-use timeframe: N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The datasets will be stored in the premises of the respective pilot leaders and (if needed) on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** These datasets may contain personal information of end-users, but they will automatically anonymized or at least become pseudonymous if a part of them needs to be shared. **Is informed consent for data sharing and long term preservation given:** Informed Consent Form (electronic or written) will be shared to all participants (designers, influencers and customers) for the participation in the pilots. These forms will describe all the rights of the participants and will make clear that the treatment of the data is confidential, complies with GDPR and is carried out exclusively for analysis and statistical purposes. If there is a need for data sharing and long-term preservation, the participants will be informed accordingly in the Informed Consent Forms.
Other issues	N/A

4.2.4 Data Sources Questionnaire

NAME	**PrismArch_WP6_004_PILOT_DATA-SOURCES-QUESTIONNAIRE**
Data summary	**Responsible partner:** ETH **Purpose:** Small groups of selected experts will be engaged in experimental settings for evaluating the user experience and spatial cognition aspects while using the PrismArch platform. Besides behavioural Data (cf. 4.1.5 Behaviour Measurement Data) we collect questionnaire-type of data based on participants' feedback, their navigational and general spatial abilities, demographic data etc. These questionnaires are typically part of the pre- and post-test in a psychological experiment.

	Type/format: Excel documents containing questions and user responses, using mostly numeric codes for anonymization and data-analysis reasons. **Re-use of existing data:** The data is original **Data origin:** Questionnaires filled by the participants of surveys and psychological experiments. **Expected size:** A few MB in total **Data utility:** This data will be used to report the cognitive issues and UX and usability guidelines in VR-aided design environments in WP3. Their insights will then be shared to technical partners for developing PrismArch's platform. Therefore, these questionnaires address all technical partners.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: The questionnaires and the consent forms are stored by the PI whose team is collecting the data (ETH) in which only the partners have access. Raw data will not be discoverable for third parties outside the project. The aggregated knowledge from this data will be made accessible to the public through D3.2 [12] and D3.3 [13]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. No sensitive information was gathered since the questionnaires were anonymous and only demographic non-identifying questions were included (age and gender). **How it will be accessible:** Stored by ETH's PI whose team is collecting the data and it is **only internally accessible** by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Web-browser (only by project partners) **Repository:** The raw files are stored in ETH's PI. D3.2 [12] and D3.3 [13] will be available on the project website. **Restrictions on access:** Shared among project partners with access to project's files.
FAIR Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A

	Use of standard vocabularies: N/A **Mappings to commonly used vocabularies**: N/A
FAIR Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be only used by partners to define PrismArch's technical features **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5. Moreover, given the small sample of the questionnaire a manual control was performed by ETH to ensure data quality.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data will be made publicly available through D3.2 [12] and D3.3 [13]. **Is informed consent for data sharing and long term preservation given:** N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires were informed on it by adding the data usage information at the start of the questionnaires)
Other issues	N/A

4.2.5 Dissemination Files

NAME	**PrismArch_WP7_005_PILOT_DISSEMINATION-FILES**
Data summary	**Responsible partner:** CERTH **Purpose:** This dataset will be related to PrismArch dissemination and communication activities, to allow better organization of events and offer better services to attendees. Video content and photos from participants will also be used for creating dissemination content. **Type/format:** Photo, video and audio files **Re-use of existing data:** No **Data origin:** Files created by the usage of PrismArch from end-users in the context of pilot evaluation **Expected size:** Depends on the number of participants **Data utility:** This data will be used in the context of WP7 for the dissemination and communication activities
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable:** The data will be public. No sensitive information will be gathered. The aggregated knowledge from this data will also be made accessible to the public through D7.5 [15]. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will be openly accessible. If there is a need for publishing some of the data (for instance as a part of a deliverable) all the sensible information will be removed. The aggregated knowledge from this data will be made accessible to the public through D7.5 [15].

	How it will be accessible: N/A **Methods/software tools to access data:** N/A **Repository:** The raw files are stored in Google Drive. D7.5 [15] will be available on the project website. **Restrictions on access:** N/A
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies:** N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies:** N/A
Fair Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project:** N/A **Re-use timeframe:** N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The datasets will be stored in the premises of the respective pilot leaders and (if needed) on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).

Ethical aspects	**Possible ethical and legal aspects preventing sharing:** These datasets may contain personal information of end-users, but they will automatically anonymized or at least become pseudonymous if a part of them needs to be shared. The aggregated knowledge from this data will be made accessible to the public through D7.5 [15]. Third parties may use it after giving appropriate credit. **Is informed consent for data sharing and long term preservation given:** Informed Consent Form (electronic or written) will be shared to all participants (designers, architects and engineers) for the participation in the pilots. These forms will describe all the rights of the participants and will make clear that the treatment of the data is confidential, complies with GDPR and is carried out exclusively for analysis and statistical purposes. If there is a need for data sharing and long-term preservation, the participants will be informed accordingly in the Informed Consent Forms.
Other issues	N/A

5. Conclusions

This DMP identifies the datasets managed by the PrismArch consortium, organized by the several struts of work in the project. In the present document, the first deliverable related to data management plan, we discussed the different datasets, both pre-existing ones and newly-created within the duration of the project. A template was shared to all partners in order to fill in the information related to the datasets aligned with the "Template for Horizon 2020 Data Management Plan (DMP)".

We clustered the datasets in the following two categories depending on their purposes:

- Supporting research and technical development (RTD),
- Resulting from pilot activities (PILOT),

The consortium guarantees that will take all the suitable measurements to make the data FAIR. In any case, PrismArch datasets will either be openly shared (by uploading them in open repositories) or shared internally among specific partners (stored on the project Google Drive). Datasets to be openly shared, will be deposited in certified repositories like Zenodo that have in place strong mechanisms and protocols for data security and long-term data preservation. Similar mechanisms exist in both the Google Drive repository and the partners' servers to ensure data protection.

6. References

[1] D8.1 - Project management and quality assurance plan, January 31st, 2021

[2] D9.1 - H - Requirement No. 1, PrismArch Consortium, April 30th, 2021

[3] D9.2 - POPD - Requirement No. 2, April 30th, 2021

[4] D1.2 -Elaborated report of cross-discipline principles-rules-constraints, and interfaces definition for cross-disciplinary and multi-simulation perspectives in VR, August 31st, 2021

[5] D1.3 – Final version of the report on principles-rules-constraints and interfaces definition for cross-disciplinary and multi-simulation perspectives in VR, April 30th, 2022

[6] Chappell D. & Dunn M., (2016), The Architect in Practice, Wiley Blackwell; 11th edition

[7] The American Institute of Architects, (2016), Architectural Graphic Standards, Wiley; 12th edition

[8] Collective Work, (2013). The Architect's Handbook of Professional Practice, Wiley; 15th edition

[9] D2.2 - Integration-ready version of AI algorithms to traverse the parametric solution space, October 31st , 2022

[10] D2.3 - Final revised version of parametric space of design, algorithms for AI assisted editing/design in VR, and algorithms for designer modelling, June 30th , 2022

[11] D3.1 - Report of cognitive issues in VR-aided design environments, May 31st, 2021

[12] D3.2 - Initial report on UX and usability guidelines in VR-aided design environments, October 31st , 2022

[13] D3.3 - Final report on UX and usability guidelines in VR-aided design environments, June 30th , 2022

[14] D7.4 - Engagement strategies: plan and material, October 31st , 2022

[15] D7.5 - Dissemination and communication plan (Update), April 30th, 2022

[16] D6.1 – Usage scenarios specification, May 31st, 2021

[17] D6.4 - Report on evaluation plan and usability study (Update), August 31st, 2022

[18] D6.5 - Report on testing and evaluating the PrismArch platform (Update), October 31st , 2022

[19] D6.3 - Report on testing and evaluating the PrismArch platform, February 30th, 2022

[20] Wu, Wenming & Xiao-Ming, Fu & Tang, Rui & Wang, Yuhan & Qi, Yu-Hao & Liu, Ligang. (2019). Data-driven interior plan generation for residential buildings. ACM Transactions on Graphics. 38(6).

[21] D4.1 Two-way communication protocol for interconnecting PrismArch with BIM/CAE-Simulation software, May 31st, 2021

7. ANNEX

- **Dataset Information Form – to be filled from partners**

NAME	**PrismArch_<WPno>_<Serial number of dataset>_<data type>_<dataset title>**
Data summary	**Responsible partner:** Partner responsible for producing and/or using the specific dataset **Purpose:** Short description of data (include a sample of the features of the dataset if possible). Also, what is the purpose of data collection/generation (and its relation to project objectives) in the context of PrismArch? **Type/format**: What is the type/format of the data generated/collected? **Re-use of existing data:** Are existing datasets reused and how? **Data origin:** What is the origin/source of the data? **Expected size:** What is the expected data/dataset size (if known)? **Data utility:** To whom will this data be useful and how? (inside the project, for instance WP and/or partners, and also to third parties, if applicable)
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: Are the data produced in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? **Search keywords:** Will search keywords be provided that optimize possibilities for re-use? **Versioning:** Will clear version numbers be provided? **Metadata creation:** Specify standards for metadata creation (if any). If there are no standards in your discipline, describe what type of metadata will be created and how.
FAIR Data: **Accessibility**	**Data openly accessible:** Will data produced in the project be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. **How it will be accessible:** How will the data be made accessible

	(e.g. by deposition in an open repository)? **Methods/software tools to access data**: What methods or software tools are needed to access the data? Also, is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? **Repository:** Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible. **Restrictions on access:** If there are restrictions on use, how will access be provided?
FAIR Data: **Interoperability**	**Interoperability:** Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organizations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)? **Data and metadata vocabularies**: Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability **Use of standard vocabularies:** Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? **Mappings to commonly used vocabularies**: In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
FAIR Data: **Reusability**	**License:** Specify how the data will be licensed to permit the widest reuse possible. E.g. Open Data License (Creative Commons CC0 License, Creative Common Attribution License-CC-BY v4.0, etc.). **Availability for re-use:** When will data be made available for re-use. If applicable, specify why and for what period a data embargo is needed **Usable by third parties after end of project**: Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

	Re-use timeframe: Specify the length of time for which the data will remain re-usable **Data quality assurance process:** Describe data quality assurance processes
Allocation of resources	**Costs for making the data FAIR:** Covered by the individual partners that will share the data **Costs for long-term preservation:** TBD
Security	**Security measures:** Security measures implemented for data protection (incl. controlled access, user authentication, firewalls, VPNs, encryption, back-ups, etc.)
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Are there any ethical or legal issues that can have an impact on data sharing? **Is informed consent for data sharing and long term preservation given:** Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?
Other issues	Refer to other national/funder/sectorial/departmental procedures for data management that you may be using (if any)