

COGITO

CONSTRUCTION PHASE  
DIGITAL TWIN MODEL

[cogito-project.eu](http://cogito-project.eu)

## D1.2 Data Management Plan



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958310.

## D1.2 – Data Management Plan

Dissemination Level:	Public
Deliverable Type:	ORDP: Open Research Data Pilot
Lead Partner:	Hypertech
Contributing Partners:	Hypertech, UPM, ASM
Due date:	30-04-2021
Actual submission date:	30-04-2021

### Authors

Name	Beneficiary	Email
Katerina Valalaki	Hypertech	k.valalaki@hypertech.gr
Giorgos Giannakis	Hypertech	g.giannakis@hypertech.gr
Raúl García Castro	UPM	rgarcia@fi.upm.es
Agnieszka Kowalska	ASM	a.kowalska@asm-poland.com.pl

### Reviewers

Name	Beneficiary	Email
Martín Bueno	UEDIN	<a href="mailto:martin.bueno@ed.ac.uk">martin.bueno@ed.ac.uk</a>
Agnieszka Kowalska	ASM	a.mikolajczyk@asm-poland.com.pl

### Version History

Version	Editors	Date	Comment
0.1	Katerina Valalaki, Giorgos Giannakis	10.03.2021	ToC and Data Elements Table
0.2	Raúl García Castro	15.03.2021	Interoperability section filled in
0.3	Katerina Valalaki, Giorgos Giannakis	30.03.2021	Initial draft
0.4	Agnieszka Kowalska	08.04.2021	IPR management
0.5	Katerina Valalaki	15.04.2021	Consolidation of partners' input
0.6	Giorgos Giannakis	18.04.2021	Internal Review & update
0.7	Katerina Valalaki, Giorgos Giannakis	23.04.2021	Full draft
0.8	Katerina Valalaki, Giorgos Giannakis	26.04.2021	Updated draft sent to reviewers
0.9	Katerina Valalaki, Giorgos Giannakis	29.04.2021	Updated version addressing comments
1.0	Katerina Valalaki, Giorgos Giannakis	30.04.2021	Final version submitted to EC

### Disclaimer

©COGITO Consortium Partners. All right reserved. COGITO is a HORIZON2020 Project supported by the European Commission under Grant Agreement No. 958310. The document is proprietary of the COGITO consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights. The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use, which may be made, of the information contained therein.

## Executive Summary

In order to support the potential for future sharing and reuse of the data generated during COGITO implementation, a Data Management Plan (DMP) has been created and the first version is presented in this document. The DMP has been designed following the EC recommendations and best practices and will be regularly updated outlining how the research data collected or generated will be handled during and after the project.

In this first version, the COGITO methodology for data management is presented along with a DMP template that will be used to detail all the COGITO data types. An initial attempt to list the data types that will be generated / collected during the project lifecycle has been made by the Consortium partners and the results are summarised in this report. Information about the purpose of data collection is provided while the measures that COGITO has in place towards handling Findable, Accessible, Interoperable and Re-usable (FAIR) data are described.

A DMP is required for any project participating in the extended Open Research Data Pilot (ORDP). COGITO would contribute towards this direction. Thus, this is a preliminary version of the DMP provided early in the project (M6). Subsequently, the DMP needs to be updated throughout the project whenever significant changes arise, such as (non-exhaustive list):

- new data are generated;
- changes are made to the consortium policies (e.g. new innovation potential, decision to file for a patent);
- changes are made to the consortium composition (e.g. new consortium members joining or old members leaving).

Thus, the COGITO DMP has been designed to be a living document that should be regularly updated - at a minimum once prior to each periodic evaluation/assessment of the project.

## Table of contents

Executive Summary .....	2
Table of contents .....	3
List of Figures .....	5
List of Tables.....	6
List of Acronyms.....	7
1 Introduction .....	8
1.1 Scope and Objectives of the Deliverable.....	8
1.2 Relation to other Tasks and Deliverables .....	8
1.3 Structure of the Deliverable .....	9
2 Methodology.....	10
2.1 Guidelines and DMP template.....	10
2.2 COGITO Methodology.....	11
3 Data Summary.....	13
3.1 Purpose of collected data .....	14
3.2 Research datasets.....	15
3.2.1 Public Software Components .....	15
3.2.2 Public Report Deliverables .....	17
3.2.3 COGITO data types .....	19
3.2.4 Scientific Publications.....	23
4 COGITO FAIR Data .....	24
4.1 Making data findable .....	24
4.2 Making data openly accessible .....	25
4.2.1 Open Access to scientific publications .....	25
4.2.2 Open access to research data .....	26
4.3 Making data interoperable.....	27
4.4 Increase data reuse .....	28
4.4.1 IPR Management.....	28
4.4.2 Copyright Licenses.....	29
5 Allocation of resources .....	30
5.1 Human Resources.....	30
5.2 Financial Resources .....	31
6 Data Security .....	32
6.1 GDPR and DPIA .....	32
6.2 Measures to ensure data protection .....	32
6.3 Data collection, storage and destruction .....	33

7	Ethical Aspects .....	35
8	Conclusions .....	37
9	References.....	38

## List of Figures

Figure 1: DMP relation to other WPs of the COGITO work plan .....	9
Figure 2: COGITO Data Management methodology and versions of the Data Management Plan .....	11
Figure 3: High level view of COGITO conceptual architecture .....	14
Figure 4: COGITO purpose for collecting data .....	15
Figure 5: Data classification levels in terms of accessibility .....	27
Figure 6: Technical and organisational measures to ensure data protection. More detailed description can be found in COGITO D10.2 “POPD Requirements” .....	33
Figure 7: COGITO rules for the protection of personal data .....	35

List of Tables

Table 1: COGITO public demonstrators ..... 15

Table 2: COGITO public report deliverables ..... 17

Table 3: COGITO Datatypes (M6) ..... 19



## List of Acronyms

Term	Description
<b>API</b>	Application Programming Interface
<b>BIM</b>	Building Information Model
<b>CC0</b>	Creative Commons No Rights Reserved licence
<b>CC-BY</b>	Creative Commons Attribution International licence
<b>COGITO</b>	Construction Phase diGital Twin mOdel
<b>D</b>	Deliverable
<b>DT</b>	Digital Twin
<b>DMP</b>	Data Management Plan
<b>DoA</b>	Description of Action
<b>DOI</b>	Digital object identifier
<b>DPIA</b>	Data Protection Impact Assessment
<b>EC</b>	European Commission
<b>ENISA</b>	European Union Agency for Cybersecurity
<b>EU</b>	European Union
<b>FAIR</b>	Findable, Accessible, Interoperable and Re-usable
<b>GDPR</b>	General Data Protection Regulation
<b>H2020</b>	Horizon 2020
<b>HTTPS</b>	Hypertext Transfer Protocol Secure
<b>LL</b>	Living Lab
<b>OAI-PMH</b>	Open Archives Initiative Protocol for Metadata Harvesting
<b>PDF</b>	Portable Document Format
<b>POPD</b>	Protection of Personal Data
<b>RDF</b>	Resource Description Framework
<b>SLA</b>	Service Level Agreement
<b>SSL</b>	Secure Sockets Layer
<b>IPR</b>	Intellectual Property Right
<b>PID</b>	Persistent Identifier
<b>T</b>	Task
<b>TLS</b>	Transport Layer Security
<b>WP</b>	Work Package

# 1 Introduction

A Data Management Plan (DMP) is a living document describing the datasets that are generated and/ or reused during and after a research lifetime. A DMP aims to provide researchers with essential information to re-produce, re-distribute and re-purpose research results thus assuring for their validity and exploitation.

This document is part of the ‘Pilot on Open Research Data in Horizon 2020’, describing the data handling within COGITO, namely the procedures for data collection, storage, protection, retention and destruction and their compliance with national and EU legislation during the project course and after its end.

Various types of data will be collected in the project, such as research data i.e. test users’ responses to questionnaire surveys, opinions expressed in living labs, discussions and round tables, logged data coming from the construction pilot sites and relevant actors that will be connected to the COGITO platform and other user interfaces that will exchange data through it. Additionally, some services will generate meta-data. The data types and formats to be collected and generated per COGITO Work Package (WP) and the procedures for data storage and back up are described in the present deliverable.

The data access rules are described and data documentation is outlined, so that data can be reused and shared after the project end. The security mechanisms to be put in place for data protection are presented. Finally, data protection and ethical issues are considered.

The DMP will be implemented across the WPs, based on data delivered. The deliverable is a living document that is updated over the course of the project, depending on project developments or legislative or policy changes by the applicable laws.

## 1.1 Scope and Objectives of the Deliverable

The DMP documents the lifecycle of data. Thus, the scope of this document is to provide details on the collection, storage, access, sharing, and reproducibility of COGITO data. It describes data that will be acquired or produced during COGITO research; how the data will be managed, described, and stored, what standards will be used, and how data will be handled and protected during and after the completion of the project.

The aim of this document is to clearly address issues, such as: the overall methodology for handling the scientific outcomes of the project; the specification of data types that the project generates and/or collects; the standards that will be used; the process for how this data will be exploited and/or shared/made accessible for verification and reuse; and the data preservation and maintenance processes.

The COGITO DMP is formulated in accordance with the H2020 guidelines regarding Open Research Data, as COGITO aims to actively contribute to the Open Research Data Pilot. In alignment with the EC Guidelines for Open Access [1], the document defines how the various datasets generated/collected will be further disseminated to a wider audience.

## 1.2 Relation to other Tasks and Deliverables

This DMP is the outcome of T1.3 “Data Management”, which belongs to WP1 “Quality Assurance and Management”. It includes horizontal activities impacting and receiving feedback from all the other WPs (and their underlying tasks) of the COGITO work plan (as shown in Figure 1). It is in continuous feedback loop with WP9 “Dissemination, exploitation and standardisation activities”. It provides the guidelines and principles for data management to all the other WPs while it also receives input from them to fill in the DMP template as presented in Section 2.1. The DMP is complemented by D10.1 – H requirements No.1 and D10.2 – POPD

requirements No.2, where the ethics requirements regarding humans and the personal data protection are analysed.

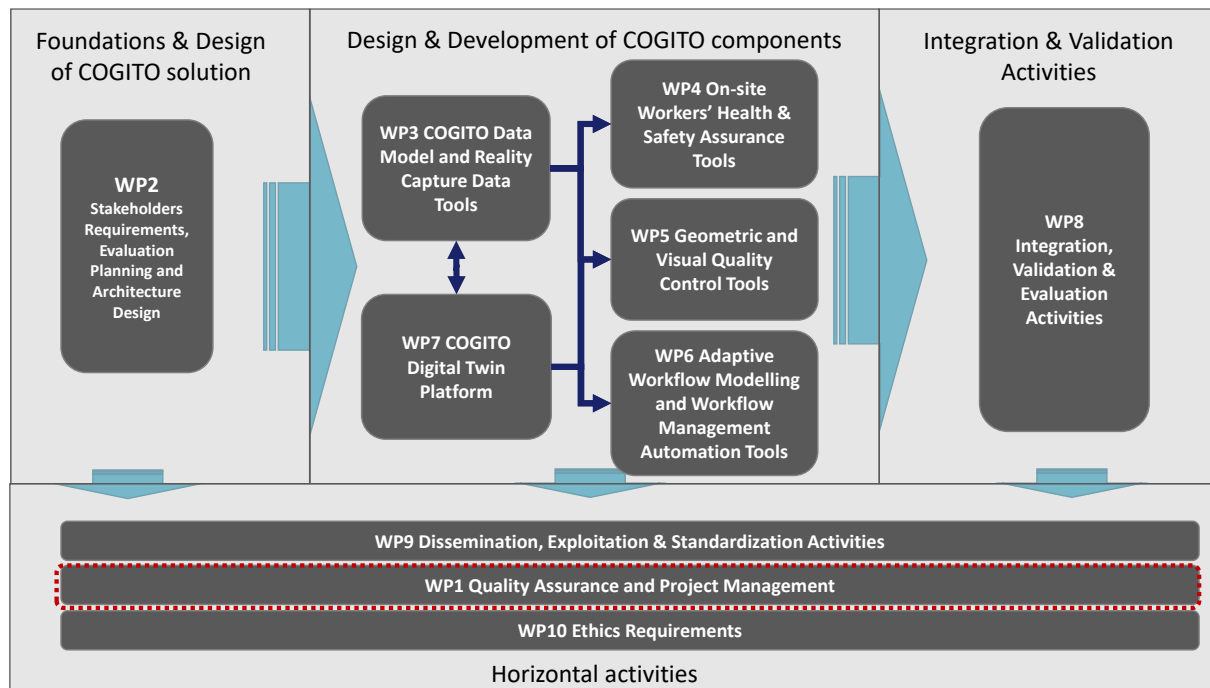


Figure 1: DMP relation to other WPs of the COGITO work plan

### 1.3 Structure of the Deliverable

The deliverable is structured as follows:

- After a brief introduction in Section 1, the methodology to be used throughout the project lifecycle for data management is presented in Section 2;
- Following the DMP template of the EC, Section 3 presents the COGITO data summary describing the purpose of collecting data within the project. Furthermore, the various COGITO research datasets are presented, including: software components, report deliverables, data types and scientific publications
- Section 4 describes how COGITO will ensure FAIR data handling making all types of data managed within the project: findable, accessible, interoperable and re-usable;
- Section 5 defines the allocation of resources for the data management and explains the allocated responsibilities within the project;
- Section 6 discusses data security issues regarding the aspects referring to the storage and transfer of data and the protection of personal and/or sensitive data;
- Section 7 focuses on the COGITO ethics, in line with the current GDPR legislation in Europe and the COGITO D10.1 "D10.1 H- Requirement No. 1 and D10.2 POPD - Requirement No. 2;
- Finally, Section 8 concludes the document.

## 2 Methodology

The COGITO DMP describes the data management life cycle for the data to be collected, processed and/or generated by the project. As part of making research data FAIR, COGITO DMP should include information about the handling of research data during and after the end of the project by answering the following questions:

- what kind of data will the project collect, process and/or generate, and to whom it might be useful later on?
- which methodology and standards will be applied?
- which metadata are required to find and understand the collected, processed and/or generated data, ideally according to the particular scientific discipline's standards?
- whether data will be shared/made open access? and
- how data will be preserved (including the period after the COGITO project has ended)? How to archive and preserve the open datasets of the project?<sup>1</sup>

### 2.1 Guidelines and DMP template

For Horizon 2020 projects, a FAIR DMP [1] template has been designed to be applicable to any project that produces, collects or processes research data. The respective activities defined as part of the methodology are structured around defining and documenting all the required information towards filling a common template. Such template contains a set of questions that should be answered with a level of detail appropriate to the project. The template includes the following sections:

- Data Summary
  - This section is associated with the COGITO's activities. In the first version of the COGITO DMP, an initial break down of datasets to WPs and relevant activities has been performed. Furthermore, the partners that will be involved in the identified activities have been specified. More details can be found in Section 3.
- FAIR Data
  - In general, COGITO data should be 'FAIR'. This means that the COGITO solution, as well as relevant activities foreseen in the DoA, will be performed in a way towards:
    - Making data findable, including provisions for metadata;
    - Making data openly accessible;
    - Making data interoperable;
    - Increase data reuse (through clarifying appropriate licenses).

More details can be found in Section 4.

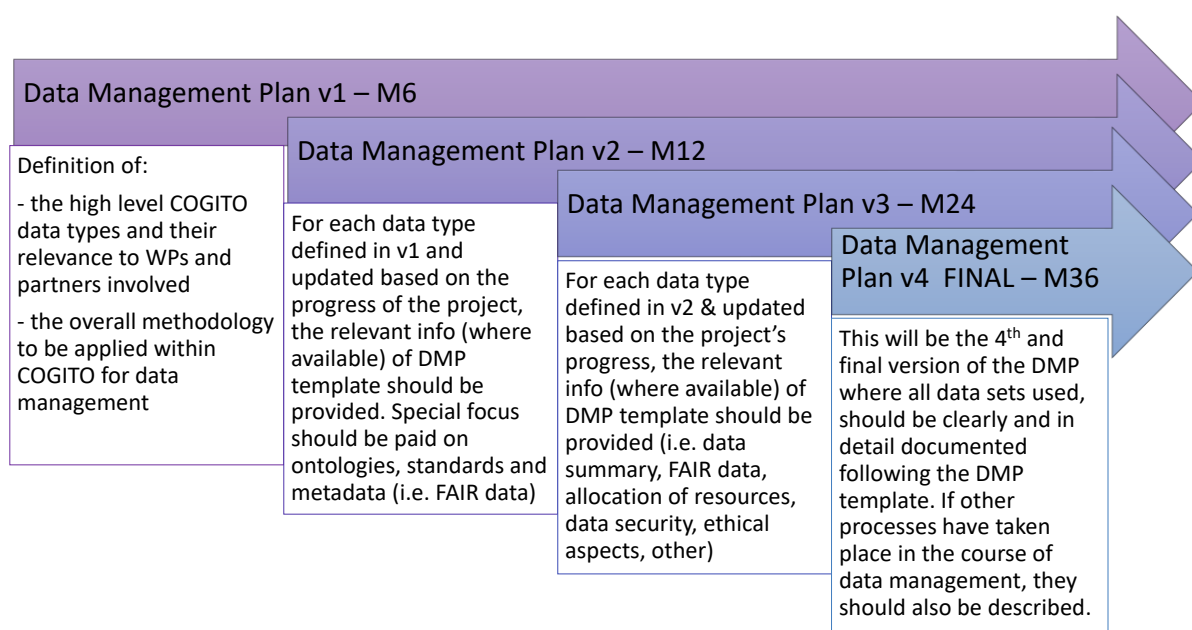
- Allocation of resources
  - In general data management can be a costly process (in terms of human and financial resources). However, specific provisions have been made in COGITO since the very beginning of the project (proposal phase) and all the relevant costs and allocation of resources have been clearly defined in the COGITO DoA and are incorporated in the project plan.
- Data Security
  - This section aims to report the security measures that will be considered for the handling of COGITO data.
- Ethical Aspects
  - This section should provide information about any potential ethical or legal issues that can have an impact on data sharing. Furthermore, the methods used for processing of personal data (if any) should be described.

<sup>1</sup> Funding bodies ensure that publicly funded research outputs can have a positive impact on future research, on policy development and societal change

- Other Issues
  - This section may refer to other national/funder/sectorial/departmental data management procedures (if any)

## 2.2 COGITO Methodology

A specific methodology will be followed within COGITO to ensure appropriate data management. Although, only one version is foreseen in the COGITO DoA, the DMP will be treated as a living document and being updated whenever deemed necessary and, at least once prior to the end of each reporting period. Thus, at least four versions of the DMP will be created during the course of the project. The process to be followed along with the information that should be provided in each version is shown in Figure 2. The methodology has been based on the OpenAIRE guides for researchers [2] adjusted to COGITO scope and aligned with the project Gantt Chart.



**Figure 2: COGITO Data Management methodology and versions of the Data Management Plan**

More specifically:

- The 1st version of the DMP sets the approach basis to be followed. It describes the basic principles the COGITO project should comply, and provides the DMP template that has been based on the guidelines and best practices provided by the EC. In addition, an initial definition of the WPs, involved partners and datasets that will be used in COGITO has been performed;
- The 2nd version should elaborate on the data formats, standards and metadata that will be used for each data type. The work should be based on the T3.2 “COGITO Data Model, Ontology Definition and Interoperability Design” and its first deliverable (to be submitted in M11). Hence, the document should include a first version of FAIR data for the defined data types which should be delivered by M12;
- The 3rd version coincides with the delivery of all the components of the COGITO architecture. To this end, this is supposed to be a relatively mature version of the DMP. All the aspects of the DMP template per data type should be defined. Input by all WPs dealing with the development of the various COGITO’s components will be required. The document should be delivered by M24;
- The 4th and final version will be delivered with the project end (M36). The final fully completed templates of all the data types should be delivered. Updates and input from the demonstration and validation activities will be required. The aim of the final COGITO DMP is to provide a complete documentation about all COGITO data elements.

From the above description, it is obvious that the COGITO DMP will be structured following an incremental approach. In case of no related available information for a given dataset, then the phrase “Non-applicable” or N/A will be used. This may be updated in the next versions of the DMP as above detailed.

In the following paragraphs, the template’s main sections and proposed contents are listed and presented, along with the way COGITO addresses each of the sections.

### 3 Data Summary

With the aim of minimising construction project durations/costs and reduce workplace accidents, COGITO targets a semantic and pragmatic alignment between various data capture techniques. A bundle of services will be delivered to facilitate the timely detection of health & safety hazards to humans and construction quality defects, as well as to provide the means for a constantly up-to-date workflow management. Methods & technologies to ensure interoperability among the different components comprising the COGITO architecture are applied within the Digital Twin platform, the core of the entire toolchain.

To achieve its goals, the project aims be tested in real-life environments: a pre-validation site in Austria (during a project's pre validation phase) and; two (2) pilot sites in Spain and Denmark . Data to be collected, processed, analysed, and generated, include (non-exhaustive list):

- **As-planned data** including 3D BIM models, scheduling, costs, business processes, workflows and defined Service Level Agreements (SLAs) for smart contracts establishment;
- **As-built data** including data obtained from sensors in the field, which can include IoT devices for location tracking, laser scanners and imagery;
- **Historical data** including data about productivity, scheduling, evaluation of processes and planning activities, etc.;
- **Quasi-static information** that captures the evolution of the “static” aspects on the site .

Furthermore, within COGITO, a wide range of the construction industry's stakeholders (e.g., Architects, Building Information Modelling (BIM) Managers, Surveyors, Contractors, and Workers, etc.) will be treated as research participants. Their engagement in all research, innovation, standardisation, demonstration, and communication activities is of paramount importance towards the successful project implementation and the realisation of its anticipated impacts. This collaborative and engaging framework mandates the collection of further data to gather opinions, experience and knowledge that will facilitate the COGITO development and ensure the delivery of an innovative solution that covers actual stakeholders' needs. Their feedback from surveys, their participation in workshops and their involvement in the project activities along with the data that will be collected and analysed through these activities will be handled in the T9.2 “COGITO Living Lab Activities”.

From the above description, it is evident that a large amount of diverse types of data will be handled within COGITO. To this end, the respective datasets have been defined and presented in Section 3.2.

A high-level view of the COGITO conceptual architecture is shown in Figure 3. It includes the identified components of the under development integrated solution. More details about each component's functionality and its contribution to the realisation of the COGITO use cases can be found in the D2.1 “Stakeholder requirements for the COGITO system”. The reason for providing this high-level view is to facilitate the data types definition and ensure that all information flows will be analysed adequately in the COGITO data management framework. The conceptual architecture illustrated below along with the list of components defined in the course of WP2 “Stakeholders Requirements, Evaluation Planning and Architecture Design” was the starting point for collecting and defining the COGITO data types (detailed in Section 3.2.3).

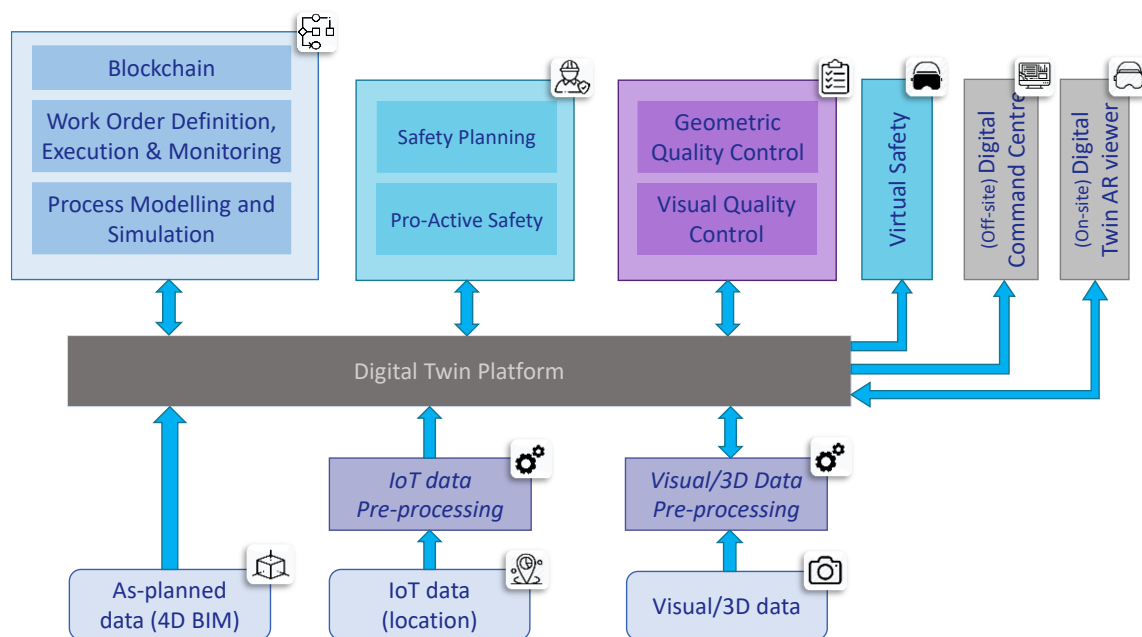


Figure 3: High level view of COGITO conceptual architecture

### 3.1 Purpose of collected data

As mentioned above, COGITO will develop an innovative tool-suite targeting the digitalisation of construction industry and provision of services to involved stakeholders. For developing the solution, user requirements are gathered at the project's early stage involving construction industry stakeholders through workshops, surveys, etc. The objective is to identify the end-users' needs and perform a process for **requirements elicitation** that will capture current status and potential improvements that can optimise the different phases of a construction project (taking into account safety, cost and time parameters). While requirements elicitation is performed at the project's early stages, a living lab is also established to create an ecosystem for **sharing knowledge, ideas and lessons learned** with the construction industry community. This is a continuous feedback process where data and information will be shared for enabling COGITO to contribute to the digitalisation of the construction field.

Apart from stakeholders' input for structuring the solution, further data will be collected, mainly from the pilot sites allowing, the further **development and configuration** of the COGITO solution. This will be the actual input that would be required to deliver the solution. As mentioned in the previous sections this include as-planned, as-built, historical and quasi-static data that are the operational pre-requisites of the COGITO solution.

After the **integration** of the whole solution, the COGITO toolbox will be tested in a controlled environment – in a RSRG's laboratory facility located in Austria whose purpose is to test new construction technologies – albeit involving actual assets and people. Thus, this **testing** and pre-validation process will require collection and processing of the data mentioned before coming from the aforementioned lab.

Then, the second phase of **demonstration and validation** will involve the application and use of the COGITO toolbox during actual construction works in two different pilot sites, one in Murcia, Spain and a second one in Copenhagen, Denmark. Thus, this demonstration process will require collection and processing of the data mentioned before coming from the pilot sites.



Figure 4: COGITO purpose for collecting data

COGITO develops the proposed solution following an agile development process. This means that the different software development lifecycle phases, i.e., requirements analysis, design, implementation, testing (as described above) are repetitively executed towards final product delivery. To this end, data are supposed to be gathered and processed throughout the project duration to facilitate the development and demonstration processes.

## 3.2 Research datasets

In this section the COGITO outputs (demonstrators and reports), that have been agreed to be publicly available, and the datasets, including the inputs/outputs of the project (as these have been defined so far), are presented. This includes public demonstrators and reports defined in the DoA and identified datasets at this phase of the project.

### 3.2.1 Public Software Components

We are considering the COGITO public deliverables as part of the DMP. Table 1 presents the list of public deliverables (type: demonstrator) of the project.

Table 1: COGITO public demonstrators

WP No	Del Rel. No	Title	Component name in the architecture	Lead Beneficiary	Est. Date	Del.
WP3	D3.5	IoT Data Pre-processing Module v1	IoT Data pre-processing	Hypertech	28 Feb 2022	
WP3	D3.6	IoT Data Pre-processing Module v2	IoT Data pre-processing	Hypertech	30 Apr 2022	
WP3	D3.7	Visual Data Pre-processing Module v1	Visual Data Pre-processing	CERTH	30 Nov 2021	
WP3	D3.8	Visual Data Pre-processing Module v2	Visual Data Pre-processing	CERTH	30 Apr 2022	
WP4	D4.1	Preventive Health & Safety Application v1	Safety Planning (SafeConAI)	AU	30 Apr 2022	
WP4	D4.2	Preventive Health & Safety Application v2	Safety Planning (SafeConAI)	AU	31 Oct 2022	
WP4	D4.5	Personalised Alerts, Prediction and Feedback Tools v1	Proactive Safety	AU	31 May 2022	
WP4	D4.6	Personalised Alerts, Prediction and Feedback Tools v2	Proactive Safety	AU	31 Oct 2022	
WP4	D4.7	Interactive Visual Material for Workforce Training v1	Virtual Safety	AU	31 May 2022	
WP4	D4.8	Interactive Visual Material for Workforce Training v2	Virtual Safety	AU	31 Oct 2022	
WP5	D5.1	Innovative Scan-vs-BIM - based Geometric QC component v1	Geometric Quality Control	UEDIN	31 Jan 2022	

<b>WP5</b>	D5.2	Innovative Scan-vs-BIM - based Geometric QC component v2	Geometric Quality Control	UEDIN	31 Oct 2022
<b>WP5</b>	D5.3	Deep-Learning -based Visual QC component v1	Visual Quality Control	CERTH	31 Mar 2022
<b>WP5</b>	D5.4	Deep-Learning -based Visual QC component v2	Visual Quality Control	CERTH	31 Oct 2022
<b>WP5</b>	D5.7	User interface and AR enabled In-situ QC Visualisation v1	DigiTAR	CERTH	31 May 2022
<b>WP5</b>	D5.8	User interface and AR enabled In-situ QC Visualisation v2	DigiTAR	CERTH	31 Oct 2022
<b>WP6</b>	D6.1	Blockchain & Smart Contracts on the Workflow Modelling and Management v1	Blockchain	QUE	31 Mar 2022
<b>WP6</b>	D6.2	Blockchain & Smart Contracts on the Workflow Modelling and Management v2	Blockchain	QUE	31 Oct 2022
<b>WP6</b>	D6.3	Adaptive Processes/Workflow Modelling and Simulation-based Optimisation Module v1	Process Modelling and Simulation tool (PMS)	BOC	31 Mar 2022
<b>WP6</b>	D6.4	Adaptive Processes/Workflow Modelling and Simulation-based Optimisation Module v2	Process Modelling and Simulation tool (PMS)	BOC	31 Oct 2022
<b>WP6</b>	D6.5	Adaptive Workflow Management and Automation Tool v1	Work Order Definition and Monitoring tool (WODM)	NT	31 May 2022
<b>WP6</b>	D6.6	Adaptive Workflow Management and Automation Tool v2	Work Order Definition and Monitoring tool (WODM)	NT	31 Oct 2022
<b>WP6</b>	D6.7	Personalised On-site Works Support Apps v1	Work Order Execution Assistance tool (WOEA)	NT	31 May 2022
<b>WP6</b>	D6.8	Personalised On-site Works Support Apps v2	Work Order Execution Assistance tool (WOEA)	NT	31 Oct 2022
<b>WP6</b>	D6.9	Workflow User Interface for Project Managers v1	Work Order Definition and Monitoring tool (WODM) UI	NT	31 May 2022
<b>WP6</b>	D6.10	Workflow User Interface for Project Managers v2	Work Order Definition and Monitoring tool (WODM) UI	NT	31 Oct 2022
<b>WP7</b>	D7.3	Extraction, Transformation & Loading Tools and Model Checking v1	Digital Twin Platform	UCL	31 May 2022
<b>WP7</b>	D7.4	Extraction, Transformation & Loading Tools and Model Checking v2	Digital Twin Platform	UCL	31 Oct 2022
<b>WP7</b>	D7.5	Data Transformation for 3D BIM Rendering v1	Digital Command Centre (DCC)	Hypertech	31 May 2022

<b>WP7</b>	D7.6	Data Transformation for 3D BIM Rendering v2	Digital Command Centre (DCC)	Hypertech	31 Oct 2022
<b>WP7</b>	D7.7	Construction Digital Twin 3D Visualisation module v1	Digital Command Centre (DCC)	Hypertech	31 May 2022
<b>WP7</b>	D7.8	Construction Digital Twin 3D Visualisation module v2	Digital Command Centre (DCC)	Hypertech	31 Oct 2022
<b>WP7</b>	D7.9	Digital Twin Platform v1	Digital Twin Platform	UCL	31 May 2022
<b>WP7</b>	D7.10	Digital Twin Platform v2	Digital Twin Platform	UCL	31 Oct 2022
<b>WP8</b>	D8.1	Integrated COGITO system v1	COGITO toolbox	UCL	31 Aug 2022
<b>WP8</b>	D8.2	Integrated COGITO system v2	COGITO toolbox	UCL	30 Apr 2023
<b>WP8</b>	D8.3	Integrated COGITO system v3	COGITO toolbox	UCL	31 Aug 2023

The purpose of creating these deliverables is to produce an accompanying documentation as regards the demonstrators / software components developed in the course of the project. The documentation should include all the information required to describe the components, including relevant API documentation describing interfaces, user manual, version, licences and so on. The template is already defined and agreed with the COGITO partners. Each document has an already defined title specified in the DoA and presented above.

The “demonstrator deliverables” are stored in the cross-platform portable document format (PDF). For open-source components, the source code may also be shared using appropriate tools (yet to be defined among the COGITO Consortium). All public deliverables will be published openly in the COGITO webpage (following E.C. review and approval). Public deliverables may also be added to Zenodo at the end of the project. For uploading documents to Zenodo the following information should be included:

- Type of publication: Project Deliverable;
- Digital object identifies (DOI): Zenodo will assign one;
- Title of the deliverable;
- Authors;
- A short description or abstract;
- License;
- Funding.

### 3.2.2 Public Report Deliverables

We are considering the COGITO public deliverables as part of the DMP. Table 2 presents the list of public deliverables (type: report) of the project.

**Table 2: COGITO public report deliverables**

<b>WP No</b>	<b>Del. No</b>	<b>Title</b>	<b>Lead Beneficiary</b>	<b>Est. Del. Date (annex I)</b>
<b>WP2</b>	D2.1	Stakeholder requirements for the COGITO system	UEDIN	30 Apr 2021
<b>WP2</b>	D2.2	Analysis of digital tools market and prevailing regulatory frameworks	ASM	30 Jun 2021
<b>WP2</b>	D2.3	COGITO evaluation methodology	UEDIN	30 Jun 2021
<b>WP2</b>	D2.4	COGITO system architecture v1	Hypertech	31 Jul 2021
<b>WP2</b>	D2.5	COGITO system architecture v2	Hypertech	30 Apr 2022
<b>WP3</b>	D3.1	Survey of Existing Models & Ontologies & Associated Standardisation Efforts	UPM	30 Apr 2021

<b>WP3</b>	D3.2	COGITO Data Model & Ontology Definition and Interoperability Design v1	UPM	30 Sep 2021
<b>WP3</b>	D3.3	COGITO Data Model & Ontology Definition and Interoperability Design v2	UPM	30 Apr 2022
<b>WP3</b>	D3.4	COGITO Data Model & Ontology Definition and Interoperability Design v3	UPM	31 Oct 2022
<b>WP4</b>	D4.3	Proactive Real-time Risk Monitoring and Detection Methods v1	AU	30 Apr 2022
<b>WP4</b>	D4.4	Proactive Real-time Risk Monitoring and Detection Methods v2	AU	31 Oct 2022
<b>WP5</b>	D5.5	BIM-based Standard Test Methods for Geometric Quality Control v1	UEDIN	31 May 2022
<b>WP5</b>	D5.6	BIM-based Standard Test Methods for Geometric Quality Control v2	UEDIN	31 Oct 2022
<b>WP7</b>	D7.1	Digital Twin Platform Design & Specification v1	UCL	30 Nov 2021
<b>WP7</b>	D7.2	Digital Twin Platform Design & Specification v2	UCL	31 Oct 2022
<b>WP8</b>	D8.4	Report on outcomes of COGITO system pre-validation	RSRG	30 Apr 2023
<b>WP8</b>	D8.5	Report on COGITO system deployment activities	CERTH	31 Jul 2023
<b>WP8</b>	D8.6	Report on the COGITO demonstration activities	FER	31 Aug 2023
<b>WP8</b>	D8.7	Holistic evaluation of COGITO system performance and impact assessment	UEDIN	31 Oct 2023
<b>WP9</b>	D9.2	COGITO dissemination and communication plan and associated material v1	ASM	30 Apr 2021
<b>WP9</b>	D9.3	COGITO dissemination and communication plan and associated material v2	ASM	31 Oct 2021
<b>WP9</b>	D9.4	COGITO dissemination and communication plan and associated material v3	ASM	31 Oct 2022
<b>WP9</b>	D9.5	COGITO dissemination and communication plan and associated material v4	ASM	31 Oct 2023
<b>WP9</b>	D9.6	COGITO Living Lab Activities Evaluation Report v1	ASM	31 Oct 2021
<b>WP9</b>	D9.7	COGITO Living Lab Activities Evaluation Report v2	ASM	31 Oct 2022
<b>WP9</b>	D9.8	COGITO Living Lab Activities Evaluation Report v3	ASM	31 Oct 2023
<b>WP9</b>	D9.9	Report and associated material on COGITO stakeholder engagement & training activities v1	ASM	31 Oct 2021
<b>WP9</b>	D9.10	Report and associated material on COGITO stakeholder engagement & training activities v2	ASM	31 Oct 2022
<b>WP9</b>	D9.11	Report and associated material on COGITO stakeholder engagement & training activities v3	ASM	31 Oct 2023
<b>WP9</b>	D9.14	Report on COGITO standardization outcomes promotion efforts	UPM	31 Oct 2022
<b>WP9</b>	D9.15	Report on COGITO standardization outcomes promotion efforts v2	UPM	31 Oct 2023
<b>WP9</b>	D9.16	Report on best practise examples of COGITO use	ASM	31 Oct 2023

The purpose of creating these reports is to document the outcomes of the project and detail the performed work along the main results. Each document has an already defined title specified in the DoA and presented above.

The documents are stored in the cross-platform portable document format (PDF). All public documents will be published openly in the COGITO webpage (following EC review and approval). Public documents may also be added to Zenodo at the end of the project. For uploading documents to Zenodo the information presented in the Section 3.2.1 should be included.

### 3.2.3 COGITO data types

We are also considering the COGITO data types as part of the DMP. Table 3 presents the defined primary / secondary data types that are collected / produced by each component of the COGITO architecture (see Figure 3). It should be noted that this is an initial attempt to structure the COGITO data types, and as such, they will be further analysed/revised in the next version of the DMP. As COGITO is still at its early stages, the table presented below may be subject to changes. The table summarises the data types handled per component, while providing further information about the WPs, and concerned partners, where the data are handled. As mentioned in Section 2.2, in the coming versions of the DMP, each data type will be further analysed using the DMP template introduced in Section 2.1.

In the table below:

- Primary data is data that have been collected but have not undergone any processing;
- Secondary data is analysed data that have been extracted through processing.

**Table 3: COGITO Datatypes (M6)**

Relevant Component (If applicable)	Data types (primary/ secondary produced from the respective component)		WPs (data generation)	WPs (data usage)	Partners involved
Location Data Acquisition tool	<b>Primary</b>	Raw data gathered by the sensorial equipment	WP3 (tools development) / WP8 (COGITO solution demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	HYP, RSRG, FER, UCL, UPM
	<b>Secondary</b>	Processed cleaned/ normalised data to be stored in the digital twin platform			
Work Order Definition and Monitoring tool (WODM)	<b>Primary</b>	Work order related data entered by the user (Project Manager / Site Manager): actions details, work instructions, workers and other resources Service Level Agreements (SLAs)	WP6 (tool development) / WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	BOC, NT, UCL, HYP, FER, OLOD, RSRG, UPM
	<b>Secondary</b>	Data transformed from the imported workflow (BPMN): schedule, links to related construction components			
Work Order Execution Assistance tool (WOEA)	<b>Primary</b>	Detailed work order execution data: work progress, execution time, reported issues Multimedia received via Remote Assistance (photos, videos, documents)	WP6 (tool development) / WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	NT, BOC, UEDIN, QUE, UCL, HYP, FER, OLOD, RSRG
	<b>Secondary</b>	N/A			

Relevant Component (If applicable)	Data types (primary/ secondary produced from the respective component)		WPs (data generation)	WPs (data usage)	Partners involved
Digital Twin Platform (DTP)	<b>Primary</b>	Configuration meta-data, user accounts, API keys, access control information. Data received from reality capture tools (IoT – time series data, raw data (point clouds), BIM models	WP3 (IoT/Reality capture) / WP8 (demonstration / evaluation)	WP4, WP5, WP6 (various tools development) / WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	CERTH, HYP, UEDIN, NT, BOC, AU, QUE, UPM
	<b>Secondary</b>	Ontological Model of 3D BIM associated with schedules and resources, H/S info, QC and workflow progress			
Process Modelling and Simulation tool (PMS)	<b>Primary</b>	Schedule data including elements IDs	WP6 (tool development) / WP8 (demonstration / evaluation)	WP6 (Adaptive Workflow Modelling and Workflow Management Automation Tools) / WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	BOC, NT, UCL, HYP, FER, OLOD, RSRG, UPM
	<b>Secondary</b>	Optimised workflow data			
Augmented Reality enabled In-situ QC Visualization (DigiTAR)	<b>Primary</b>	Raw image data (as Visual Data Acquisition Tool)	WP3, WP4, WP5 (tools development), WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	CERTH, UCL, AU, FER, RSRG
	<b>Secondary</b>	Notifications regarding task confirmation & completion, remedial work & hazard mitigation			
GeometricQC	<b>Primary</b>	Output of geometric specification control (e.g. deviation numbers, out of tolerance flags, colour maps)	WP5 (tool development), WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	UEDIN, RSRG, FER, UCL, UPM
	<b>Secondary</b>	Point clouds matched to BIM components Segmented building components			
VisualQC	<b>Primary</b>	Geo referenced visual data matched to BIM components	WP5 (tool development), WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	CERTH, RSRG, FER, UCL, UPM
	<b>Secondary</b>	Defect detection notifications to DT Platform			
Digital (Visual) Command Centre (DCC)	<b>Primary</b>	3D BIM associated with IoT data, QC defects/ annotations, H/S issues annotations, workflow progress	WP7 (tool development), WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	HYP, UCL, RSRG, FER, AU, UEDIN, NT
	<b>Secondary</b>	N/A			

Relevant Component (If applicable)	Data types (primary/ secondary produced from the respective component)		WPs (data generation)	WPs (data usage)	Partners involved
BlockChain Platform Tool	<b>Primary</b>	Service Level Agreement (SLAs) output (through design) and Smart Contract Develop & Configuration	WP6 (tool development) / WP8 (demonstration / evaluation)	WP8 (demonstration / evaluation)	BOC, NT, FER, RSRG
	<b>Secondary</b>	Data received from Word Order Monitor Tool for Smart Contract execution			
SafeConAI	<b>Primary</b>	3D BIM associated with schedule and resources Safety rules and regulations	WP4 (tool development) / WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	AU, RSRG, FER, UCL, UPM
	<b>Secondary</b>	4D BIM enhanced with safety info			
ProActiveSafety	<b>Primary</b>	Trajectory of resources	WP3, WP4 (tools development) / WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	AU, RSRG, UCL, FER, CERTH, HYP, UPM
	<b>Secondary</b>	4D BIM enhanced with safety info			
VirtualSafety	<b>Primary</b>	4D BIM enhanced with safety info Trainee feedback / input	WP4 (tool development) / WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	AU, RSRG, FER, UCL
	<b>Secondary</b>	Personalised feedback based on performance in the game generated by the tool			
Geometric (3D) Data Acquisition Tools	<b>Primary</b>	(coloured) 3D point cloud data	WP3 (tool development) / WP8 (demonstration / evaluation)	WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	UEDIN, RSRG, FER, UCL, UPM
	<b>Secondary</b>	Geo/Site-referencing data Spherical images			
Visual Data Acquisition Tools	<b>Primary</b>	Raw image data from various inputs (cameras, UAV, goggles etc.)	WP3, WP8	WP5, WP7 (Digital twin platform) / WP8 (demonstration / evaluation)	CERTH, RSRG, FER, UCL, UPM
	<b>Secondary</b>	Geo/Site-referencing data			

Relevant Component (If applicable)	Data types (primary/ secondary produced from the respective component)		WPs (data generation)	WPs (data usage)	Partners involved
N/A	<b>Primary</b>	<b>Questionnaires:</b> <ul style="list-style-type: none"> <li>- living lab (LL) participants feedback on LL workshops (in terms of organisation, content-related aspects as well as COGITO results)</li> <li>- feedback from other COGITO events from the participants</li> </ul> <b>Dissemination activities:</b> <ul style="list-style-type: none"> <li>- Data for the events recruitment (name and surname, organisation represented, email address)</li> <li>- Photo of a person that participated in the COGITO events and published on COGITO website, in newsletter</li> <li>- E-mail to subscribe to newsletter</li> <li>- Website statistics</li> <li>- Newsletter statistics</li> <li>- Number of deliverables' downloads</li> <li>- Communication materials statistics (e.g. number of people who watched video, downloaded other promotional materials)</li> <li>- Number of people participating at different events where COGITO will be present</li> </ul>	WP2 (user requirements), WP9 (dissemination and communication)	WP2 (user requirements), WP9 (dissemination and communication)	ASM, RSRG, FER, UEDIN
	<b>Secondary</b>	N/A			

### 3.2.4 Scientific Publications

Along with the dissemination of the project deliverables and datasets, further dissemination of project scientific publications is also considered as part of the DMP. At the time of writing the deliverable, there are no scientific publications based on COGITO relevant work. Once a publication is available, it should be documented in the DMP including the following information.

**Publication reference number:** The title and the DOI of the publication

**Publication description:** A short description of the content of the publication

**Standards and metadata:** The type of the document format and any type of metadata associated with the content of the document.

**Data sharing, archiving and preservation (including storage and backup):** how data will be preserved, how to archive and preserve it

## 4 COGITO FAIR Data

The COGITO DMP describes the data management life cycle for the data to be collected, processed and/or generated. As part of making research data Findable, Accessible, Interoperable and Re-usable (FAIR) [5]:

- Findable: unique, steady ID, searchable resource available, and meaningful metadata;
- Accessible: retrievable by standard methods and protocols, metadata accessible regardless of data availability;
- Interoperable: recognised by standard formats and terms;
- Reuseable: clear licenses and accurate meaningful metadata according to community standards. Content traceable,

the COGITO DMP should include information about:

- How the research data are handled for the duration of the project and how they are preserved after the end of it. The type and acquisition method of the project data as well as the parties interested to use them during and after the end of the project;
- The methodology, applicable standards, and the required metadata;
- The private/public accessibility of the collected data. The preservation of the collected data including archiving methods to ensure that publicly funded research outputs can have a positive impact on future research, for policy development, and for societal change.

How DMP covers these requirements is briefly described below.

### 4.1 Making data findable

In general, data is considered to be findable when it is [2]:

- described with metadata and vocabularies in a standardised way;
- assigned with persistent identifiers;
- registered/ indexed in a searchable resource.

In COGITO, all publicly available documents will be published at a cross platform acceptable PDF format along with meaningful metadata, reflecting the origin and content of the document. Name conventions and clear versioning should be used. Publicly available documents should be uploaded to the COGITO website (upon EC's approval) and may also be uploaded at Zenodo [3] for long term preservation after the end of the project. Zenodo follows FAIR principles. In particular, for making data findable [4]:

- A DOI is issued for every published record on Zenodo;
- Zenodo's metadata is compliant with DataCite's Metadata Schema [5] minimum and recommended terms, with a few additional enrichments;
- The DOI is a top-level and a mandatory field in the metadata of each record;
- Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing;
- Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

Scientific publications will be published, using Digital Object Identifiers (DOIs), to facilitate retrieval for potential users. Descriptive metadata, like title, authors, project, abstracts, keywords, etc, will also be used.

Datasets should have a unique, persistent ID, located in the COGITO Digital Twin, which will be the main repository of the various datasets handled within the course of the project. Specific metadata standards will be used that will contain vital information to enable exchange and interoperability (see also Section 4.3). Open file formats will be used when possible. The DMP template will be used to document related information of each data type in the next versions of the DMP.

## 4.2 Making data openly accessible

Open access can be defined as the practice of providing on-line access to scientific information that is free of charge to the reader. In the context of R&D, open access typically focuses on access to 'scientific information' or 'research results', which refers to two main categories [6]:

- Peer-reviewed scientific research articles (primarily published in academic journals); and
- Research data.

These two main aspects of data accessibility are considered in the following sections.

Following FAIR principles, COGITO will make data openly accessible when possible. Despite not the entirety of the data can be made publicly open, it can still be FAIR [7]. In general, data are accessible when uploaded in a data repository and retrieved by their PIDs. When data cannot be shared openly, metadata (mainly bibliography) should be provided. In the case of sensitive or personal data, (pseudo) anonymisation and specific provisions for managing access rights should be applied. Where accessing data requires additional tools / methods, these should be documented accordingly.

### 4.2.1 Open Access to scientific publications

The open access mandate comprises two steps:

- depositing publications in repositories; and
- providing open access to them

Open access publishing is defined as the free availability of peer-reviewed scientific publications for any user. The COGITO Grant Agreement in §29.2 specifies that:

“Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results”. In particular, it must:

- As soon as possible, or at the latest at publication date, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;
- Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications;
- Ensure open access to the deposited publication - via the repository - at the latest:
  - on publication, if an electronic version is available for free via the publisher, or
  - within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.
- Ensure open access - via the repository - to the bibliographic metadata that identify the deposited publication.

The bibliographic metadata must be in a standard format and must include the following:

- the terms “European Union (EU)” and “Horizon 2020”;
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable, and;
- a persistent identifier.

The two main routes to Open Access are:

- **Self-archiving / 'green' Open Access** – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access is granted only after the embargo period expired.

- **Open Access publishing / 'gold' open access** - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors.

To achieve this, the COGITO Consortium will use both “green” and “gold” open access routes as described before. The project coordinator commits to ensure that the outputs of COGITO research will be readily accessible. All scientific publications arising from COGITO project, with the consent of the authors and in compliance with other publishers’ policies will be hosted at the project’s website (<https://cogito-project.eu/>).

COGITO will make other public deliverables (demonstrators and reports as presented in Section 3.2), working papers and conference papers, which are not scientifically peer reviewed, openly accessible via the project website, and other online research dissemination platforms such as ResearchGate and OpenAIRE’s Zenodo repository. Zenodo follows FAIR principles. In particular, for making data openly accessible [4]:

- Metadata for individual records, as well as record collections, are harvestable using the OAI-PMH [8] protocol by the record identifier and the collection name. OAI-PMH is a free and universal protocol for information retrieval on the web [4];
- Metadata is also retrievable through the public REST API, an architectural style that uses open, free and universal standards for information retrieval on the web;
- Metadata are publicly accessible and licensed under public domain. No authorisation is ever necessary to retrieve it;
- (Meta)Data will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for, at least, the next 20 years;
- Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

Non-public documents (classification: confidential in Figure 5) generated throughout the project will be collected in an internal project repository (COGITO Nextcloud). This repository is managed by the project coordinator. Access to the platform is protected through a login and password procedure. This platform is used by project partners to store and share working documents and deliverables. Non-public documents include:

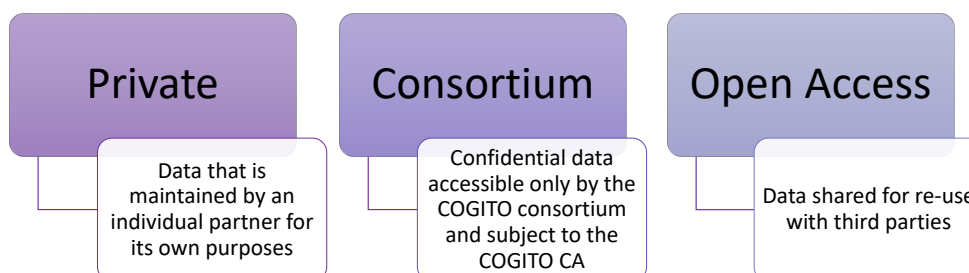
- Restricted data such as the ones needed for internal management and communication which will not be published online and will not be made available to the public;
- According to COGITO DoA, in contrast to public deliverables, deliverables that have been characterised as confidential will not be shared.

#### 4.2.2 Open access to research data

Research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion or calculation. Open access to research data is the right to access and reuse digital research data under the terms and conditions set out in the project’s Grant Agreement [6]. Research data in COGITO is assumed to be the datasets presented in Section 3.2.3 (as will be updated during the project lifecycle).

Decisions regarding data access and licensing strategies will be made in due time, once the research datasets are available. Future versions of the present document will report all relevant details. In the case of COGITO datasets containing personal data (e.g., names, addresses, etc), it will be carefully examined by the consortium whether these data will be shared after the subject's consent. To this end, datasets in COGITO should be analysed, and any ethical or legal issues (e.g. personal data, confidential data, etc.) which can impact sharing the data, should be identified. Based on this analysis, COGITO Consortium should decide whether the datasets (all, none or even part of it) will be openly accessible using the classification levels shown in Figure 5. In case there are restrictions, relevant explanations should be provided. For the openly accessible data, COGITO should provide necessary information about the data repository used for its archiving. Details about the storage security and available

backup / recovery procedures may also be provided. Any method(s) and/or tool(s) required for accessing the data should be described (e.g. visualiser software).



**Figure 5: Data classification levels in terms of accessibility**

Many data is gathered by the project for the purpose of project monitoring and delivery rather than new knowledge creation; it is therefore likely that such data are categorised as Consortium. However, the project will seek to openly disseminate its research findings, except in cases where there are defined exploitable outcomes, privacy concerns (e.g. tracking data, personal information accessed through COGITO Living Labs, etc.) or there will be a high administrative burden for a dataset or limited benefit to other users. Any data that will be made openly available will undergo anonymisation and/or aggregation where it is required to protect the identities of project participants. Before publishing data types, even after (pseudo)anonymisation, relevant consent forms must be signed.

### 4.3 Making data interoperable

Data interoperability refers to handling data that can be easily understood and shared with other platforms and systems, when they are created using standard vocabularies and include references to other data and metadata [7].

In COGITO, the publishing of publicly available documentation in Zenodo supports their interoperability through Zenodo's relevant FAIR principle [4]:

- Zenodo uses JSON Schema [9] as internal representation of metadata and offers export to other popular formats such as Dublin Core [10] or MARCXML [11];
- For certain terms, Zenodo refers to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE);
- Each referenced external piece of metadata is qualified by a resolvable URL.

Furthermore, regarding COGITO data types, one of the main goals in COGITO is to have interoperable data flowing through the COGITO Digital Twin platform; this will also support the interchange and reuse of data with other stakeholders external to the project.

In order to have interoperable data, COGITO will use open standards and formats facilitating the integration of data coming from different sources. For interchanging data, the preferred format will be Resource Description Framework (RDF), since it facilitates re-combinations with different datasets from different origins and eases the definition of links between datasets.

Data in COGITO will be described according to the COGITO ontology, which will be based on and will reuse existing standard ontologies (when possible) to support interoperability. When needed, mappings among the COGITO ontology and other standard ontologies will be defined in order to facilitate data exchange.

The metadata that will be used to describe the different COGITO datasets will also be interoperable, based on well-known standards, and available through standard interfaces.

#### 4.4 Increase data reuse

Data can be reused when the conditions about how others can make use of the data are well-described following community-standards and are communicated as specified by the data owners [7].

As indicated previously, COGITO public documentation will be made available on Zenodo. Zenodo follows FAIR principles. In particular, for increasing data reuse [4]:

- Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments;
- License is one of the mandatory terms in Zenodo's metadata, and is referring to an Open Definition license;
- Data downloaded by the users are subject to the license specified in the metadata by the uploader;
- All data and metadata uploaded are traceable to a registered Zenodo user;
- Metadata can optionally describe the original authors of the published work;
- Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

In case of software products, well-established online repositories (e.g. GitHub) can be used to disseminate the respective outcomes.

In this section, we also provide some high-level information of Intellectual Property Rights (IPR) and the licensing to be used within COGITO to make the data and the open-source products reusable.

##### 4.4.1 IPR Management

COGITO, being a market oriented H2020 project, has considered different activities related to exploitation and IPR management to make the results sustainable after the project end and to ensure that it will be further used by project partners and beyond. Exploitation and IPR management are foreseen in Task T9.3 “Exploitation Strategy and Preliminary Business Innovation Planning”, which starts in M19. However, already at proposal stage, some basic rules regarding IPR have been provided including:

- background know-how by all project partners, respecting each other's IPRs, providing patents, copyrights, or otherwise protecting IP at fair and reasonable conditions;
- agreement of all Partners on co-ownership of foreground based on the actual partner contribution in the particular result.

Moreover, COGITO partners have also agreed to respect each other's rights during project results dissemination, especially regarding the publication of foreground information. Specific processes will be implemented to consult the result with relevant partners in due time before publishing sensitive information. As already mentioned, project publications will be accessible in open repositories such as European Commission's OpenAIRE and Zenodo, and all partners have committed to provide Open Access to all scientific publications (free of charge online access for any user) under both the “green” and “gold” open access schemes.

It is also worth mentioning that key issues regarding the IPR are also included in the Consortium Agreement, including:

- definition of the foreground and background knowledge concerning the project;
- assignation of rights shares of the anticipated results and access granted to the other participants;
- conditions for licensing to third parties to use the foreground;
- confidentiality regarding other non-COGITO project members;

- responsibilities for detecting and taking civil and criminal action against third parties that in any way injure the rights conferred by the IP rights;
- means and conditions for internal and external exploitation of project results.

Different tools and activities will be performed from M19 to manage the IPR issues and ensure proper exploitation of project results. These include:

- collaborative tools for IPR management (IPR control spreadsheet) and risk assessment;
- questionnaires for detailing Partners' plans and strategies with regards foreground and background;
- webinars to discuss the joint exploitation opportunities especially with regards the main outcome (BIM-Digital Twin harmonization).

Individual project partners will build software products which may be protected by the specific partners. Patents Landscape Analysis will be updated during the project to ensure that competing entities do not hold IPRs conflicting with the COGITO toolbox. At the proposal preparation stage, partners have performed due diligence on the freedom to operate, and have verified that sufficient freedom exists to penetrate the market and claim a significant share without IP protection barriers from competitive companies.

#### 4.4.2 Copyright Licenses

When material is widely shared, copyright licences protect the authors of work and grant specific rights to publishers and others to use this work. The European Commission [12] encourages authors to retain their copyright whilst disseminating it as open access. Creative Commons provides legal tools to enable open access in these circumstances, with CC-BY (Creative Commons Attribution International licence) and CC0 (Creative Commons No Rights Reserved licence) enabling reuse by third parties. The data sets may be given different licenses according to their specificities<sup>2</sup>.

Where research findings are published in a journal, or other scientific outlet, there should be consideration of the copyright agreement with the publishers, which may involve an embargo period. Submission in project repository and website requires the author to agree to a non-exclusive distribution licence, and a Creative Commons licence may be added at this point.

In this initial stage it is not possible to define the copyright arrangement for each project dataset. The most appropriate licencing arrangements for each of the project datasets will be investigated later in the course of the project as they are better characterised by their respective work packages and the COGITO Exploitation strategy. The data and ethics management reports, and their updates, also address this aspect.

<sup>2</sup> The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data (<https://eudat.eu/services/userdoc/license-selector>)

## 5 Allocation of resources

### 5.1 Human Resources

The DMP presented in this deliverable aims to identify which project outputs are to be disseminated, as well as, to decide on the way and means of storage and their Open Access (if applicable). To ensure it, a dedicated time slot will be reserved at each of the project plenary meetings and, if needed, at selected consortium teleconferences. EC and project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsibilities on data management in the project consortium are:

- Data Management Plan Leader (HYPERTECH) – Hypertech has assumed this role as project coordinator and leader of the WP1 “Quality Assurance and Project Management”. Hypertech is responsible for preparing and leading related discussions at the relevant project meetings and for maintaining the channels for dissemination of project outcomes.
- Ethics Monitoring Board (HYPERTECH, UEDIN, ASM) – An ethics monitoring board was established during the Kick-off-Meeting of the project to study and address any legal, ethical, and privacy issues for the technologies developed by the Consortium, and overview the preparations and execution of the COGITO demonstration and validation activities. The Ethics Monitoring Board will address privacy issues related to data collection and handling, providing guidelines and feedback to the COGITO partners, as well as interface with the research participants in relation to potential ethical and privacy concerns. The Ethics Monitoring Board will be assisted by further external experts (or the commission), if necessary.
- To ensure that the national legislation of each pilot site will be considered, during the demonstration and validation phase of the project, each pilot site (incl. the pre-validation site) will have its own ethics committee and one person will be nominated per site as responsible for adhering to the project’s recommendations, as well as the National and European legislation.
- Technical Manager (UEDIN) – UEDIN has assumed this role. Technical Manager is responsible to identify data collected / handled by the project and technical project outcomes eventually suitable for publication.
- Scientific Manager (UCL) – UCL has assumed this role. Scientific Manager is responsible to coordinate with the Technical Manager the process of identifying data collected / handled by the project and project outcomes eventually suitable for publication.
- Dissemination and Exploitation Manager (ASM) – ASM has assumed this role. Dissemination and Exploitation Manager is responsible to collect published work and ensure their placement in the considered repositories while maintaining COGITO inputs for Open Access.
- Each individual partner – to identify own project results suitable for publication.

All COGITO partners must respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP.

Backing up data for sharing through Open Access repositories is the responsibility of the partner possessing the data.

Quality control of these data is the responsibility of the relevant WP leader, supported by the Data Management Plan Leader.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.

## 5.2 Financial Resources

At this stage of the project, it is not possible to define whether additional financial resources need to be allocated for the data management within the project and if such costs arise, they would be associated with the cost of storage of collected data or fees to be submitted to public repositories. At this time, however, no special costs are expected to occur.

## 6 Data Security

### 6.1 GDPR and DPIA

According to the European regulation (EU) 2016/679, the new General Data Protection Regulation (GDPR), is a set of rules defining how individuals, companies and organisations collect, store and process personal data related to individuals in the EU [13]. As a key instrument of the GDPR, the Data Protection Impact Analysis (DPIA) is recommended to enhance the accountability of Data Controllers and demonstrate their compliance with the GDPR. The main purpose of a DPIA is to enable the entities responsible for data collection and processing to identify possible risks for data protection, and to establish rules, in particular, concerning the amount of collected data and the purpose of processing.

A DPIA is required whenever processing is likely to result in a high risk to the rights and freedoms of individuals. A DPIA is required at least in the following cases [14]:

- a systematic and extensive evaluation of the personal aspects of an individual, including profiling;
- processing of sensitive data on a large scale;
- systematic monitoring of public areas on a large scale.

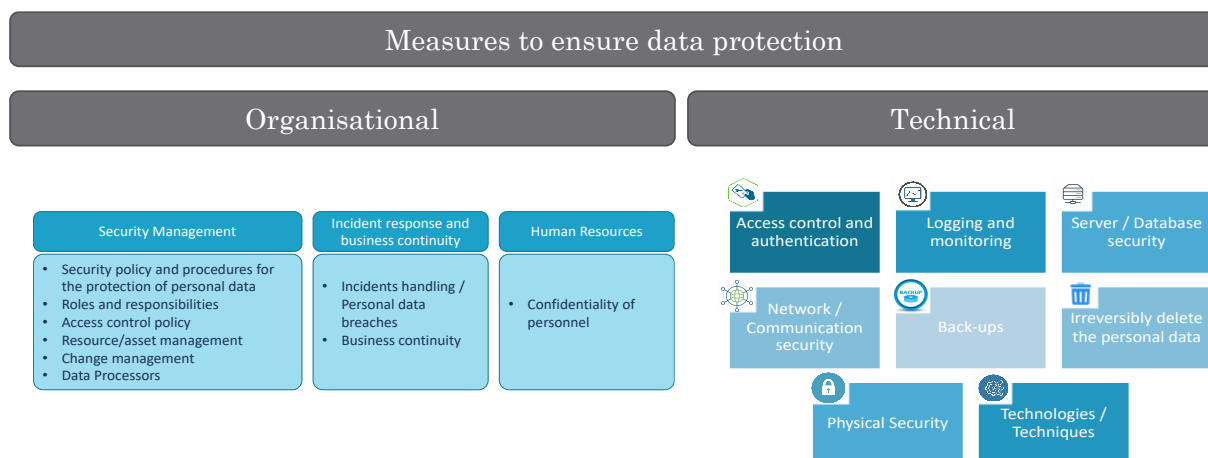
At the point of writing this document, the abovementioned cases are not met in the context and planning of work execution in the COGITO project. Furthermore, no high risks to the rights and freedoms of individuals have been identified. This will continuously be evaluated during the course of the project and re-assessed whenever deemed necessary. In the next versions of the DMP, when the project will be at a more mature stage and the Consortium will have a clearer view of all data types that will be generated, processed and stored throughout the project implementation, this section may need to be updated.

### 6.2 Measures to ensure data protection

To innovate ethically and responsibly, following the guidelines and best practices of European Commission [6], COGITO applies the concept of ‘privacy by design’, utilising a framework where systems, databases and processes are designed and developed in way that safeguards the fundamental rights of data subjects. The wider concept of ‘data protection by design’, now included in the GDPR, requires data controllers to implement appropriate technical and organisational measures to give effect to the GDPR’s core data-protection principles (articles 5 and 25 of GDPR). COGITO measures to achieve data protection by design, as part of the research and development context, include:

- data minimisation - COGITO will collect only the data that are needed for achieving COGITO objectives. Data in COGITO will be collected on a ‘need to know’ basis. More specifically, COGITO will collect and manage only the data which are required for a specific purpose that is relevant and limited to its objectives and aligned with the work plan description;
- technical and organisation measures description (shown in Figure 6) - considering that COGITO may process personal data, an evaluation of the relevant risk level should be performed. Following this risk level evaluation, COGITO will proceed with the selection of appropriate security measures for the protection of personal data. Then, the COGITO Consortium will describe the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants. Guidelines provided by ENISA for the security of personal data processing [8] will be followed;
- arrangements that enable data subjects to exercise their fundamental rights (e.g., as regards direct access to their personal data, consent to its use or transfer, make people aware of any tracking/profiling, etc.);
- pseudonymisation or anonymisation of personal data;

- detailed description of COGITO data handling including, for example description of the applied cryptography (e.g., encryption).



**Figure 6: Technical and organisational measures to ensure data protection. More detailed description can be found in COGITO D10.2 “POPD Requirements”**

### 6.3 Data collection, storage and destruction

During the proposed system lifecycle, a holistic security approach will be followed to ensure information security (confidentiality, integrity, availability) and to protect from a potential misuse. Tasks mainly focused on the definition and development of the data security framework to be applied for data types’ collection and storage within the COGITO solution include:

- T3.4 “IoT Data Pre-processing Module”;
- T3.5 “Visual Data Pre-processing Module”;
- T7.1 “Digital Twin Platform Design & Interface Specification”; and
- T7.2 “Extraction, Transformation & Loading Tools and Model Checking”.

So, a detailed analysis will be performed in the aforementioned tasks to ensure that the data exchange in COGITO is constantly protected by transmitting encrypted information through secure channels. The key security measures that will be finally used for protecting COGITO data types will also be detailed in the next versions of the DMP.

To protect the collected data types and control access to the COGITO toolbox, only authorised users will have access to use case-specific collected data. COGITO solution will provide secure access using security protocols, for example OAuth2 for role-based authorisation and secure encrypted TLS / HTTPS calls. The user interfaces will offer security services like identification, authentication and authorisation of the relevant actors, while introducing data security and privacy awareness to every service in the system.

Each data owner partner involved in managing a specific dataset will be responsible for the archiving and preservation of the corresponding data. As a general rule, responses to questionnaires and interviews will be stored on paper and/or computer files protected by physical and technical security measures.

Regarding, the document storage (classified as Public or Consortium level), Nextcloud and Zenodo repositories will ensure secure and safe storage of open and confidential data. More specifically:

- Zenodo provides clear security guarantees. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files and metadata are backed up on a nightly basis. Files are regularly checked against their checksums (using MD5 algorithm) to assure that file content remains

unvarying. In case of closure of the repository, Zenodo ensures that efforts will be made to integrate all content into suitable alternatives [3].

- Nextcloud aligns with industry standards such as Clause 14 of ISO/IEC27001-2013 and related standards, guidance and security principles. It is built around combined assurance layers consisting of rich security features, applied best practices governed by policy and the design itself validated by industry standard testing processes [15].

Regarding data destruction, as computerised data (hard disk drives) will be used for data storage, existing methods for permanent and irreversible destruction of the data will be utilised (i.e., full disk overwriting and re-formatting tools).

## 7 Ethical Aspects

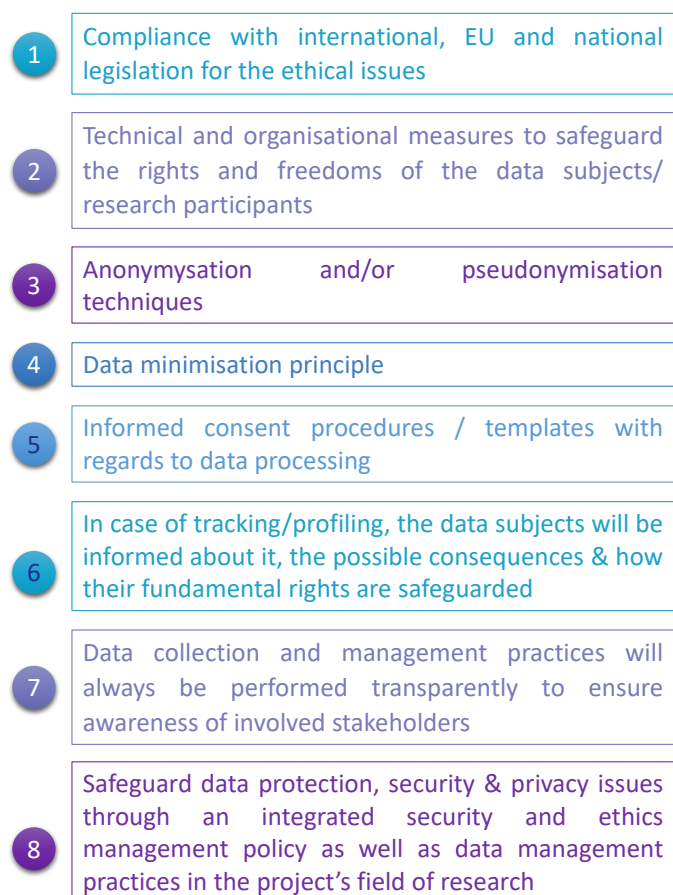
The COGITO consortium is fully aware that a number of ethical, privacy and data protection issues will be raised by the activities to be performed in the scope of the project. For avoidance of repetitions, the reader is advised to refer to the following already submitted deliverables (only a brief summary is included here):

- D10.1 H- Requirement No. 1;
- D10.2 POPD - Requirement No. 2.

More specifically, in the D10.1 H- Requirement No. 1 ethics requirements about humans are presented. More specifically, the document includes information about:

- the procedures and criteria that will be used to identify/recruit research participants;
- the informed consent procedures that will be implemented for the participation of humans before the start of the ethically impacted task;
- the template of the informed consent/assent forms and information sheets (in language and terms intelligible to the participants); and
- clarifications on whether children and/or adults unable to give informed consent will be involved and, if so, justification for their participation.

This deliverable is complemented by D10.2 POPD - Requirement No. 2, which sets out ethics requirements considering protection of personal data (POPD) that the COGITO project must comply with. The COGITO rules for POPD are summarised in the figure below.



### Ethics POPD Requirements Principles

Figure 7: COGITO rules for the protection of personal data

In these documents, the ethics methodology that will be applied in COGITO to ensure compliance with all the applicable ethics requirements is described. This way, COGITO ensures that the ethical standards and Horizon 2020 framework's related guidelines are always respected. The aforementioned deliverables D10.1 and D10.2 complement the DMP. All COGITO partners must fully respect their principles and rules.



## 8 Conclusions

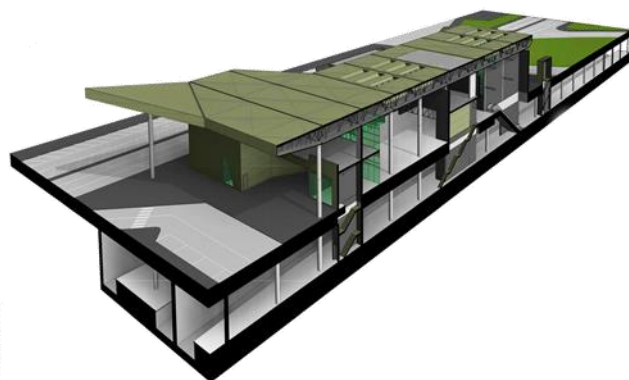
The first version of this document provides the methodology to be used within COGITO for data management and the DMP template that will be used to detail the identified data types that will be managed in the COGITO solution.

This document presents also the principles and procedures to be followed in COGITO for data management proper handling of the research data. Main topics covered are: open access policy, data handling and sharing, and procedures that will be followed for secure data collection, storage, retention and destruction.

The document has been based on guidelines and EC's recommendations for managing FAIR data. A first attempt to identify research data is presented. More information and a complete filled DMP template will be provided in the next versions of the DMP – which is supposed to be a living document – when COGITO partners will have a clearer view of the various data types that the COGITO platform will deal with.

## 9 References

- [1] "H2020 Programme Guidelines on FAIR Data Management in Horizon 2020," 2016.
- [2] "how-to-create-a-data-management-plan." [Online]. Available: <https://www.openaire.eu/how-to-create-a-data-management-plan>. [Accessed: 20-Apr-2021].
- [3] "Zenodo - Research. Shared." [Online]. Available: <https://zenodo.org/>. [Accessed: 26-Jun-2019].
- [4] "Zenodo." [Online]. Available: <https://about.zenodo.org/principles/>. [Accessed: 25-Apr-2021].
- [5] "DataCite Schema." [Online]. Available: <https://schema.datacite.org/>. [Accessed: 25-Apr-2021].
- [6] European Commission, "Open access & Data management - H2020 Online Manual." [Online]. Available: [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm). [Accessed: 25-Apr-2021].
- [7] OpenAIRE and EUDAT, "Open DMP." [Online]. Available: <https://argos.openaire.eu>. [Accessed: 25-Apr-2021].
- [8] "Open Archives Initiative Protocol for Metadata Harvesting." [Online]. Available: <https://www.openarchives.org/pmh/>. [Accessed: 25-Apr-2021].
- [9] "JSON Schema | The home of JSON Schema." [Online]. Available: <https://json-schema.org/>. [Accessed: 29-Apr-2021].
- [10] "DCMI: Home." [Online]. Available: <https://dublincore.org/>. [Accessed: 25-Apr-2021].
- [11] "MARC 21 XML Schema." [Online]. Available: <https://www.loc.gov/standards/marcxml/>. [Accessed: 25-Apr-2021].
- [12] "Open access - H2020 Online Manual." [Online]. Available: [http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm). [Accessed: 26-Jun-2019].
- [13] "What does the General Data Protection Regulation (GDPR) govern? | European Commission." [Online]. Available: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern_en). [Accessed: 26-Jun-2019].
- [14] "When is a Data Protection Impact Assessment (DPIA) required? | European Commission." [Online]. Available: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_en). [Accessed: 25-Apr-2021].
- [15] "Nextcloud." [Online]. Available: <https://nextcloud.com/>. [Accessed: 26-Jun-2019].



# COGITO

CONSTRUCTION PHASE  
DIGITAL TWIN MODEL

[cogito-project.eu](http://cogito-project.eu)



THE UNIVERSITY  
of EDINBURGH



POLITÉCNICA



ferrovial  
construction



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958310